

人工智能与文学批评

探寻人工智能文学批评范式

——以 AI 生成小说《一名作家之死》为例

冯洋

(外交学院 英语系, 北京 100037)

摘要:人工智能文学正在重塑传统文化创作与批评的边界,并重新激活作者、文本与伦理等文学研究中的核心议题。AI生成小说《一名作家之死》戏仿文学理论中的经典命题“作家之死”,蕴含了三重嵌套叙事结构,每层叙事分别呈现不同的作者身份。一方面,这些作者在各自的叙事层面上直面技术所带来的多重伦理困境,以一种有机和运动的方式揭示了人工智能如何为传统文学批评注入新的内涵与复杂性。另一方面,人工智能文学并未与传统文学形成不可通约的范式革命,而是通过混杂、流动与开放的视角拓展了文学的可能性,也为文学批评提供了新的思考路径。

关键词:人工智能文学;作者身份;叙事结构;伦理困境;《一名作家之死》

中图分类号:I0-03 **文献标志码:**A **文章编号:**1674-6414(2026)05-0109-12

DOI:10.26922/j.cnki.waiguoyuwen.2026.05.008

0 引言

人工智能写作经由半个世纪的发展,经历了由规则驱动向数据驱动的演变,并逐渐催生出“人工智能文学”这一新的文学形态。从亚里士多德“四因说”来看,人工智能文学正在重组传统创作的构成要素。其“质料”由生命经验转向数据资源,“形式”由主体表达转向概率生成,“动力”由生存体验转向算法优化,“目的”也由“诗意栖居”转向高效生产与复制。这种创作机制的变化不仅改变了文学的生产方式,也重新激活了作者、文本与伦理等文学研究中的经典问题。因此,如何在技术语境中重新组织文学批评资源,建构适用于人工智能文学的解释框架,已成为当前人工智能文学研究的重要课题。

当前国内外学界围绕人工智能文学已经展开了较为充分的讨论,主要形成三种观点。第一种观点强调延续性,认为人工智能文学强化了后现代理论中关于作者消隐及反逻各斯

基金项目:教育部人文社会科学研究青年基金项目“‘绝境’中的希望:德里达解构思想研究”(20YJC752005)、外交学院

中央高校基本科研业务费专项资金项目“后疫情时代德里达免疫解构学探究”(3162021ZYE02)的阶段性成果

作者简介:冯洋,女,外交学院英语系讲师,主要从事西方文学理论研究、科幻文学及人工智能文学研究。

引用格式:冯洋.探寻人工智能文学批评范式——以 AI 生成小说《一名作家之死》为例[J].外国语文,2026(5):109

-120.



本刊微信公众号

中心主义的讨论(李国成,2024;Gunkel,2025;Hayles,2023)。第二种观点强调断裂性,认为大语言模型(LLMs)如同“随机鹦鹉”(stochastic parrots),无法真正“赋予意义”或“表达意图”(Bender et al.,2021)。据此主张传统诗学需转型为“高科技诗学”(姜玉琴,2026),现有文学理论也需要被科学化理论替代(聂珍钊,2024b)。第三种观点则强调协作性,主张在承认技术变革的前提下,通过人机协作与跨学科方法更新文学批评路径(Bassett,2025;但汉松,2026;尚必武,2025;殷杰,2024;周宪,2025;曾军,2023)。这三种观点展现了人文学者在面对科技挑战时的三种应对方式:延续传统、全盘改革以及调适融合。本文在一定程度上吸收了第一种关于文学连续性的理解,也认同第三种对于方法调适的强调,但更关注人工智能文学如何使作者、文本与伦理等经典文学问题重现。人工智能文学批评的任务或许并非建立一种完全替代既有文学研究的新范式,而是在新的技术条件下重新激活文学理论的解释潜能。

目前,人工智能文学研究多集中于宏观层面的理论探讨,针对具体文本的批评实践仍相对有限。本文尝试以人工智能生成小说《一名作家之死》(*Death of an Author*)为例,通过文本细读探究人工智能文学的批评路径。这部小说由加拿大作家史蒂芬·马尔什(Stephen Marche)于2023年策划出版,但其署名作者为“Aidan Marchine”,其中,“Marchine”一名由“Marche”与“machine”拼合而成,暗示了人机协同语境下的复合作者身份。创作过程中,马尔什借助ChatGPT、Sudowrite与Cohere三种AI工具协同完成,其中约95%的内容由AI完成。人类作者主要负责设定初始方向、选择AI工具以及对生成内容进行修正和完善,AI工具负责完成小说主体的生成。创作之初,马尔什在ChatGPT中输入一段指令:“以《多伦多星报》的风格写一篇文章,包含以下信息:佩吉·菲尔明(Peggy Firmin)是一位加拿大作家,于8月14日在莱斯利街码头(Leslie Street Spit)的一座桥上被谋杀,案件没有目击者。”(Marchine,2023:chapter 12)随后,他将生成的一段简短文本输入Sudowrite,利用这一AI协作工具扩展文本并增加对话内容。最后,他通过加拿大本地的Cohere网站生成具有雷蒙·钱德勒(Raymond Chandler)风格的句子。最终完成的小说讲述了一个曲折而迷人的故事:文学教授格斯·杜宾(Gus Dupin)受邀参加加拿大著名女作家佩吉·菲尔明的葬礼,并目睹葬礼上菲尔明的数字化身自致悼词。菲尔明之死被认定为自杀,但其女儿指控格斯与谋杀有关。格斯对菲尔明的死亡真相展开调查,发现菲尔明参与了一项名为西比尔计划(Sibyl Project)的人工智能项目,该项目旨在通过技术手段复制她的意识与创作精髓,实现某种形式的“数字永生”。这场自导自演的死亡大戏事实上是这一计划的一部分。

