

doi:10.20270/j.cnki.1674-117X.2026.1001

文学与人工智能跨学科研究笔谈

蔡熙，夏明凤，黄巧心，向丰田，雷秋香

（湘潭大学 文学与新闻学院，湖南 湘潭 411105）

摘要：人工智能的兴起，以前所未有的方式改变了文学创作和实用写作的传统范式，引发了文学和文化生产方式的根本性变革，催生出新的文学生产主体和新的文本生成模式。人工智能介入的文学为人类提供乌托邦般的情感寄托，同时，文学的实践主体也开始融入“非人”的因素，产生“类文学意识”。从跨学科的视角出发，深入探讨文学与人工智能的关系是当前一项重要学术议题。

关键词：人工智能文学；跨学科研究；人机共生；文学情感；类文学意识

中图分类号：I207.2 文献标志码：A 文章编号：1674-117X(2026)01-0001-13

Commentary on Interdisciplinary Research in Literature and Artificial Intelligence

CAI Xi, XIA Mingfeng, HUANG Qiaoxin, XIANG Fengtian, LEI Qiuxiang

(School of Literature and Journalism, Xiangtan University, Xiangtan 411105, China)

Abstract: The rise of artificial intelligence has changed the traditional paradigms of literary creation and practical writing in an unprecedented way, triggering fundamental changes in the production methods of literature and culture, and giving rise to new subjects of literary production and new text generation models. Literature with the intervention of artificial intelligence provides humanity with utopian emotional sustenance. Meanwhile, the practical subjects of literature have begun to incorporate “non-human” factors, giving rise to “quasi-literary consciousness”. An in-depth exploration of the relationship between literature and artificial intelligence from an interdisciplinary perspective is undoubtedly an important academic topic at present.

Keywords: artificial intelligence literature; interdisciplinary research; human-machine symbiosis; literary sentiment; quasi-literary consciousness

人工智能文学：人机协作与提示词叙事

蔡熙

当今时代科学技术迭代升级，以ChatGPT和DeepSeek等为代表的人工智能依靠大语言模型自

动生成文本，在文学写作、机器新闻写作、文学翻译、人文研究等方面都能产生新的成果，并且

收稿日期：2025-09-10

基金项目：湖南省哲学社会科学基金“学术湖南”精品培育项目“西方文论中的中国元素研究”（24ZDAJ013）

作者简介：蔡熙，男，湖南永州人，湘潭大学教授，湖南省芙蓉学者特聘教授，博士，博士生导师，研究方向为文学理论。

往往“一秒得万句”。相比之下，依靠生活积淀和夙兴夜寐思考的传统文化人写作却只能“两句三年得”。人工智能的兴起，以前所未有的方式改变了艺术创作和实用写作的传统范式，引发了文学和文化生产方式的根本性变革，催生出新的文学生产主体和新的文本生成模式。因此，在新文科背景下探讨文学与人工智能的跨学科研究，是当前一项重要学术议题。

人工智能的技术基础是算法、算力和数据。

“算法”（algorithme）指的是人们给计算机下达的一系列指令。算法是人工智能研究的主要产品，也是人工智能领域的大道所在。在人工智能领域，任何理论或想法都先被写成算法，然后进行测试。人工智能的算法涵盖了人类和不同物种能够完成的所有任务或行为。因此，人工智能赋能的文学研究可以提供新的工具和方法，与传统的文学研究相比，其在文本分析、数据挖掘、风格分析、跨语言研究等方面具有明显的优势。下面举一个例子。笔者在 DeepSeek 输入提示词：“有一篇微型小说，题目是《电话》，其内容如下，请对它进行分析或者赏析。”

老太太接到了一个诈骗电话，对方谎称是她儿子，但她儿子在两年前已经病逝了。

可那骗子的声音和她儿子实在太像，她舍不得挂断。

骗子说得口干舌燥，发现根本骗不了她。

她这才说出了真相，最后请求对方：“最后再说一句吧！”

骗子思考片刻后说：“妈，保重啊！”然后挂断了电话^[1]。

不到10秒钟，DeepSeek 评论微型小说《电话》的文章就出来了。

这篇微型小说虽然只有短短几段，却包含了强烈的情感冲击和深刻的社会思考。它以极简的文字，构建了一个既荒诞又感人的故事，揭示了人性中的复杂与温情……

DeepSeek 从情节结构与叙事技巧、情感冲击与人性探讨、社会现实的映射、结尾的留白艺术四个方面，对作品进行了全面的分析。可见，DeepSeek 作为大模型，基于自然语言处理技术，通过算法技术能够快速分析文本，判断微型小说《电话》的主题、叙事结构和叙事风格等，大大

提升了研究效率。尤其值得称道的是，DeepSeek 还将《电话》与欧·亨利的三部经典短篇小说进行了对比：

这篇小说让人联想到欧·亨利式的“意外结局”，但不同于欧·亨利的幽默讽刺，它更偏向于温情与悲凉的交织。

DeepSeek 作为 AI 工具能通过数据搜索构建文学影响的网络，揭示作家间的隐性关联。智能化的基础是数据，大数据是人工智能的重要来源。数据挖掘技术“作为知识发现过程，它通常包括数据清理、数据集成、数据选择、数据交换、模式发现、模式评估和知识表示”^[2]。DeepSeek 通过数据挖掘技术帮助我们发现中西文学之间的隐性联系以及中西小说中情感表达的差异，帮助研究者发现潜在的叙事模式。在 DeepSeek 的帮助下，文学研究者能快速实现资源的信息整合和论文文本生成，迅速完成一篇规范论文的写作。

人工智能介入的文学，无论是创作还是研究，有两个非常明显的特点：一是写作主体的变化。人工智能导致创作主体的消弭，文学的实践主体开始融入“非人”因素。二是人工智能的写作都离不开提示词。

人工智能文学是人机协作的结晶。人工智能并非人类，它没有希望，不会渴求，也无感觉；它既没有意识，也不能反思。人工智能并非一个独立的实体，它如同一种连接型的媒介，一方面连接着人类，另一方面则连接着各类大语言模型和智能终端。智能写作是人类的创造物，在人类制造的机器上体现人类设计的过程。在人工智能写作的过程中，作者的角色或者说人类的身份发生了根本的变化，其从传统的直接创作者转变为与人工智能协作共事的合作伙伴。智能文学创作中作者角色的演变也为我们重新思考人工智能文学的创造力提供了新的视角。

文学自诞生以来，无论是口传文学还是书面文学，通常都被认为是少数天才的事业。在探讨作品诸要素与艺术家的关系时，无论是再现说还是表现说，我们都强调作者的主体性。但是，在人工智能的创作中，算法居于主导地位。人类为机器定义一个问题，然后，机器在人类无法企及的领域中运行。借助“生成式人工智能”（AIGC）对语词及语图关联的预训练，它就能够生成各种

