

现当代文学研究

“强人工智能”视野下“机器写作”的思想渊源、理论困境及发展进路探析

黄文虎

摘要：人工智能写作涉足文学领域引发了舆论的广泛关注。从“强人工智能”的视角来看，基于算法的“机器作者”甚至可以被视为一种存在于虚拟世界的“类智能体”，这种观点源于主张“心身分离”的计算主义，它是强人工智能的思想基础。由于割裂了身体与心智之间的内在关联，“机器写作”同样无法回避语言表达、语义理解和元语言等认知难题。从发展进路来说，“机器写作”存在两种不同的演进方向，从“无身认知”的视角来看，“机器写作”如同数字化世界中的“符号组合游戏”，它是一种“算法式存在”。从“具身认知”的视角来看，“机器写作”必须依托人类主体的肉身体验来重构“人机协作”这一新的创作生态链。

关键词：强人工智能；计算主义；机器写作；无身认知；具身认知

DOI:10.16100/j.cnki.cn32-1815/c.2021.04.017

引言

随着人工智能技术的突飞猛进，近年来的“机器写作”已经涉足到文学领域，甚至出现了引发舆论追捧的“机器诗人”。就人工智能技术层面而言，计算机在处理非虚构写作以及虚构文体方面还无法做到“得心应手”，因为基于数据和程序的机器写作不等于“创作”。

无疑，在人工智能时代，“机器写作”具有不

可忽视的理论和实践意义，甚至代表了文学领域的“认知革命”，关于“机器写作”的话题也引发了文学研究者的热议。总体来看，当前有关人工智能写作的一个热点话题是比较人工智能与人类智能在文学写作领域的优劣。主流观点认为人工智能虽然具有强大的计算和模仿能力，但在创造力方面远不如人类智能。然而，从“强人工智能”的立场来看，文学创作的本质不过是一套高级算法。既然是可还原的算法，那么“机

器写作”超越人类智能似乎就只是时间问题。

但“强人工智能”视野下的“机器写作”仍然需要面对一系列“认知难题”,此类“症结”的关键在于当前基于深度学习的人工智能写作的实质是基于“无身认知”这一基本范式,忽视了“具身化”(Embodiment)对于高级认知活动的不可替代的价值。因此,“机器写作”的发展进路实际上存在“无身”范式和“具身”范式两种可能性。前者趋向的是替代人类作家的“数字化拟主体”,而后者有可能导向“人机协作”创作模式。由此来看,要深入探讨“人工智能写作”,就必须正视“数字化机器写作”的局限性,重新审视“心身合一”的具身观对人工智能写作未来发展可能产生的深刻影响。

一、“机器写作”的思想渊源追溯

按照技术乐观派的预测,不仅记者的“写稿能力”可以被替代,作家的“创作能力”同样能被机器模仿甚至超越。“人工智能恐慌论者”认为人工智能迟早能够替代人类作家,后者可能不久将面临“失业”的威胁。这一观点正是“强人工智能”的立场。自20世纪50年代以来,人工智能逐步作为一门跨界型的综合性学科崭露头角,经过半个多世纪的发展,它已经形成两个基本的研究纲领,分别是“强人工智能”(Strong AI)与“弱人工智能”(Weak AI)。

“强人工智能”其实是计算主义(Computationalism)的另一种表述,它代表了人工智能的主导性立场,该学派认为机器智能能够像人类一样思维。早在20世纪70年代,就有人工智能研究者极为乐观地估计,不出十年,就会诞生与人类智能相当的“通用智能”(General Intelligence)机器,它甚至能够阅读像莎士比亚戏剧这种复杂的文学体裁。^①在20世纪80年代,约翰·塞尔(John Searle)把此类将“人脑视为计算机”的主张称为“强人工智能”。^②而“弱人工智能”则只是将计算机作为模拟人脑部分功能的工具。

按照技术乐观派的主张,“机器写作”或“智能写作”绝不是冷冰冰的数据和程序所预设的产物,而是有朝一日也能形成与人类心灵相媲美的“机器之心”。在计算主义看来,算法是可

通过计算所得出的规律。由这种“一切皆可转化为算法”的主张来看,作诗机通过有穷的步骤根据特定的规则作出一首完整的“五言诗歌”,那么该程序就可被视为“能行可计算”的有效程序。

从认知科学的视角来看,当前人工智能在文学写作领域的应用主要是基于大数据和深度学习为基础的智能算法,这一智能技术的哲学基础是以“无身认知”(Disembodied Cognition)为主导的计算主义。“无身”意味着心智与肉身的分离,因此,机器的“无身写作”正是试图摆脱肉身的束缚,而尝试通过可操控和可表征的算法来生产文学作品。而机器诗人“小冰”正是“无身写作”的代表,因为它是一种超越肉身之上的“算法式存在”。