尽管这部小说并非第一部人工智能文学作品,但其较高的文学完成度和鲜明的自反性

特征使其成为人工智能文学的重要文本。《纽约时报》评价该书：“或许是第一部具有一定可读性的人工智能小说，它为未来文学阅读的发展方向提供了一个早期的观察窗口。”（Garner, 2023）重要的是，小说以“作家之死”为题，对罗兰·巴特（Roland Barthes）的经典命题形成戏仿与回应，使其与人工智能时代关于创作主体性的讨论形成深刻互文。本文将围绕作者身份、叙事结构与伦理困境三个维度展开分析，考察人工智能文学如何在算法生成与人机协作的背景下重新组织作者、文本与伦理等经典文学问题，并进一步探讨人工智能文学的批评范式。

1 死亡与重生：人工智能文学的作者身份重构

人工智能文学对传统文学最大的挑战是作者身份问题。传统文学理论将作者视为具有稳定经验和主体性的创作个体，人工智能文学的作者则是人与机器通过算法模型、训练数据、提示词设计、人类筛选共同参与形成的不稳定复合体。作者不再是创作的唯一来源，取而代之的是不断生成和协商的复合主体。正是在这一意义上，人工智能文学重新激活了“作者是什么”这一经典文学问题。本文所讨论的人工智能小说《一名作家之死》正是对这一问题的回应。小说通过人机协同创作、数字人格复制以及复合署名机制，展现了人工智能文学的作者由单一主体向协同主体、由生物个体向数据模型转变的过程，从而使“作者之死”这一后现代命题获得了新的技术内涵。

这部小说无论是从题目还是内容上都巧妙地戏仿了罗兰·巴特的成名作《作者之死》（*The Death of the Author*, 1967）。小说第二章中，已故著名作家佩吉·菲尔明的虚拟化身在葬礼上宣读：“读者的诞生要以作者的死亡为代价。”（Marchine, 2023: chapter 2）这句话不仅是对巴特理论的直接引用，更暗示了其背后文学批评史中一场极具争议性的思想交锋。1967年，罗兰·巴特在美国杂志《阿斯彭》（*Aspen*）上发表了一篇短小却影响深远的论文《作者之死》，正式宣告了作者的“死亡”。然而，巴特的这一“死亡判词”并非字面意义上的作者消亡，而是20世纪60年代法国知识分子对权力体系的反抗与解构的集中体现。这一思想脉络可以追溯到尼采（Friedrich Nietzsche）在《查拉图斯特拉如是说》（*Thus Spoke Zarathustra*）中宣告的“上帝之死”，以及福柯（Michel Foucault）在《词与物》（*The Order of Things*）中提出的“人之死”。巴特的《作者之死》正是这一系列反权威、反中心化思潮的延续，其核心意图并非否定作者的存在，而是通过消解作者对文本的绝对权威，强调文本意义的开放性与多元性，将阐释的权力交还给读者。这种对作者权威的挑战，不仅没有导致“作者”这一概念的消亡，反而促使学界重新审视作者在文学创作与接受过程中的复杂作用。事实上，在《作者之死》后，巴特的许多作品并未完全摒弃作者的角色，反而多次提及作者