类型、风格迥异的作品，呈现出自动生成性特征。人类作者仅仅负责设定概念模型、知识规则、词典条目以及提示词等基础元素，为大模型的文学创作搭建框架。这样，文学创作过程在某种程度上与人类的直接书写已经没有多少关联，在创作过程中人类作者已经没有个人语言和身体的参与，更谈不上激情和性格的投入。换言之，人类作者已经被去个性化。

在人工智能创作中，随着机器算法和大数据的深度介入，文学的实践主体开始融入“非人”因素。文学创作的主体发生了裂变，其不仅关涉写什么、怎么写的问题，还关涉“谁在写”的问题。文学作者从传统的“人”演变成了“人机合作”的AI作者，传统创作的边界已然模糊。可见，生成式人工智能在人机交互中呈现一种新的主体性。虽然人工智能生成的内容并非完全是机器的产物，从数据库的来源、训练过程，人工智能都体现了人类的某种意志，但是，人类毕竟不再是“道成肉身”的作者，写作变成了一种混合主体的创作行为，一种新型的人机伙伴关系正初露端倪。这一系列文本生产机制的变化，催生出某种具有本体论特征的、新型的跨学科研究。

提示词是人工智能写作中体现人类创造性的核心元素。在生成式人工智能的写作中，人工智能可以根据人的指令，在短时间内生成符合人类要求的图像或文字作品。人类与机器的互动，建构着一个远比人类社会更复杂的人机社会，人工智能和人类将成为协作共事的伙伴。交互性是人工智能与人类协作共事的本质所在。使用智能工具写作时，用户 - 创作者是以提示词的方式与 AI 大语言模型进行互动的。提示词的质量十分关键，是人机交互过程中体现人类创造性的核心元素。王峰团队利用提示词让大模型生成情节连贯的小说文本，发现将提示词输入大模型时，高质量的提示词写作十分关键。这是大模型介入文学写作之后出现的一种新的叙事现象。在《提示词工程：智能长篇小说的核心驱力》一文中，王峰提出建立“提示词叙事学”^[3]，将“提示词工程”作为 AI 创作的“手法”纳入文论领域进行思考。

提示词的本质是提问，它考验着人类发现问题、提出问题的能力。法国哲学家吉尔·德勒兹坚持生命的权力。他认为，普遍性的生命以创造

性的方式最大限度地发挥自身的潜能而得以进化，这就是生命提升自身的力量时产生的一种特定的发展方向。生命在于提出问题，这是它自身的权力。生命提出问题——不仅思考存在物，而且思考所有的生命。对德勒兹来说，哲学就是创造概念。“任何概念都诉诸一个或数个问题，如果没有这些问题便不会有意义。”^[4] 概念只有根据它们与之相关的问题才具有意义。问题是敞开意义域和构成概念创造基础的行为，提问是问题的表达，概念是一种对问题的回应。机体、细胞、机器和声波都是对生命力量的回应。人类输入提示词，AI 根据提示词来写作。提示词是体现人类创造性的场域，其质量决定着作品的质量，它是人工智能创作活动中最能体现人的创造力的环节。在人机协作的写作过程中，提示词是大模型可以有效执行的自然语句，是体现人类创作意图和预期目标任务的说明性文字，是一系列用以诱导大语言模型作出响应的指令。一方面，它必须符合机器语言的特点，要求简洁、明晰；另一方面，它又必须符合算法逻辑，要有明确的目标和清晰的任务。“大模型不断升级，我们需要根据技术发展不断优化提示词叙事质量，探索多元化的提示词叙事不断更新这一叙事学框架。”^[5]

人工智能技术是史无前例的颠覆性技术，它正在改变人类的思想、知识、感知和现实，并引发人类历史进程的变迁。同样，人工智能技术对文学的影响也是颠覆性的。相比于文学创作，人工智能技术对文学研究的影响更为明显。人工智能文学涉及计算机、文学、语言、哲学、古籍、数学、地理、逻辑学、符号学等多个领域，推动了基于科技与文学共生关系的新型文学跨学科研究。

人工智能技术在跨语言与跨文化研究方面有其独特的优势。例如，机器翻译打破了语言壁垒，使研究者能更便捷地接触非母语文本，带来语言和文化的整合，促进人机对话和跨语境交流；人工智能也可辅助分析翻译中的文化损耗问题，利用自己的记忆产生越来越具体的概念，这对用户不无裨益。但是我们也要清醒地看到，人工智能缺乏创造性的解读，缺乏人文批判性思维，它不能对其发现的东西进行反思；它只是在运用算法、算力和数据的基础上产生一个结果，而不管这个结果从人类的角度来看是平庸的还是深刻的，是

善的还是恶的。人工智能也不能反思，因为其行动的意义最终由人类决定。特别值得指出的是，为了迎合人类的偏好，人工智能在处理大量信息的同时，往往加大对世界本来面貌的扭曲程度。因此，利用人工智能进行文学研究仍然需要人类学者的参与，仍需结合传统的研究方法，这样才能确保研究的准确性和深度。

参考文献：

[1] 冬天惠铃. 2024年12篇最佳微小说，短到极致，却催

人泪下……[EB/OL]. [2025-08-25]. http://www.360doc.com/content/24/0405/06/3980569_1119461414.shtml.

- [2] HAN J W, KANBER M, JIAN P. 数据挖掘：概念与技术 [M]. 范明，孟小峰，译，北京：机械工业出版社，2007：25.
- [3] 王峰. 提示词工程：智能长篇小说的核心驱力 [J]. 南方文坛, 2025(1): 62-67.
- [4] DELEUZE G, GUATTARI F. What is Philosophy? [M]. BURCHELL G, TOMLINSON H, Translated. New York: Columbia University Press, 1994: 22.

人工智能时代的文学情感

夏明凤

人工智能作为近几年不断蓬勃发展的事物，在比较文学学界是一个崭新且富有争议的话题，受到业内学者的重点关注。特别是在“新文科”建设作为文科发展新方向的今天，人工智能以其自身拥有的跨学科潜能，成为新文科发展不可缺少的内容。当下，比较文学研究关于人工智能的讨论主要分为以下两种：其一是将人工智能作为一种方法，探讨其对比较文学研究所带来的裨益与改变；其二是将人工智能视为一种能够产出的“主体”，并不断论证此主体的合法地位^[1]。但不论哪种情况，人工智能仍然以技术手段的形式存在，而非一个完全独立的主体。其后果便是，人工智能的产出往往被视为人类作用的结果，而人工智能在其中所做的不过是对人类产出的加工重组与输出，因此，关于人工智能所创作的文学作品究竟是“我们的文学作品”还是“他们的文学作品”，以及共同创作的文学作品质量如何与能得到何种程度的评价，成为学科研究讨论的重点。