“无身写作”的哲学基础是计算主义的“无身认知”,其核心思想是将“机器写作”视为一种高级心智功能的呈现,它可以脱离于肉身通过算法实现。计算主义认为,心智可以与肉身相分离而不会受后者的影响,正如计算机硬件与软件的关系。“机器写作”与人类写作最根本的区别在于是否需要存在一个肉身的主体。在强人工智能的观点来看,创作者的创作思维与创作能力是一种特殊的认知功能或认知状态,而不是与肉身密不可分的主体存在。因此,无论是机器、人类或其他有机体,只要能够模拟出一套同一或相似的心智运作模式(如图灵机模拟人类处理符号的编码模式),它就能够被“迁移”到“无身”的机器之中。因此,人类的意识、理智、情感等认知活动都可以被还原为脱离身体而存在的人工算法。

对人工智能影响深远的“图灵测试”可以被视为“心身分离论”观念影响下的一个思想实验。^③“图灵测试”的核心目的是要探讨“机

① Christian Lexell. *Artificial Intelligence Versus Human Intelligence: Are Humans Going to Be Hacked?* Springer-Briefs in Applied Sciences and Technology, 2019, p.9. <https://doi.org/10.1007/978-3-030-21445-6>

② [美] 约翰·塞尔著,杨音莱译:《心、脑与科学》,第20页,上海:上海译文出版社,1991。

③ Turing, A.M. "Computing machinery and intelligence. Mind", LIX(236), 1950, pp.433-434.

器能否(像人一样)思维”。但图灵却暗含了一个前提条件:即思维的呈现完全可以脱离身体而存在。从心身关系的角度来看,“图灵测试”所预设的场景是完全“去身体化”的心智对话。在测试中,文字符号是识别人类心智与机器心智的主要媒介。无论是机器还是人类,都必须通过文字符号作为心智活动的外在化表征。也就是说,图灵所说的“思维”其实是一种可以超越肉身之外的心智活动的规则化表征。因此,在“模仿游戏”中,“身体表意”在符号传达中的重要性被最大限度的清除了。^①

2019年的一则报道称,机器诗人“小薇”是“中国第一个通过图灵测试的作诗机器人,入选中央电视台《机智过人》节目。”^②对于机器诗人“小薇”是否通过“图灵测试”我们可以暂且存疑。进一步可以提出疑问,若读者的确无法分清一首诗(或其他文学体裁)是人类创作还是机器创作的时候,那么这是否就证明智能机器作家已经“懂得”创作思维了呢?问题的关键在于如何理解人类的思维。

“机器写作”的发展思路沿袭了这一强人工智能的立场,机器诗人/作家只要在数字计算机中“制作”出与人类创作者“不相上下”或“难辨真假”的作品,就可以被认定为这些“无身机器”同样具有创作思维或文学创作的能力。从符号学的角度来看,这一观点确实具有强大的说服力,如卡西尔所言:“人是符号的动物。”^③人类的认知活动必须通过符号这一载体加以表征,既然数字计算机能够模仿人类的语言符号构筑“拟人化”的文学世界,那么至少从效果上来看,机器诗人的作品证明了一个事实:即文学写作可以被还原为规则化的机器算法。

既然人类的一切心智活动(包括文学写作)都是程度不同的可量化计算过程,那么“文学写作”自然可以被表征为形式化的算法规则。2018年,机器诗人“小冰”的学习对象已经多达上千位诗人,它在短时间内就创作了数以万计的诗歌,从效率上来看,“小冰”无疑显示出了远超于人类诗人之上的“迭代学习的能力”。但也有批评家指出“小冰”的诗歌中存在不少带有明显拼贴痕迹或令人费解的诗句。^④不过,按照计算主义的思路,机器诗人或机器作家作

品中的缺陷是由于算法规则还不够完善,比如可以通过进一步扩大数据库或优化情感算法来提升“机器人”在写作能力方面的“拟真度”,使其不断贴近人类作家,直至有一天能够“骗过”最挑剔和最富有经验的人类读者。之所以强人工智能信奉者如此相信算法的力量,其认识论基础就在于他们坚信“人类认知和智能活动实质就是计算。”^⑤

在强人工智能看来,广义上的计算就是主体或“拟主体”(计算机)的一种自动或被动的选择能力。人类理智活动中的各种“算计”往往基于一种理性的选择,而非理智活动则是一种基于特定心理状态(激动、愤怒等情绪导向)之下的选择。计算主义主张貌似缺乏规则的非理智活动也能够被表征为一套非随机性的算法规则。所谓“愤怒出诗人”,形容的就是情绪化状态下所激发的超越规则之外的灵感思维。不过,计算主义者主张理智与情感只是基于复杂程度不同的规则,本质上都可以被形式化为一套算法。在“机器写作”领域,这体现在“人工智能小说”的出现。在2016年,人工智能创作的小说《电脑写小说的那一天》参加了日本“新星一奖”,该事件曾引发舆论热议,甚至有人担心计算机是否会代替网络写手。^⑥

就目前来看,人工智能小说必须依靠人类的协作才能完成,并且存在不少难以避免的“硬伤”。但是按照计算主义的理论,当前“机器写作”所无法或难以处理的非形式化内容都必

① Claude Draude. "Intermediaries: Reflections on Virtual Humans, Gender, and the Uncanny Valley", *AI & Society*, 2011(26): 319-327.