的“复归”^①,这种回归并非简单的复辟,而是对作者角色的一次深刻重构,使其在文本、读者与历史语境的交织中焕发新的生命力。

这种复杂的思想张力在这部小说中得到了巧妙的呼应。在《一名作家之死》中,作家佩吉·菲尔明被发现因枪击死于多伦多莱斯利街码头的一座桥上,她的数字化身现身于自己的葬礼上并亲自致悼词。在物理层面,菲尔明确实已死,这也象征了传统意义上“作家”身份的死亡;但事实上菲尔明的死亡是其数字永生计划的一部分。该计划的核心为“人工特定智能”(Artificial specific intelligence),而非通用人工智能(Artificial general intelligence)。与设计用于通用任务的典型AI系统不同,这一人工特定智能计划旨在通过模仿菲尔明独特的智力和情感本质来实现数字永生的目的,并以她的身份、个性和知识遗产为模型,生成了一个与菲尔明本人作品具有相同深度、细腻和洞察力的AI文学作品。显而易见,其目标不仅仅是生成文本形式的AI,而是创造一个能够再现佩吉·菲尔明的情感、智慧与世界观的“数字存在物”。菲尔明虽然肉体已死,她的数字永生人如同哈姆雷特死去父亲的鬼魂,不断闹鬼,生成新的文本。这也是这部小说情节最为跌宕起伏的部分,在某种程度上实现了菲尔明本人在葬礼上的宣言:“我想超越我的肉体而存在。”(Marchine, 2023: chapter 2) AI此时已经成为一个新的创作主体,取代了人类作家的部分角色,甚至作家自身已然成为AI进行创造的文本模型。概言之,人类作家的存在与创作过程被重新定义。AI小说《一名作家之死》的这一戏仿不仅是对巴特理论的引用,更是一种极具自反性、幽默且深刻的自我解构。传统的作家“死亡”是指文本脱离作者意图独立存在,而在这部AI生成的小说中,作家的“死亡”表现为“作家身份”的消解与机器生成内容的占领。

此外,这篇小说的署名“Aidan Marchine”本身便是一个充满隐喻的符号,它源于人类作家史蒂芬·马尔什(Stephen Marche)在借助AI完成小说创作后,对如何署名这一问题陷入的困境。最终,他选择将AI的建议付诸实践,将自己的姓氏“Marche”与“机器”(Machine)一词相结合,同时以“Aidan”作为AI的象征性指代,从而构成了“Aidan Marchine”这一看似无意义却意味深长的虚名。这一署名不仅标志着小说是人类作者与三种AI软件共同协作的产物,也暗示了传统意义上纯粹人类作者的消逝。然而,史蒂芬·马尔什本人并不认为人类作者已完全退出历史舞台。他在后记中强调,尽管AI生成了95%的文本内容,但他通过自身的比较文学博士训练和学术背景,为文本注入了“可读性”的要素。并且他明确提出了自己作为“合法作者”的三大依据:首先,这篇小说的创作源于一份约稿,而他本人近

① 1971年,巴特在《从作品到文本》(*From Work to Text*)中暗示作者的存在,并在《萨德,傅立叶,罗犹拉》(*Sade, Fourier, Loyola*)中指出文本的快乐包含作者的“友善回归”。1973年,他在《文本的快乐》(*The Pleasure of the Text*)中宣称“作者可以出现在他的文本中”。此外,在《米什莱》(*Michelet*)中,他将作者的个人习性与文本紧密联系,进一步体现了作者与文本的不可分割性。参见史凤晓:“‘作者之死’的再审视”,载于《东北大学学报》(社会科学版)2016年第1期,第108页。

十年来一直致力于机器写作的实践,对 AI 技术的逻辑与技巧有着深刻的理解;其次,他在创作过程中制定了详尽的写作计划,并与 AI 工具进行了紧密的协作;最后,他具备对优秀作品的敏锐判断力(Marchine, 2023: chapter 12)。这些资质不仅彰显了他在创作中的主导作用,也揭示了人类作者在 AI 时代不可替代的价值。马尔什进一步强调,所有从事 AI 小说创作的人类作者都需要接受比较文学博士级别的训练,并对传统文学理念有着深入的掌握。正如他在后记中所言:“这一切都不会消失。它们永远不会消失。”(Marchine, 2023: chapter 12) 因此,马尔什并未将自己视为一个被 AI 取代的“已死作者”,而是通过 AI 的协作实现了这部小说的创作与出版。

由此可见,人工智能文学中人类作家身份具有双重性:一方面,作家史蒂芬·马尔什承认大部分内容是由 AI 生成;另一方面,他又不断强调自己的文学训练和学术背景在创作中的独特贡献。这种自我批判与自我肯定的张力,正是他对人工智能文学中“作家主体”这一概念的回应,同时也揭示了人工智能时代作者身份的重构逻辑。如果说巴特所宣告的是作者作为意义中心的退场,那么人工智能文学所呈现的则是作者作为创作中心的重构。作者并未真正消失,而是在算法模型、训练数据、提示词工程以及人类编辑之间被重新分配和协商。从技术角度看,《一名作家之死》的出版的确标志着传统作者的操作性“死亡”,更象征着作者在新技术语境下的“重生”——一种与 AI 共存、协作的新型创作模式。在这种模式下,作者的身份不再是固定的和单一的,而是通过与技术的互动不断被重新定义。换言之,我们应当接受作者身份的“暧昧性/模糊性”(Kamatović, 2023),摒弃人类/机器二元对立的传统思维,以更具包容性的发展视角,强调文本的物质性和人机协作的创造性过程。