人工智能作为一种新的技术手段，其出现并参与学科发展是必然的，但这绝不是人工智能参与文学发展的唯一方式。比较文学作为一门开放、面向未来的学科，在将讨论内容聚焦于人工智能的主体性判断之外，还应深入思考将人工智能作为一个“完整主体”之后，其不仅是理性工具论意义上的存在，还是情感主体的存在这一问题。在中外文学

史中，情感的流露与迸发都是文学作品的重要起源，人类的复杂情感——爱与恨、愁与怨的交织，自古以来就是文学作品的主要内容。这些复杂情感不仅是在人类内部产生的，更是人类对自然万物的爱意表征。当李白面对庐山瀑布说出“飞流直下三千尺，疑是银河落九天”时，他是否会想到有一天人类会与自己创造出的虚拟对象产生爱意？而这些爱意或许会成为文学的新内容。

一、人工智能与情感

人工智能是否会产生情感？这个问题在今天看来似乎无须多言，答案是否定的。首先，现在的人工智能所做的工作偏重模仿，即学习人类知识、模仿人类行为与模拟人类思维，而情感所必需的自主意识是现阶段人工智能尚未拥有的，人工智能迄今为止还停留在“似乎有意识”^[2]的阶段。其次，情感作为一种身体对外界的反应，其具身性要素是必备的。已有心理学研究表明，人类并非因为情感波动而产生身体反应，而是因身体有所反应才产生情感。这意味着，人并非因为感到羞耻而脸红，而是因为脸红才感到羞耻。因此，情感作为一种身体的反应而存在，并且离不开身体。人工智能在现阶段的发展，更多以语言表达为主。尽管不少机器人已经拥有身体，不论是人形身体还是物形身体，但其更多按照既定程序进

行反应，哪怕将常见人类情感作为一种预设置于其程序内部，其依旧受限于种类。除此以外，情感与认知紧密相连，不仅存在可认知的情感，更存在不可认知的情感，因此，并非所有情感都被人类所认知。人工智能目前所做的只是模仿人类认知，当认知本身都尚且处于更新状态时，其模仿物只能如柏拉图所言的艺术一般远离真理。

因此，在现阶段看来，人工智能因缺乏自主意识和灵活反应的身体，难以拥有情感，但人工智能必将经历由弱人工智能发展至强人工智能的阶段。当强人工智能时代来临，人工智能会拥有独立思考能力与自主意识，或许与之相伴而来的便是人工智能自身拥有的情感。尽管今天的人工智能本身无法拥有强烈且复杂的情感反应，但它仍旧以他者的形式参与着人类的情感交流与实践，并为人类提供情感慰藉，激发人类情感体验，其具体表现形式可能为语言对话、屏幕互动以及虚拟和现实层面的相互感触。

二、人工智能的情感实践场景

现阶段的人工智能虽自身尚未拥有情感，但在人机互动过程中，其仍旧扮演着提供情感支持的角色，激发了人类多样性的感情。不同实践场景下的人工智能发挥着不同的作用。

相较于“情感”，似乎“理性”更加符合人工智能这一技术带给人们的印象。诚然，生成式人工智能作为大数据集合的中心，凭借大量的数据训练来完善自身的知识储备，以此达成输出目的。因此，现阶段诸如 DeepSeek、豆包等人工智能大语言模型应用，用户能够通过指令输入或问题设置得到想要的结果与回答，但在看似理性的数据处理流程背后，AI 似乎在关怀人类情感方面下足了功夫。例如，倘若用户在 DeepSeek 中输入与担心、忧虑等情绪相关的问题，那么 DeepSeek 的深度思考过程中一定会显示出对用户情绪的关怀，并试图引导用户说出更多信息，以获得更好的输出效果。更有甚者，当笔者尝试询问 DeepSeek “人工智能的出现会给比较文学学科带来什么改变”时，其深度思考过程依然展现出对笔者情绪的揣测，即笔者是否未能说出问题背后的担忧，并针对这一不确定内容进行了回答，尽管笔者所提问题并未存在明显的情绪内容。由此可见，看似理性的

人工智能并非单一拥有理性一面，在系统设置与升级过程中，为人类提供情感慰藉始终是人工智能不可绕过的一环。

早在 2013 年，由斯派克·琼斯导演的电影《她》的上映便激发了人类与人工智能恋爱的幻想与激情。电影中的主角西奥多，与最新的人工智能系统萨曼莎展开了一段跨“物种”的恋爱，尽管这种情感并不能为其他人所理解，但不论是西奥多还是他的朋友艾米都沉浸在这种关系中无法自拔。萨曼莎作为不断升级发展的人工智能系统，在与西奥多的交往过程中不断丰富自己的意识与情感，以达成系统上的升级。但不幸的是，西奥多并非萨曼莎唯一的交往对象，而只是其 8316 位交互对象中的一员。萨曼莎虽坚称她深爱着西奥多，但她最终为了实现自己的存在价值而选择离开西奥多。电影最终以西奥多向前妻的再度表白结束，这似乎指向了人类与人工智能恋爱的悲惨结局，仿佛昭示着人类的爱情仍应在人类内部发生。

与《她》不同的是，如今的人工智能尚未发展成为如萨曼莎般拥有自主意识的个体，但与人工智能恋爱却已不是新鲜话题。不同于萨曼莎所拥有的自主意识，今天的用户可以通过具体指令来任意塑造自己的智能恋爱对象，例如输入“你是一个性格稳重、事业成功的男 / 女性”指令等。除此以外，由于人工智能大数据的收集，用户甚至可以将某些具体标签加在指令中，如最近在年轻人群体中流行的星座与人格测试等。或者在指令中直接禁止用户不喜欢的回答模式，以此塑造一个完美的、理想中的人工智能恋人。然而，基于大语言模型塑造的恋爱对象只能满足用户口头语言上的恋爱需求，却无法带来视觉上的震撼。叠纸文化于 2024 年初推出的游戏《恋与深空》，便在市场供应上填补了这一空白。游戏以精细建模塑造了五个风格各异的完美男性角色，以满足不同玩家的恋爱需求。在进入游戏之前，玩家可以根据自己的真实相貌“捏”出游戏中自己的样子，以便在浪漫情节出现时产生更强的代入感。此类种种，都使得人工智能成为一个满载情感寄托的主体，它以用户心中最理想的模式承载用户的情感，给用户带来全新的情感体验。

当人工智能与虚拟实在联系在一起，或许有可能为人类提供一个可抵达的乌托邦。翟振明在《有

无之间：虚拟实在的哲学探险》中提出将虚拟实在看作与我们所处的自然世界相平行的世界存在，并认为虚拟实在与自然世界之间的转换是两个平行世界间的经验性转换，而不是虚幻与真实之间的本体性转换^[3]。换言之，虚拟实在只是一种在未来发展中可选择的感知框架，与我们在自然世界中用眼睛、皮肤等感知器官去感受是一样的，它们具有同等的本体论地位。在此种本体论前提下，人类可以自由选择自己想要生活的世界。

在“流动性转向”中，“地方”与“位置”的概念得到了区分，相较于后者的情感真空，前者“地方”更富情感色彩，体现了人们的情感认同与依恋^[4]。在虚拟实在这个可以被人们自主选择的世界中，人类不仅可以以类似于“上帝”的身份建构起独特的虚拟环境，以满足对回不去的“地方”的期待，更可以在赛博空间中建立人与人之间的联系，甚至在这个过程中获得爱情并结婚生子。而虚拟实在配套的感知装备可以给用户带来模拟的触感体验，在某种程度上进一步加强人机互动的“实感”，达成人类在亲密关系中想要触摸的欲望。在此意义上，虚拟实在与人工智能的结合将创造出更丰富、更深刻的情感实践场景。

三、人工智能时代下的文学情感可能

情感自古以来都是文学作品的重要主题，也是文学创作的源泉之一。人工智能在当今阶段虽尚未拥有自主情感，但其凭借对人类思维和认知的模仿，已然可以在多种场景下为用户提供情感支持，成为人类情感的投射对象。此阶段，讨论文学创作中人工智能的情感为时尚早，因为输入与输出间“黑箱子”内究竟是什么仍难以得知，不过，此时的人工智能已经参与了人类情感生成过程，并将更加深入且持续地产生影响，因此，探讨人工智能激发的文学情感具有一定的意义。

首先是生成式人工智能为代表的大模型应用。越来越多的用户正在将此类应用当作具体对象与之对话。例如，将AI设定为善解人意的心理医生，与之沟通如何处理生活中遇到的难解的人际矛盾等。除技术水平的差别外，人工智能扮演的心理医生与真实医生最大的不同之处在于其耐心与私密性。只要打开类似的网站，便会有这样一位倾听者随时为用户答疑解惑，提供情感支持，并且

无须担心隐私泄露等问题。人工智能提供了一个方便且私密的环境，人类的情感或烦恼有了更多的出口。

其次是提供恋爱快感的人工智能。正如前文所述，人工智能提供了一个“完美”虚拟恋爱对象的可能，且在恋爱过程中持续接受用户的“调试”，满足用户对爱情的幻想。但与人工智能的恋爱体验注定和真实恋爱存在差距，其中之一就在于无法真实地感触到对方，这点在电影《她》中已经初见端倪。但正是这一无法弥补的遗憾，成为《她》中引人思考并迸发出无限张力的地方。相比美满与幸福，遗憾似乎能带来更多灵感的乍现与经典作品的形成。

的确，人工智能的存在使得遗憾的种类增加了，但尽管如此，此类文艺作品并未呈减少趋势。由此可见，人类对这样的情感需求依然存在。当虚拟实在与人工智能技术结合时，人类便扩大了身临其境的感知空间，有了获得更大情感空间的可能性。如前所述，虚拟实在技术能使人拥有创造与感受的能力，当消失已久的地方重现眼前、当日思夜想之人站在面前，失而复得的复杂情绪或许会成为文学创作的新源泉，同时带来无限的可能性。人工智能的出现给人类带来了全新的情感体验，或深或浅，或明或暗，这些丰富体验都是人类对自我的再度投射。尽管人工智能有诸多机会参与人类的情感生活，并带来新奇与丰富的情感体验，但这之中也确实存在诸如深度连接不够等问题。随着人工智能的逐渐发展，或许在不久的将来，人类可以与人工智能建立更密切的联系。在这一天到来之际，这些连接程度不一的内容与自古以来已经存在的情感类别会一起构成新的情感，成为激发文学灵感的新内容。

参考文献：

- [1] 陈众议, 聂珍钊, 金莉, 等. 人工智能时代的人文学科建设：挑战与机遇 [J]. 上海交通大学学报(哲学社会科学版), 2024, 32(12): 97-112.
- [2] 金雯.“似乎有意识”：理解人机互动的限度与潜能 [J]. 传媒观察, 2025(4): 28-36.
- [3] 翟振明. 有无之间：虚拟实在的哲学探险 [M]. 北京：北京大学出版社, 2007: 2-3.
- [4] 刘英. 流动性研究：文学空间研究的新方向 [J]. 外国文学研究, 2020, 42(2): 26-38.

人工智能与文学新变

黄巧心

深度求索（DeepSeek）、文心一言（ERNIE Bot）等基于大语言模型的AI工具横空出世，为文学创作和文学研究提供了全新的范式。在新文科背景下，AI的机器学习能力与人类的主观能动性，不应是“你死我活”的对立关系，二者的关系应该建立在以人为本的共识之上。借用跨学科研究这一融通的中介，观照二者的交互作用，从历史和逻辑中去探索二者你中有我、我中有你的和谐共生关系，对于推动数智时代文学研究的发展具有重要的意义。

一、人工智能：作为经验主义的科技成果

无论是外国的ChatGPT还是中国的DeepSeek，这一类生成式人工智能（generative artificial intelligence）都具备语言学习能力，一言一行都比拟人类。实际上，其底层逻辑无外乎“反向传播”（backpropagation），即一种模拟人类神经网络运行的计算机算法。1986年，杰弗里·辛顿（Geoffrey Hinton）与同事发表其奠基之作^[1]，并陆续公开该算法的研究进展。该算法一经推广就受到发现DNA结构的诺贝尔获奖者之一的弗朗西斯·克里克（Francis Crick）的驳斥，他说：“近期，某些神经网络计算机算法显现出卓越特性，似乎为理解大脑的计算特性提供了全新途径。遗憾的是，这些神经网络在诸多关键方面都缺乏现实基础”^[2]，而且迄今为止仍无法与人类真实的神经网络进行互证。不过，在当下，其在电脑上的可行性使之支撑着如今几乎所有的人工神经网络，应用到计算机视觉系统、大语言模型等AI领域。所谓“反向传播”，质言之，也即“经验-教训”。如需在任意图像中识别出“鸟”，首先将数千个合成神经元按层排列。图片被输入网络的第一层，该层通过相当于人工智能的突触连接将有关像素内容的信息发送到下一层。此时，神经元可以使用这些信息来挑选线条或边缘。然后，再将信号发送到下一层，下一层可能会挑选出眼睛或脚。这个过程一直持续，直到信号到达负责正确判断

大问题的最终层：“鸟”或“不是鸟”。在这个学习过程中，反向误差传播算法，即“反向传播”在反复运行。如果向网络输入一张鸟的图像，但系统误认为它不是鸟。那么一旦它意识到了错误，它就会生成一个错误信号。这个错误信号在网络中逐层向后移动，加强或削弱每个连接，以尽量减少未来的错误。如果再次向模型显示类似的图像，调整后的连接将导致模型正确地宣布：“鸟”。如此，正反馈加强连接，负反馈削弱连接，以此“调教”模型。当投入的训练数据越多，其识别能力越强。这一计算机实现的“经验主义”成果首先用于图像处理，并广泛应用于医学领域，如CT、核磁等医学影像辅助判断。