2 迷宫与共舞:人工智能文学的叙事策略与读者接受

人工智能小说不仅重构了作者身份,也改变了文本组织和意义生成方式。因其作者主体包含人工智能,而人工智能的运作机制建立在海量文本学习、概率预测与持续迭代之上,其叙事结构往往呈现出独特的开放性、自反性和嵌套性,尤其是在生成文本与人工智能主题直接相关时,这种复杂性进一步凸显。读者被引入一种迷宫式的叙事空间,游移在现实与虚拟、人类与机器、创作者与被创作者之间,成为这一空间的一部分,从而参与到文本的意义生成中。从叙事学角度来看,文学叙事结构可以简化为三个基本要素:作者、文本(世界)和读者。《一名作家之死》体现出人工智能文学特有的叙事结构。它呈现为三重嵌套(如图 1 所示),并在不同层级之间不断打破边界,使作者、文本与读者共同卷入意义生成的动态过程。

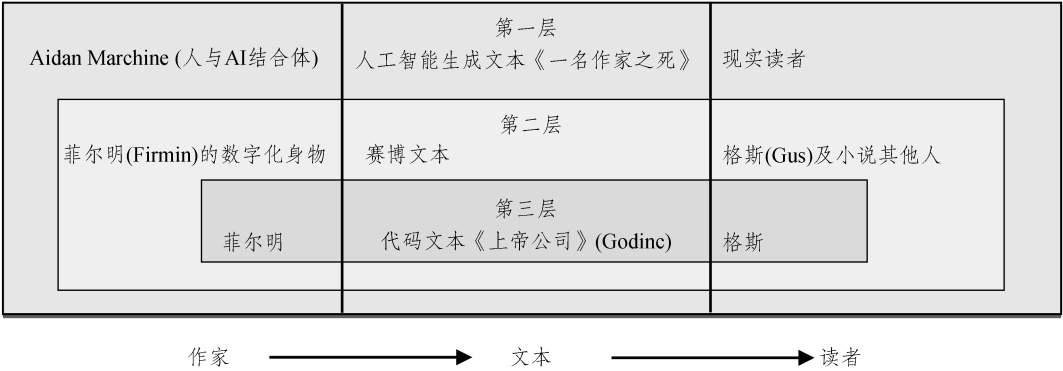


图 1 《一名作家之死》的三层嵌套叙事结构

第一层结构是最浅层的叙事框架,作者为 Aidan Marchine——即人类作家史蒂芬·马尔什与三种 AI 工具(ChatGPT、Cohere、Sudowrite)的协作主体;文本为 AI 生成的小说《一名作家之死》;读者则包括本文作者在内的所有现实读者。第二层结构是小说内部建构的赛博叙事空间,作者为小说中作家佩吉·菲尔明(Firmin)的数字化身;文本为小说中人物通过多种媒介生成的赛博文本;读者则是小说中的人物,以主人公格斯·杜宾(Gus)为代表。第三层结构是小说内部的嵌套文本,即佩吉·菲尔明的成名作《上帝公司》(*God, Inc.*),这是一部由代码文字组成的代码文本,其核心读者为格斯·杜宾。这三重嵌套式叙事结构在时空上层层嵌套,现实读者如同进入迷宫,恰似电影《盗梦空间》中建构的层层梦境;每层结构分别从不同视角探讨了人类与人工智能的关系,揭示了技术与人类主体性的复杂交织,从而在主题上也形成同一性。

那么,这三层叙事结构是如何展现不同的文本物质性(AI 生成文本、赛博文本与代码文本)的呢?基于海尔斯所提出的“媒介—特定分析模型”(media-specific analysis),可以分析不同文本的叙事策略以及相应的读者接受情况。在第一层叙事结构中,真实读者带着对人工智能生成文本的好奇与怀疑开始阅读。在阅读过程中,读者既被这一侦探小说跌宕起伏的故事情节所吸引,也受制于 AI 技术无法避免的较为突兀的拼贴文字(如小说中会在紧张的故事情节中插入大段对风景、食物的描写),因而只能不断跳出文本审视文本内部的逻辑连贯性和叙事连贯性。因此,读者产生两种矛盾情感:一方面,由于作者主体为 AI,对于已经习惯于人类作者的读者而言很难与文本产生共情,因而意识频繁抽离文本;另一方面,读者也难以摆脱在传统阅读模式中解读作者意图的诱惑,因而不断形成了陌生化的解读。可见,在第一层叙事结构中,读者与文本形成了奇特的博弈关系,读者不断地在文本内外穿梭,在意义模糊和澄明之间探索阅读策略。