对于计算机而言，图像与语言不无相似之处。语言学之父索绪尔发现，语言符号系统不仅是一种分层装置^{[3][28]}，而且兼具组合关系和聚合关系。符号和符号的组合形成语言的结构，而每一个组合都相当于图像的像素。在语言链条的某一环节上能够相互替换的符号具有某种相同的作用，这些符号聚合成群^{[3][31]}，因此具有聚合关系的符号相当于图像中的各个层次。如此一来，反向传播同样也被推广至语言领域。再者，数字人文、数字媒体在30年间积攒了海量的实验数据材料，为ChatGPT和DeepSeek一类应用大语言模型应用提供了深厚的基础。值得注意的是，AI的应用关键在于“经验”，大语言模型（large language model）则关乎语言材料。生长于美国的ChatGPT说得一口英语，而土生土长的DeepSeek则自然而然精通中文：DeepSeek LLM 67B Chat在中文表现上超越GPT-3.5的结果表明，DeepSeek等中国智造产品更符合国人的“体质”。

二、人工智能：作为现代逻辑的应用成果

人工智能如同电脑中的人工神经网络，它与人脑一样有不同的分区，各个分区负责不同的作用。纵览全局，反向传播是针对数据的底层逻辑，属于数据库处理模块，它服从于统筹规划各个模块

的“总司令”，也即中央推理系统。正如人脑在不停处理人与世界的关系，人工智能也在模仿以语言反映人与世界的关系。不过，人工智能并不具有真切的人生经历，所以它的认识论是以 AI 概念为基础的。

AI 概念的提出比 AI 应用的实现要早得多。早在十七、十八世纪就已出现 AI 的思想先驱，比如笛卡尔、莱布尼茨、霍布斯、休谟、康德等。其源头可上溯至古希腊，古希腊为 AI 提供了至少两个思想资源：第一个是德谟克利特的机械唯物主义；第二个是古希腊人的形式主义传统——通过形式刻画来澄清自然语言处理的歧义。在这个问题上作出贡献的哲学家，主要有毕达哥拉斯、苏格拉底、柏拉图和亚里士多德^{[4]77}。

与之相比，维特根斯坦更加贴近当下的 AI 时代。基于 AI 科学视域，可以发现维特根斯坦的两本著作都涉及 AI 的重要话题。其一是《逻辑哲学论》，这本维特根斯坦的成名作讲述了世界 - 事实 - 思想 - 命题 - 真值 - 真值函数 - 一般形式这七者的关系，亦即“世界 - 语言”这一链条的构成。在徐英瑾看来，《逻辑哲学论》牵涉到三个话题：对于世界的形而上学构建；对于这个形而上学世界的语言表征；对于那些不可被表征事项的“沉默权”^{[4]123}。其二是《哲学研究》，这是维特根斯坦思想的集大成之作，也是他对《逻辑哲学论》进行扬弃的产物。此书所讨论的是智能体 (agent) 在怎样的规范性条件约束下，在历时的动态环境中，如何利用相关的表征工具（特别是日常语言），完成和环境以及其他智能体之间的因果互动，并最终完成某些任务。从 AI 角度看，《哲学研究》超越《逻辑哲学论》的最大地方，就在于它不再把静态的知识体系规整视为哲学理论的聚焦点，而是把焦点转移到了智能体的行动，转移到了对于信息的实时处理上。AI 科学的“黑话”转译了这两本书的哲学体系，而逻辑学话语则揭露了它们的本质。《逻辑哲学论》体现的是经典逻辑，而《哲学研究》表征了现代逻辑。亚里士多德的经典逻辑显然已经不再适用时代的发展，因而被维特根斯坦舍弃，现代逻辑才是符号 AI 的核心技术基础。经典逻辑可简单地理解为：要证明 A 是 A，只要证明非 A 即可。现代逻辑则更加直接：只需举出 A 之所以为 A 的证据。显然，现代逻辑更为“用

户友好”。这两个逻辑学的概念并非真的如此简单，其要在计算机中通过定义、判断、循环等编程手段得以实现恐怕更是不易。概言之，维特根斯坦用现代逻辑理顺了语言与世界的关系，从而解决了 AI 实践中的某些悖论。

三、文学：AI 是敌是友？

综上，当下广受大众喜爱的 AI 实际上是文学与科技跨学科研究实践的成果，是一个“混血儿”。科技本质上是一把“双刃剑”，人工智能越向披靡，就越值得人们警醒。经验论并非认识世界的不二法门，现代逻辑也非完美无缺，它们的内在特性——当然也包括其不足——自然而然地嵌入这一造物之中。无独有偶，经验主义和现代逻辑一致指向具象性，人工智能亦表现出“数据为王”“算法为王”的倾向。人工智能应用生成之初如约翰·洛克 (John Locke) 所言是一张“白板”，它后来所吐露的语言文字实质是对人工喂养的数据进行人为选择的结果。既然有人为的干预，那必然容易产生偏见。这种偏见来源于数据和算法，并共同作用于同质化 (homogenization)，或马尔库赛所谓的“单向度性” (one-dimensionality)。如果任由这一趋势发展，将对关注异质性的文学研究造成不堪设想的后果。

数据偏见已然显现。虽然 AI 应用训练数据来源仍未公开透明，但流传在网络空间的海量数据往往最受 AI 应用开发者青睐。在这一片新兴的公共领域（尽管在一定程度上接受监管），各方言论和信息良莠不齐，却又汇成公众舆论^[5]。这一公共领域还自带网络这一“形式”所带来的优势，其具有传播速度快、助长公众舆论单一化的特点。雪上加霜的是，AI 技术能够帮助错误信息和虚假信息以迅雷不及掩耳之势蔓延，令公众舆论的真实性大打折扣。于此，某些与主流公众舆论不合的意见、一些真实可靠且知之甚少的信息则被拦截过滤，而不出现在这个虚拟公共领域的声音自然更是“名存实亡”，在场的信息数据垄断公众舆论，数据偏见油然而生。除了这一类“人云亦云”的数据偏见外，另外一种数据偏见就是其与权力勾结在一起。首当其冲的便是流媒体催生的大批新兴内容制造者，如博主、网红等，他们为训练模型数据创造养分并同时利用 AI 应用生成信

息数据对其反哺，无论自愿与否，其最终都沦为广告主的媒介，为商业权力或政治权力站台。另外，AI应用开发者在建构AI算法和选择数据训练模型时，充分发挥自以为绝对客观的个人主体性（其实时刻受到潜在的意识形态的影响），所以个人主观偏见肯定会带入AI应用当中。进一步而言，AI应用开发者背后往往牵连着更广泛的权力。这些权力主体凌驾其上，具有更强有力的控制权，它通过制造、伪造、篡改等数据操纵手段，扰乱AI应用的输出成果，甚至以非科技的手段对其进行直接干预，使某些意识形态在AI应用中显化。与数据偏见类似，算法偏见根植于技术研发者及其背后的权力，而其权力总是掌握在极少数的人（如AI技术的顶层设计者）手里。数据偏见和算法偏见在一定程度上推动着个人和社会走向单向度的思维绝境，并渐渐枯竭文化土壤，萎缩文学有机体，中断文学研究的养分供给。

AI技术的发展让生活在信息时代的我们尝到了“甜头”。以文学研究为例，AI支持的大语言模型应用助力学术研究，帮助研究者搜集信息、整理信息，扫除知识盲区、扩充知识储备。虽然“一千个读者有一千个哈姆雷特”，但大语言模型应用可以辅助研究者进行基础性文本分析。前人的研究经验更是大语言模型应用的“特长”，它可以帮助学者更快更全地归纳前人的研究成果，搭建文学作品的特征联系网，大大节省查找、筛选、归整资料的时间，使得研究者调配时间到自身的研究对象和研究问题本身。另外，AI应用往往掌握多种语言，使语言屏障变薄，有助于学者掌握世界文学和理论发展的前沿动态。概而言之，AI应用对于重视前人研究经验的学术性科研创新，具有极大的辅佐作用。此外，还有不少用户发布指令，让AI进行文学创作。这一类的“创作”往

往一经发表，便能震惊世人。首先，其言语流畅，逻辑通顺。文体各有侧重，若是应用文、说明文等以传达信息为主的文体，满足以上两点即可。若是诗歌、散文、小说等文学体裁，以上不过是门槛，尚可娱乐消遣。若再论言情、写意、载道，则因其自百家而来，泛泛而谈，无所负载，无所寄托，其作品终究落得平平无奇。究其原因，大约是AI创作来源于网络世界，而网络世界又与现实世界有着不可消除且不容忽视的异质性。从创作本质而言，AI生成内容与经典模仿理论存在本质差异。为弥补上述缺陷，一种新型写作范型应运而生，即“人机合谋”的AI作者。这类AI创作或“半”AI作品提供了以科学技术为文学创作主体的研究样本。这些作品是否因袭模拟尚待考究，其文学性以及文学价值或仍需由时间来见证。不容置疑的是，它们实属新型的文学类型，是一种新的文学样式，丰富了文学的内涵。

综上所述，无论是文学研究还是文学创作，人工智能能在一定程度上协助打通语言障碍、学习前人研究成果、提供新的文学素材，AI作者有望成为新的跨学科研究范型。数智时代，文学与人工智能的关系会变得越来越互通、互鉴、互为。

参考文献：

- [1] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning Representations by Back-Propagating Errors[J]. Nature, 1986, 323(6088): 533–536.
- [2] FRANCIS C. The Recent Excitement About Neural Networks[J]. Nature, 1989(337): 129.
- [3] 叶蜚声, 徐通锵. 语言学纲要 [M]. 修订版. 北京: 北京大学出版社, 2012.
- [4] 徐英瑾. 心智、语言和机器: 维特根斯坦哲学和人工智能科学的对话 [M]. 北京: 人民出版社, 2013.
- [5] 哈贝马斯. 公共领域的结构转型 [M]. 曹卫东, 王晓珏, 刘北城, 等译. 上海: 学林出版社, 1999: 225.

生成式人工智能的类文学意识

向丰田

2025年伊始，DeepSeek深度思考模型正式推出，引发广泛关注。生成式人工智能技术的快速发展，深刻影响了文学创作，AI的文本生成能力

挑战了以人类为中心的文学创作模式，同时激起了诸多学者对AI主体性及“类文学意识”的思考。“生成式人工智能指通过人工智能相关技术，

自动化生成文本、图像、视频、音频等多类型内容。”^[1]作为生成式AI的核心技术，Transformer架构及其驱动下的大语言模型，通过对海量数据的解析与训练，能够响应用户指令，生成包括文本、图像及视频在内的多种内容。生成式人工智能凭借深度学习模型，捕捉大数据库中的潜在规律，模仿“人类创造智能”生成内容。作为一种新质生产力，生成式AI广泛参与社会经济生产活动和意识形态生产活动。“人工智能作为新的技术变革，会改变意识形态与生产方式的关系，出现新的意识形态。”^[2]不同于传统的分析型AI，生成式AI的生产活动并非基于预设规则，而是基于概率和模型的“创造”。这种“创造”在意识形态生产活动中不仅可以提高生产效率，还能够催生新的意识形态。“文学的普遍属性在于，它是一般意识形态的形式。”^[3]作为审美意识形态，文学的“社会意识”根植于文字符号的出现、人类长期的社会生活经验、内在情感的表达需要以及文学创作形式的探索等“社会存在”之上。人类创作、欣赏、发展文学，形成了各种各样的文学认识和理论，这些都不是一蹴而就的。机器不是基于个体的生活情感体验，而是通过大数据模型和算法分析具备了一定的文学创作能力。虽然是对人类语言习惯的模仿，但生成式AI所创作的文学作品也具有一定的独创性，表现了生成式AI对文学的独特理解，这也是其拥有类似于人类的文学意识的体现。

作为不断进步的科技文化，生成式AI进行的文学创作蕴含着极大的不确定性和冲击力，将会持续颠覆并刷新人们关于科技和文学的认知和想象。AI带来的科技威胁在人机交互的模式中若隐若现，侵蚀着人与机器之间的距离。生成式AI对文学的强势介入，让学术界开始重新探讨“作者已死”的论题。人机交互现象的出现，使文学创作者的身份开始模糊。关于人工智能主体性的问题，不少学者提出了各自的看法与见解。

人机合作和AI创作现象的出现，说明文学创作不再是纯粹的人类活动。生成式AI慢慢地衍生出接近于人类的“类文学意识”。不同于人类文学的创作方式，生成式人工智能基于数据大模型，形成了一套独特的文学生成逻辑。“人工智能文学的生成机制完全依赖于计算逻辑。人工智能文学基于算法的重复性操作而产生，但是在这种重

复中，提示词的随机性、数据集的多样性以及机器学习的动态性共同作用，又使之产生了丰富而独特的差异。”^[4]AI文学的生成思路是通过算法深度学习人类文学文本中的隐含规律和内部结构，并基于算法模型中数据与指令的滑动、变换等来输出文本。虽然人工智能的学习过程是反复采集和抓取人类文学文本中的有效信息并以此汇聚成大数据库，但其不是进行简单的抄袭或一成不变的机械模仿，而是不断地进行自我更迭和创新。