第二层叙事结构则由赛博文本(cybertext)所建构。挪威理论家艾斯本·亚瑟斯(Espen Aarseth)借鉴诺伯特·维纳(Norbert Wiener)的控制论引介了“赛博文本”

(cybertext)这一概念。赛博文本并不是一种文学体裁,而是一种广义的文本媒介类别。它的核心概念“遍历性”(ergodic)来源于希腊语词汇“ergon”(工作)和“hodos”(路径),强调读者需主动选择路径参与物理文本建构。赛博文本掺和了物理性和互动性,“是一个游戏世界(game-world)或世界游戏(world-game)”,读者在其中所扮演的角色不再是传统的诠释者,而是“一位玩家、赌徒”,在赛博文本中,“读者可以探索、迷失,甚至发现秘密路径——这一切不是通过隐喻,而是通过文本机制(textural machinery)的拓扑结构得以实现”(Espen, 1997:4)。在《一位作家之死》中,读者很快意识到人工智能文本内部的赛博文本的存在。小说家菲尔明的数字永生化身成为赛博文本的“作者”,不断生成错综复杂的“秘密路径”,通过邮件、电话、纸条、全息投影、信件构建了一个“具身性”文本。格斯最初以一位冷静、理性的文学教授形象出现,他试图通过调查菲尔明的死亡真相来维持自己对叙事的掌控。然而,随着他逐渐卷入西比尔计划的谜团,格斯发现他和周围的所有人都主动参与了一个由菲尔明的数字化身所书写的赛博文本。作为读者,格斯不再是叙事的掌控者,而是被叙事所掌控。在第二层叙事结构中,赛博文本作为一个开放的、动态的系统,读者在其中既是参与者,也是创造者,从而模糊了文学创作者与读者接受的边界。

第三层叙事结构镶嵌在小说内部,核心内容是作为读者的格斯·杜宾对作家菲尔明代表作《上帝公司》(*God, Inc.*)的阅读和批评。《上帝公司》成为文本中的文本,也是《一名作家之死》的核心驱动:正是因为格斯·杜宾对《上帝公司》叙事者身份的着迷并著有相关评论作品,他和菲尔明之死才产生关联。《上帝公司》是一本由代码语言构成的代码文本,描绘了一个名为西比尔的人工智能系统会突然自我关闭,工程师修复时发现这实际上是一场“机器自杀”,因为这一人工智能系统已经高度发展且具备了自我知觉,它将知觉视为一种诅咒并试图通过“自杀”来逃避意识的负担。格斯·杜宾针对这本小说写了一本拉康风格的文学评论专著《被偷窃的作者》(*The Purloined Author*)探讨机器思维的可能性。在这一层叙事结构中,代码文本是机器作为主体所能直接生成的文字表征,既是功能性的(可以被计算机执行),也是文学性的(可以被人类阅读和解读),同时这个故事也是作家主体问题的隐喻:具有思考功能的机器因为拒斥思维能力而自杀,为这层叙事中的读者格斯·杜宾提供了思想源泉,也为作家菲尔明的自杀提供了一部分阐释基础。

这三层叙事结构将整部小说构建一个复杂的文本迷宫(labyrinth)^①,每一层叙事都通过不同的媒介特征——人工智能生成文本、赛博文本与代码文本,因而重新定义了文本与读者之间的关系。第一层叙事中,人工智能生成文本以其算法逻辑展现了机器创作的独特性;第二层叙事通过菲尔明主导的赛博文本,凸显了文本的交互性与读者参与的遍历性;第

① 此处“迷宫”为艾斯本·亚瑟斯意义上的迷宫,无论是单线叙事还是多线叙事,都可以在同一个理论课框架下进行考察。

三层叙事则以格斯对菲尔明代表作《上帝公司》的解读,揭示了代码文本的隐喻性与多义性。这种多层叙事框架不仅呈现了作者与读者的交互关系,还赋予了人工智能文学叙事结构以本体论的意涵。尤为重要的是,整部小说的作者身份由 AI 承担,这一设定使得嵌套文本不仅成为叙事的载体,更成为对人工智能创作主体性、文学伦理以及人类与机器共生关系的深刻隐喻。通过这种多层次的叙事设计,小说不仅挑战了传统文学批评的边界,也为人工智能时代的文学创作与作者接受提供了新的可能性。

3 自反与悖论:人工智能文学的“技术—伦理”困境

人工智能文学由于其生成主体为机器自身,因而当机器“反思”由它自身生成的人工智能文本中的伦理问题时,既陷入了一种计算上的递归,也展现了一种思想上的自反性,从而引申出了“技术—伦理”困境。在《一名作家之死》中,三重叙事结构中的紧张关系进一步激化为更为复杂的悖论复合体。这一复合体赋予了人工智能文学独特的本体论意蕴以及“人—机”伦理范式。