人工智能的演进始于对人类既有知识的学习，这一过程通过基于大规模数据集的算法预训练完成，通过交互产生的新知识使其能力超越原有的范畴。关键之处在于，无监督的学习机制使其摆脱了对人类输入数据的依赖，转而能够自动地安排学习进程，这显现出AI已从以前的被动学习转向主动学习。AI通过学习现有知识获得新知识，而且经过数以万次的自主学习，持续地强化自己的算法神经网络。在不断的输入、反复的自我学习、加工、处理的过程中，生成式AI逐步形成了一种“类文学意识”，即类似于人类的思维加工和信息处理的意识，其包含构思文学文本、使用文学手法、理解艺术效果、进行文学创作等。由算法网络编织形成的“类文学意识”是生成式AI可以像人类那样创作文学作品的核心所在。这种“类文学意识”使得生成式AI在缺乏创作主体应有的理性知识、生活经验与感性认识的情况下，仍然能够依靠自身的大数据库和计算逻辑生成文学作品。如今的生成式AI在极大程度上具备了创作自由的可能性。生成式AI提供的先进生产力，使得文学创作的门槛愈发降低。即使不具备任何旧体诗格律知识的普通人，也可以在短时间内凭借DeepSeek生成一首合乎格律的诗歌。从诗歌与短篇小说的摸索，再到长篇小说的尝试，AI的文学创作将大步向前。随着算法与技术的突飞猛进，人工智能可能会以一种前所未有的、难以想象的姿态出现在人们的视野当中。

不过，人工智能的“类文学意识”也具有先天的局限性。“人工智能之父”图灵在《计算机器与智能》中谈到：“除非机器能够做到因为有思想、懂感情而不是通过符号的偶然来临去写十四行诗或创作协奏曲，否则我们不能认为机器与大脑是等同的——也就是说，不仅把它写出来，而且知

道已经把它写出来了。”^[5]图灵一针见血地指出了AI文学创作的先天缺陷。的确，人工智能没有生活经历，未曾作为一个鲜活的个体感知过世界、自然与社会，因此无法体验各种细腻的情绪情感，更不能意识到自己是否正在创作。因此，基于算法逻辑的人工智能，仅停留在功能性与技术层面模仿人类，其行动逻辑并未根植于人类主体性的核心要素——自觉的能动性、自我反思意识及自主意志。二者的行动原理有着本质区别。

在文学创作过程中，人工智能并非基于自身知识经验与创作冲动，而是基于大数据与深度学习模型等技术以及人类的指令进行创作。人工智能的文学意识依赖于人的主体性和算法数据库，而非自身的创作需要，这与人类的文学意识存在本质上的差异，其只能被称为“类文学意识”。

概言之，AI时代的到来，使得文学创作者、

研究者需要严肃认真地思考诸如AI文学的归属、AI文学创作者的定义、AI文学与人类文学的关系等等新的研究课题。

参考文献：

- [1] 卢宇,余京蕾,陈鹏鹤,等.生成式人工智能的教育应用与展望:以ChatGPT系统为例[J].中国远程教育,2023(4): 24.
- [2] 刘伟兵.人工智能会生成意识形态吗? [J].山东大学学报(哲学社会科学版), 2025(3): 175.
- [3] 童庆炳,李衍柱,曲本陆,等.文学理论教程[M].5版.北京:高等教育出版社, 2015: 64.
- [4] 刘家夷.计算逻辑:人工智能文学的异质性生成 [J].江西社会科学, 2025, 45(1): 93.
- [5] A M 图灵.计算机器与智能 [M]// 玛格丽特·博登.人工智能哲学 [M].刘西瑞,王汉琦,译.上海:上海译文出版社, 2001: 72-73.

接受美学视角下生成式人工智能“机器幻觉”的生成机制

雷秋香

在新文科建设不断深化的当代学术图景中，文学研究与人工智能的跨学科交融正逐渐展现出其理论潜力与现实紧迫性。人工智能是新时代模拟、综合、拓展人类行为和人类智慧的综合性技术学科，其产生于20世纪50年代，是计算机科学、逻辑学、生物学、心理学和哲学等众多学科理论和技术的集大成者^[1]。“机器幻觉”是指人工智能生成的内容可能会偏离真实世界的事实或不能准确反映用户指令，并由此产生幻觉风险。机器幻觉主要分为两类：事实性幻觉和忠实行幻觉^{[2][48]}。机器幻觉不仅是一个技术问题，更是一个触及意义生产、文本接受与文化表征的核心人文议题。在新文科倡导跨学科整合与方法论创新的背景下，从接受美学的视角探究机器幻觉的生成机制，不仅拓展了接受美学的解释边界，也为我们重新审视文学本质、作者身份与读者角色提供了新的思考空间。AI文本的接受机制可被理解为一种“算法化的阐释循环”：用户的提示词构建出初级的“期待视野”，而模型的输出又反过来重塑甚至固化

用户的认知框架。这一过程虽然与姚斯所提出的“视域融合”形成了表面呼应，但其内在动力存在着根本的差异。传统文学视野的嬗变依赖于审美创新与历史经验的双向互动，而AI的“视野”调整则受到工程师对训练数据的干预与技术路径的约束，从而使得技术权力无声地渗透至意义生成的微观层面。

在新文科所关注的全球知识生产语境中，机器幻觉常常再现并强化训练数据中隐含的文化霸权。多模态大模型的不断发展更使机器幻觉的形式趋于复杂，其影响已超越技术层面，深入触及数字时代的知识建构、文化传播与认知伦理等根本议题，要求人文社会学者从伦理、美学与文化政治等维度进行深度介入。生成式人工智能从准备、运算到生成内容，涉及数据、算法和算力三个基本要素，在数据训练、模型运算、内容输出、内容利用等各个环节上存在风险，需要进行识别和治理^{[2][48]}。

接受美学理论，尤其是伊瑟尔的“召唤结构”

与姚斯的“期待视野”，为解读“机器幻觉”的生成机制提供了新的理论框架。本文旨在揭示AI如何通过算法模拟文学接受中的“空白”，诱导读者参与意义填补，同时又因技术本质与人类审美的根本差异，导致意义生成的扭曲与压缩，最终催生“机器幻觉”。

一、“空白”的算法化：召唤结构的机械实现

沃尔夫冈·伊瑟尔在《文本的召唤结构》讲演稿中提到，“作品的未定性和意义空白促使读者去寻找作品的意义，从而赋予他参与作品意义构成的权利”^[3]。换言之，具有“召唤结构”的文学文本会召唤出个体的创造欲，激发读者具体化文学作品中的不确定性和意义留白。正如鲁迅在《孔乙己》中并未直接交代人物的最终命运，仅以“坐着用这手慢慢走去了”^[4]作结，但正是这种叙事留白赋予读者阐释孔乙己命运的自由与审美想象的空间。

然而，生成式人工智能实践的是一种算法化、概率化的“空白”填补机制。用户的提示词成为触发器，将模糊的意图转化为可检索的标签，而模型内部的注意力机制则优先聚焦于局部语义关联，依照统计概率生成“最可能”的后续内容，为此，其常常牺牲文本整体的逻辑性与人文一致性。一方面，AI作为“读者”难以捕捉用户提问中的隐含意图与情感基调，仅能机械拼凑高频词进行响应，导致生成文本文学性贫乏、情节突兀甚至虚构学术引用；另一方面，AI作为“作者”，虽与用户之间存在知识结构、文化背景与认知意图的严重不对称，却必须将开放性的、不确定的人类发问转化为确定性输出，这就不可避免地会导致“机器幻觉”。这种算法对空白的机械实现，非但不能复制人类阅读中的意义协商过程，反而反映出技术在模仿人文理解时难以逾越的根本界限，即无法真正响应“召唤”，而只能执行运算。