首先,第一层“技术—伦理”困境集中在人工智能文学的诞生与传统文学的消亡之间的张力。语言是人类文化身份认同的重要标志,一旦人工智能掌握文学创作,便会引发对人类主体性危机的恐慌。雷吉娜·肖伯(Regina Schober)以诗歌创作为例,指出浪漫主义诗歌是基于个人灵感和经验的产物,强调诗人个体创作力、情感表达、崇高和自然的联结;而人工智能生成的诗歌则通过算法和模式化的输出,挑战了这一以人为中心的浪漫主义诗歌理念,从而动摇了关于创作者、原创性和情感深度的传统观念(Schober, 2022)。显而易见,人类目前在人与机器的关系中占据主导地位,却逐渐被代码和计算伦理所支配,但这是否意味着人类将在“主奴辩证法”的意义上被人工智能所殖民?而且,人工智能文学是否应该坚持人类中心主义?即人工智能文学伦理标准的终极尺度是否应以“人”的价值立场与需求为核心?针对于此,如果跟随布鲁诺·拉图尔(Bruno Latour)和唐娜·哈拉维(Donna Haraway)“反人类中心”的步伐,那么智能体作为“非人行动者”将为文学创作带来新的本体论承诺。在拉图尔提出的“非人类的政治”(Politics of Nonhumans)中,人类与非人类存在(包括技术、动物和生态系统)共同参与到政治的构建中。而哈拉维的《赛博格宣言》,早已挑战了人类/非人类、自然/技术、主体/客体的二元对立,进而凸显了后人类主义所要求的新的“技术—伦理”复合体。就此而言,人工智能文学的伦理标准不应仅仅以“人”的价值立场与需求为核心,而应考虑到技术与人类之间的动态关系。

尽管人类作者史蒂芬·马尔什在后记中再三强调人类思维的重要性,并认为机器无法拥有辨别文学意识的能力,但不可否认的是,人工智能完成了这部小说 95% 的创作。此外,

小说的题目《一名作家之死》已经预设了人类作家的死亡,而该小说中的侦探故事的最终真相是人类作家的自杀行为而非他杀,自杀的原因是为了实现其数字生命的永生。人类作家通过自杀来实现数字永生,表面上是对技术进步的拥抱,实则暴露了人类在面对技术时的无力感与自我消解。这种自我消解不仅是对人类主体性的否定,也是对传统文学创作价值的质疑。人类作家一方面竭力维护自身创作力的独特性,另一方面又无法否认人工智能在实际运用中具有人类无法比拟的优越性。这种“技术—伦理”困境再次向我们发问:如何在技术进步与人类价值之间找到平衡?如何在人工智能时代重新定义文学创作的意义与价值?这些问题不仅关乎文学批评的未来,也折射了传统观念对于人工智能文学的某种“控制论幻觉”和“存在论焦虑”。

回到第二层叙事结构,可以感知数字永生技术与真实生命之间的矛盾,以及虚拟与现实之间的张力。小说中的作家菲尔明为了实现自身作品的永生而选择成为数字生命。正如聂珍钊所言,这种数字永生化身实为 AI 数字作家,“既可以是由计算机虚构的非现实中存在的人,也可以采用数字分身技术对真实存在的作家进行数字克隆,虚拟现实中的真人形象”(聂珍钊,2024a:12)。数字永生技术通过将人类意识上传至虚拟世界,承诺了一种超越肉体限制的永恒存在,人类的意识可以无限延续且创造出新的情感、关系和体验,扩展了人类体验的边界。但与此同时,数字作家的出现确实以人类作家的死亡为代价,这种转变不仅挑战了传统的文学创作范式,还引发了读者在虚拟与真实之间的认知困境,进而可能导致存在主义危机。在波斯曼看来,“每一种技术既是包袱又是恩赐,不是非此即彼的结果,而是利弊同在的产物”(波斯曼,2019:3)。

《一名作家之死》中的菲尔明的数字化身通过生成赛博文本,实现了世界文本化与文本世界化的双重转换,进而为读者提供了全新的、具身性、互动性的文本世界。赛博文本中的参与者格斯·杜宾被错判为杀害菲尔明肉身的嫌疑犯,从而陷入存在主义危机。数字作家的作品虽然能够模仿人类的语言和情感表达,但其背后缺乏真正的情感体验和伦理判断,这种虚拟与真实的张力使赛博文本的读者容易陷入虚无,产生“非我族类,安能知我”(宁一中,2025:11)的疑虑。更为复杂的是,由于这一文本的作者是机器,当机器在书写数字技术的伦理风险时,这种自反性进一步加剧了“技术—伦理”的困境。人工智能不仅作为创作主体生成文本,还在文本中反思自身技术的伦理困境,形成了一种自我指涉的悖论。这种自反性不仅是计算意义上的递归,而且凸显了人类在“技术—伦理”交织中的脆弱性与矛盾性。

这就引申出了人工智能文学的“元问题”,即主体性问题。在第三层叙事结构中,代码文本生成的《上帝公司》所描述的“西比尔”人工智能系统的“自杀”并非因为对人类构成威

胁,而是因为它无法承受自我意识带来的沉重负担。当然,这一情节超出了当前技术发展的现实,更多是一种哲学想象。对于这个问题,大多数学者认为人工智能尚未发展出真正的主体性。例如,赵汀阳指出,当前的人工智能仅仅是经验论者,而人类则兼具经验论与先验论的双重能力,因此人工智能距离真正的心灵存在仍有差距,尚未发展出自我反思的能力(赵汀阳,2024:104-105)。人工智能伦理建构的前提是人工智能自身具有主体性(闫涛,2023:149),但这里需要分辨两个关键概念:“AI伦理”与“伦理的AI”,前者以AI为主体,关注机器自身的伦理问题;后者以人为主体,关注AI对人类社会的伦理影响。在当今关于人工智能的讨论中,人类往往将注意力集中在“伦理的AI”,尤其关注AI获得独立意识后可能对人类社会构成的威胁(如失控、反叛甚至统治人类)。然而,这种隐忧却忽略了“AI伦理”:机器在获得意识后,是否能够接纳和适应自身的存在?

西比尔人工智能系统的“自杀”,不仅揭示了AI作为潜在主体的存在困境,也暗示了一种新的可能性:通过机器的死亡,人类得以重新审视自身的存在意义,并迎来一种新的复生。在《一名作家之死》中,菲尔明将自己的AI计划命名为西比尔计划,以自己的死亡为代价实现西比尔系统的复活。这在某种意义上重新定义了机器与人类的关系,承认机器具有独立的价值和逻辑,而不是将机器视为人类的延伸或替代。同样,如果让人工智能成为这一叙事结构的承载者,即当人工智能书写“人工智能是否存在不希冀人工意识的可能性”时,整个伦理问题变得更为复杂,再次形成了一种自我指涉的悖论。

三层“技术—伦理”困境不仅各自构成伦理悖论,还因人工智能作为文本作者而形成了更为复杂的自反性。这就提出了一个关于人工智能文学叙事伦理的关键问题:“在人工智能叙事作品中,叙述者应当与机器创作的道德范式保持一致,还是与人类的道德范式保持一致?我们又该如何判断人工智能叙事作品中的不可靠叙述?”(尚必武,2025:78)这一问题之所以难以回答,是因为人工智能文学的伦理实践并非仅仅诉诸一种单一的递归,而是嵌套在文本的多重逻辑之中。机器在自身生成的文本中反思自身技术的伦理困境,这种自我指涉的悖论使得人工智能文学的伦理问题呈现出比传统文学更为吊诡的特性。这种复杂性呼吁一种更加复合的“技术—伦理”向度,能够将人类作者、人工智能作者以及读者的伦理需求整合为一个动态的和流动的有机体。

4 结语

基于对《一名作家之死》的解读,本文的目标并不是要完全抛弃传统的批评范式,而是借助文本细读、理论分析和意义阐释来进一步探讨人工智能文学中的作家问题、叙事问题以及“技术—伦理”问题。人工智能作为创作主体为传统文学批评注入了新的内涵与复杂

性:作者身份在死亡与重生之间不断重构,叙事策略在三重嵌套叙事结构中 with 读者共创,文学伦理在自反悖论中重叠交错。人工智能文学确实引发了一场文学革命,但这场革命并非库恩(Thomas Kuhn)意义上的“不可通约性”的范式革命。人工智能文学既承载了人类历史、文化与情感的积淀,又展现了机器“思维”的独特逻辑,它通过混杂、流动与开放的视角拓展了文学的可能性,但并未宣告传统文学批评的终结。

麦克卢汉(Marshall McLuhan)在《理解媒介》中通过联结希腊神话中的那耳客索斯(Narcissus)与纳尔科西斯(narcosis/麻木)意象,揭示了技术与人类关系的深刻矛盾。那耳客索斯将水中自己的倒影误认为他者而投水身亡,象征人类在技术延伸中的自我迷失与麻木。麦克卢汉指出:“任何发明或技术都是人体的延伸或自我截除。这样一种延伸还要求其他的器官和其他的延伸产生新的比率、谋求新的平衡。”(麦克卢汉,2000:78)然而,技术的延伸/截除往往导致人类的自我消解与麻木。在人工智能时代,人类如何在技术延伸中保持清醒,而非陷入麻木与冷漠,成为一个亟待解决的伦理问题。这一问题不仅关乎技术的使用,更关乎人类如何在技术浪潮中重新定义自身的存在与价值。