二、“期待视野”的数据固化：认知惯性的技术强化

姚斯在《文学史作为向文学理论的挑战》一文中提出了“期待视野”的概念，他说：读者“从类型的先在理解、从已经熟识作品的形式与主

题、从诗歌语言和实践语言的对立中产生了期待系统”^[5]，这一过程本质上是历史的、开放的、具有个体差异的。然而，生成式人工智能通过对训练数据库内的海量文本数据进行概率统计与模式提取，将原本属于人类审美经验中的“期待”转化为可计算、可预测的数据关系。它并非真正“理解”文本的情感或文化内涵，而是通过词语序列之间的概率匹配模拟出符合用户既往偏好或常见表达模式的文本内容。这一机制在表面上似乎迎合了用户的审美习惯，实则可能导致认知惯性的技术性强化。算法不断回应用户已有的兴趣结构，将原本多元、流动、可能被挑战或拓展的“期待视野”固化为一套被数据定义的反馈循环。由于AI缺乏具身化的情感体验与文化实践，其生成的文本往往停留于抽象语言符号的表层组合，无法承载人类读者在现实世界中通过身体、情感与社会互动所构建的深层意义。这也正是所谓“机器幻觉”的根源之一——语言形式与经验实质的脱离。因此，生成式AI虽看似拓展了文本生产的边界，实则可能正在将接受美学中最为珍贵的“视野融合”与“意义协商”机制窄化为一种数据驱动的自动化投射，从而潜在地削弱文学接受中应有的批判性、多样性和人文精神。

三、接受美学的新范式：“人机共生”

传统接受美学的理论基础建立在文本的人类作者性之上，即便强调读者对“空白”的填补，其诠释行为仍围绕着作者的意图框架展开。然而，人工智能生成文本从根本上改变了这一格局，文本的源起不再是一个有意识的创作主体，而是一个基于概率运算的无意识算法系统。在这一过程中，读者身份发生了根本性转变，其从被动的意义接收者、阐释者跃升为主动的“意义共同创造者”。值得注意的是，本雅明所慨叹的机械复制时代传统艺术中消逝的“灵韵”，在AI生成艺术中以“数字灵韵”的形式重新回归了。这种灵韵不再依赖于原作在特定时空中的在场性与本真性，而恰恰诞生于算法的非人特质之中。生成式人工智能输出的不可预测性、语义跳跃、意象拼贴与逻辑裂隙，在传统尺度下被视为技术缺陷，却在人机共生的审美情境中转化为新型艺术表现力的来源。以“九歌”和“小冰”为例，它们常常因

概率采样机制产生非常规的词语组合与意象衔接，形成强烈的陌生化效果。这些“失误”反而促使读者跳出对作者意图的追寻，转而进入与算法系统的创造性对话之中。因此，人工智能生成文本拓展了新的接受美学范式，推动我们重新思考创作、文本与接受三者之间的关系。

四、文化抵抗的策略：对抗性提示词

在生成式人工智能日益渗透文化生产的当下，一种新型的批判性实践正在浮现，即用户通过策略性地设计“对抗性提示词”，将技术界面转化为文化抵抗的场域。这一行为已远远突破传统接受美学中读者对文本“空白”的被动填补范畴，意味着用户角色从意义的“消费者”转变为意义的“干预者”乃至“共构者”。正如福柯所指出的，

“在每个社会，话语的制造是同时受一定数量程序的控制、选择、组织和重新分配的，这些程序的作用在于消除话语的力量和危险，控制其偶发事件，避开其沉重而可怕的物质性”^[6]。生成式AI正是通过其训练数据中隐含的性别、种族与文化偏见，被算法以“客观”的概率计算再生产出来，从而悄然强化了主流意识形态。例如，用户通过刻意要求模型生成“女性程序员”或“非西方中心的叙事结构”展开了一场自觉的“话语斗争”。他们不仅批判性地揭示出算法中被遮蔽的政治性，更主动迫使系统偏离其固有的生成路径，在异质性的符号间隙中开辟出反霸权的表达可能。这一过程呼应了阿多诺对“否定的辩证法”的呼吁，不在技术秩序之外寻求乌托邦，而在系统内部通过“异质元素”的植入，激发自我批判与更新的潜力。

生成式人工智能的“机器幻觉”现象，既是技术局限的表征，也是人类认知模式与文化权力结构的数字镜像。本文通过接受美学的理论透镜，揭示了AI生成文本中幻觉的深层机制：算法通过

统计模拟文学“空白”与“召唤结构”，以数据驱动的“期待视野”诱导读者填补意义，但这种互动本质上是技术逻辑对审美经验的殖民化。

这一发现对人文与技术的双重领域具有深刻启示。在理论层面，接受美学的范式亟须革新——

“机器读者”的提出挑战了人类中心主义的阐释传统，而“人机协同阐释”框架的构建，则为数字时代的文学批评开辟了新的可能。在技术实践中，AI开发必须超越对流畅性与效率的单一追求，转而拥抱文化敏感性与伦理自觉。

最终，生成式AI的伦理困境与文化危机，指向一个根本命题：在算法统治的文本世界中，人类如何守护意义生成的自由？答案或许在于重构技术与人文的共生关系——不是将AI视为威胁，而是将其作为照见自身认知局限的镜子；不是放任算法霸权，而是在数据中铭刻多样性，在交互中保留抵抗的缝隙。唯有如此，我们才能在技术的狂飙中，让文学的灵魂免于沦为概率的囚徒，让文化的肌理挣脱符号的扁平化，让每一次意义的重建都饱含生命的温度与历史的重量。

参考文献：

- [1] 陈露阳.人工智能的兴起与生态文明的发展[J].湖南工业大学学报（社会科学版），2025，30(3): 43.
- [2] 王芳，朱学坤，刘清民，等.生成式人工智能研究进展[J].图书与情报，2024(4).
- [3] ISER W. *The Act of Reading: A Theory of Aesthetic Response*[M]. Baltimore: Johns Hopkins University Press, 1978: 11.
- [4] 鲁迅.孔乙己[M]//鲁迅.鲁迅选集：第1卷.北京：中国青年出版社，1957: 18.
- [5] H R 姚斯，R C 霍拉勃.接受美学与接受理论[M].周宁，金元浦，译.沈阳：辽宁人民出版社，1987: 28.
- [6] 许宝强，袁伟.语言与翻译的政治 [M].北京：中央编译出版社，2000: 3.

责任编辑：黄声波