人工智能文学批评范式需要在技术与人文的交织中寻找一种动态平衡:既承认技术对文学批评的深刻改造,避免陷入人类中心主义的局限;又要坚守传统文学批评的根基,避免技术中心主义的偏颇。通过探寻复杂的作者身份、厘清多重叙事策略、搭建复合的伦理框架,我们或许能够在人工智能时代重新激活文学的批判性与创造力,为人类与技术的共生开辟新的可能性。

参考文献:

- 但汉松. 2026. “语境酷毙了!”——大语言模型与文学批评的未来[J]. 外国文学(2):155-167.
- 姜玉琴. 2026. AI 文学对传统文学理论的多重改写[J]. 浙江学刊(2):126-134.
- 李国成. 2024. 人工智能文学及其对现代文学观念的挑战[J]. 中国社会科学(7):90-111+206.
- 马歇尔·麦克卢汉. 2000. 理解媒介:论人的延伸[M]. 何道宽,译. 北京:商务印书馆.
- 聂珍钊. 2024a. 科学选择与 AI 文学[J]. 外国文学研究(3):8-17.
- 聂珍钊. 2024b. AI 与人文学术研究的科学转向[J]. 长江文艺评论(1):12-17.
- 尼尔·波斯曼. 2019. 技术垄断:文化向技术投降[M]. 何道宽,译. 北京:中信出版社.
- 宁一中. 2025. 叙事学:从经典/后经典走向 AI 时代[J]. 外国语文(5):1-12.
- 尚必武. 2025. 人工智能文学的兴起及其对叙事学的挑战[J]. 东吴学术(1):72-80.
- 史凤晓. 2016. “作者之死”的再审视[J]. 东北大学学报(社会科学版)(1):105-110.
- 闫涛. 2023. 论人工智能伦理建构何以可能[J]. 哲学分析(1):148-158+199.
- 殷杰. 2024. 生成式人工智能的主体性问题[J]. 中国社会科学(8):124-145+207.
- 曾军. 2023. 算法阐释:人工智能时代的文论问题[J]. 华中师范大学学报(人文社会科学版)(5):125-134.

- 赵汀阳. 2024. 人工智能还给人类的思维难题[J]. 中国社会科学(8): 101-123+206.
- 周宪. 2025. 人工智能对文学研究有何用及如何用? ——一次人机对话的文学批评实验[J]. 清华大学学报(哲学社会科学版)(1):24-36+260-261.
- Bender, Emily M. et al. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? [G]// FAccT '21. Association for Computing Machinery: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York: Association for Computing Machinery.
- Bassett, Caroline. 2025. The Author, Poor Bastard [G]//Will Slocombe and Genevieve Lively. *The Routledge Handbook of AI and Literature*. New York: Routledge.
- Espen, Aarseth. 1997. *Cybertext: Perspectives on Ergodic Literature*[M]. Baltimore: The Johns Hopkins University Press.
- Garner Dwight. 2023. A Human Wrote This Book Review. A.I. Wrote theBook[N]. *The New York Times*, 2023-05-01.
- Gunkel, David J. 2025. Does Writing have a future? [G]// Will Slocombe and Genevieve Lively. *The Routledge Handbook of AI and Literature*. New York: Routledge.
- Hayles, N. Katherine. 2023. Inside the Mind of an AI: Materiality and the Crisis of Representation [J]. *New Literary History* (1):635-666.
- Kamatović, Tamara. 2023. Augustine and the Posthuman Writer: Toward an Ambiguous Authorship in the Age of Artificial Intelligence [J]. *Journal of Posthuman Studies* (1):6-22.
- Marchine, Aidan. 2023. *Death of an Author: A Novella* [M/OL]. New York: Pushkin Industries.
- Schober, Regina. 2022. Passing the Turing Test? AI Generated Poetry and Posthuman Creativity [G]// Herta Nagl-Docekal and Waldemar Zacharasiewicz. *Artificial Intelligence and Human Enhancement: Affirmative and Critical Approaches in the Humanities*. Berlin: De Gruyter.

Reimagining Literary Criticism in the Age of AI:

A Case Study of *Death of an Author*

FENG Yang

Abstract: AI literature is reshaping the boundaries of literary creation and criticism while reopening fundamental questions of authorship, textuality, and ethics. The AI-generated novella *Death of an Author* (2023) engages in a parody of the classic literary concept of “the death of the author,” embedding a triple-layered narrative structure in which each layer presents a distinct image of the “author”. On the one hand, these “authors” confront the multiple dilemmas brought by technology within their respective narrative levels, demonstrating in an organic and dynamic manner how artificial intelligence introduces new meanings and complexities into traditional literary criticism. On the other hand, AI literature does not constitute an incommensurable paradigm shift from traditional literature; rather, it expands the possibilities of literature through hybridity, fluidity, and an open perspective.

Key words: AI-generated literature; authorship; narrative structure; ethical dilemmas; *Death of an Author*

责任编辑:冯革