② 经济日报微信公众号:“小薇”为何能通过图灵测试? 2019-03-29, <https://baijiahao.baidu.com/s?id=1629319184902138220&wfr=spider&for=pc>

③ [德] 恩斯特·卡西尔著,甘阳译:《人论》,第46页,北京:西苑出版社,2003。

④ 白亮:《技术生产、审美创造于未来——基于人工智能写作的思考》,第41页,《南方文坛》2019年第6期。

⑤ 酈全民:《用计算的观点看世界》,第10页,广州:中山大学出版社,2009。

⑥ 阅听文学微信公众号:《人工智能已能写出完整小说!网文作者即将被AI代替?》,2018-10-11, <https://baijiahao.baidu.com/s?id=1614005882416515360&wfr=spider&for=pc>

将在未来依靠一种更为高阶的表征体系加以规则化。假如仍然有不能规则化的创作内容,那么就必须进一步创造更为强大的算法来解决。

也就是说,强人工智能的底层逻辑是:“无身”的机器完全可以将“不可表征的认知活动”(如创作灵感体验)转化成可表征的形式。基于这种“表征主义”的思路,“机器写作”的过程实际上就是一种“去情境化”的信息加工和处理过程。以算法形式存在的机器诗人无需背负“沉重的肉身”,它如同一个存在于数字化文学世界中的“超级大脑”。

因此,从思想渊源来看,计算主义的“无身认知”与“表征主义”奠定了“机器写作”的本体论基础。在“无身”的机器之中,文学主体不是被异化或消解,而是完全不存在了。对于机器而言,文学的意义不再关注基于生命体验的创作过程,而是聚焦于如何通过规则化的精巧算法来模拟文学符号的批量生产。在“表征主义”观念来看,需要通过形式化的高级算法来为文学创作的神秘性“去蔽”。与新闻报道等纪实型体裁相比较,虚构型的文学体裁的重心是需要优化“难以表征”的情感算法。

二、“机器写作”所面临的理论困境反思

针对文学这种关涉人类想象能力的虚构型文体,“无身的机器写作”所暴露出的缺陷并不是可以通过持续扩容的数据与不断迭代的算法就能轻易解决的技术性问题,而是涉及强人工智能对于人这一“具身性”主体的“忽视”。计算主义思路下“机器智能”最根本的缺陷在于切断了思维、身体与世界之间的有机联系,它仅仅将智能视为一种“离身性”的孤立于周遭环境之上的“算法式存在”。

正因为这种认识论层面的缺陷,“机器写作”在语言表达、语义理解和“元语言”三个层面很难抵达人类智能的认知高度。

第一,“机器写作”的语言表达困境。文学是语言的艺术,而由语言文字所构筑的文学世界是一个充满无限可能的“符号世界”。因此,无论是传统的符号主义抑或与大数据结合的神经网络算法,仍然无法回避一个基本问题,即算

法是否能够构建文学语言表达的普遍规则?以“作诗机”为例,假如要“创作”一首五言绝句,程序员可以通过数据挖掘技术来创建一套五言绝句的海量数据库,同时制定一套五言绝句的作诗规则,然后机器算法再通过深度学习不断“迭代”,从而“把握”作诗的基本规则,进而能够在短时间内模仿并批量生产出难分真假的“五言绝句”。实际上,机器诗人小冰就是以这种“数据+神经网络算法”的运算模式作为基础。

问题在于,由于文学写作原本是一套天马行空、超越规则之上的语言艺术,因此,当面临复杂多变的文学语言表达之时,当前任何算法都无法从海量的数据中建构出一套“放之四海而皆准”的文学规则。因此,拥有海量数据的机器智能在处理文学语言搭配和组合方面很容易暴露出缺乏常识性的“荒谬”。

以机器诗人小冰的诗作《黄昏里来了一碗茶》为例。诗作如下:

“黄昏里来了一碗茶,回家一齐看——嘴里的妻子已失去了/让野火的人们/风景如风车里一碗茶凉/是少年的故事/回家一年的时候/我猜我也一例有敌骑的呼声响。”^①

首先,此诗中存在几处不通顺之处,如“我猜我也一例有敌骑的呼声响”,这属于句法层面的常识错漏。其次,从全诗的主要意象来看,茶、妻子、风车、少年、敌骑之间缺乏内在的关联,存在“拼凑”的痕迹。最后,由于先锋派诗歌强调对语言规则本身的“反叛”,这使得“常识性错误”和“陌生化手法”的边界显得暧昧不清。由于人类诗人天然具有追求“陌生化”表达形式的冲动,因此,诗句的组合方式往往呈现“反常识”的特征,但这种“反常识”依然是建立在对常识的认知基础之上。显然,我们很难想象一个人或机器在不了解常识的情况下能够判断什么叫“反常识”。有鉴于此,“缺乏常识”的机器算法与“反常识”的诗歌创作之间其实存在一道难以跨越的鸿沟,其主要原因在于当前的神经网络算法仍然没有克服与常识密切相关的情境难题。

① 小冰:《阳光失了玻璃窗》,第14页,北京:北京联合出版公司,2017。

辨别非表征化的常识离不开具体的历史、社会和文化情境,而对于机器诗人而言,之所以它总会犯下某些明显的常识性错误,其重要原因是它的计算过程是一种“去情境化”的算法。就文学写作而言,人类个体的认知活动都必须依托主体所面临的特定情境。尽管作家创作时的真实心境很难“还原”,但作家创作作品之“意”必然基于具身化的特定情境,如李白的“呼儿将出换美酒,与尔同销万古愁”与诗人好酒放达的豪爽性情直接相关。而杜甫的“国破山河在,城春草木深”与诗人爱国主义情怀密不可分。而读者之所以能够“感同身受”也是因为他/她和作家都是基于情境化的具身存在。而对于从未“尝过”酒味和没有国别身份的机器而言,它显然缺乏通过肉身来感知情景的能力。

第二,“机器写作”的语义理解困境。不可否认,当前的机器智能的确在文学语言的表达层面可以做到一定程度上的“以假乱真”,但这无法证明机器智能具有语义层面的理解能力。

为了驳斥机器能够“理解”故事这一说法,塞尔设计了一个名为“中文屋”的思想实验。“中文屋”假设一个只懂英文的人被关在一间封闭的房间,此人通过一本“万能”的中文宝典可以准确地回答房间外面所提出的任何中文问题。其结果是屋外的人以为这位不露面的神秘人物精通中文,但实际上他对中文一无所知。房间中不懂中文的人隐喻的正是计算机程序。塞尔认为基于形式系统的计算机无法“理解”意义,正如“中文屋”中不懂中文的人只能处理基于信息输入与输出这一句法或语法层面的符号表征体系,而无法上升到语义层面。塞尔认为:“如果心当真是一台计算机,它的运算必须用语法来定义,然而,意识、思想、感情、情绪以及心理的所有其他特征远非语法所能包容。”^①按照塞尔“中文屋”的观点,人工智能写作不可能拥有某种意向状态(如意图、愿望和信念),它可以模仿“我爱你”这样的语言表达形式,但它不可能真正理解“我爱你”背后的语义内涵。尽管塞尔“中文屋”这一思想实验的提出已经有三十余年,但争议并未结束,它对如何理解时下流行的人工智能写作仍然具有重要意义。

以人工智能诗人小冰的《阳光失了玻璃窗》

为例。该诗集的出版团队认为小冰是一个兼具IQ与EQ的人工智能创作主体,“她”创造诗作的过程可以对应于人类作家的创造性活动。^②从“中文屋”的视角来看,小冰的诗作“生成”的底层逻辑是一套基于大数据为基础的高级算法,它只能在一套预设的算法规则基础之上生产和重组已有的符号组合关系,因此它无法“理解”文学语言对于创作主体或接受主体的意义何在。也就是说,小冰所谓的创造性活动仍然是在句法和语法层面对于诗歌符号表达形式的“拼合”,所以它对于符号世界中“爱”与“恨”的模拟不代表它能够区分“情感”一词在语义层面的涵义。而从“中文屋”反对者的观点来看,人工智能诗作的“生成”显示出一种“自我迭代”的认知过程,这一过程是“小冰”通过对海量信息的筛选、过滤、整合和建构来实现的,这正类似于人脑对于信息的加工和处理过程,它不是机械式的程序运转,而是充满创造性色彩。

主张机器诗人具有创造性的观点与强人工智能所持的计算功能主义的观点密切相关。这种功能主义认为“精神状态随附于物理状态,因为精神状态是功能状态,它们通过物理状态来实现。”^③也就是说,心智是一种可以脱离肉身的可迁移的功能和状态,它同样可以“随附”于不同的物理载体(如计算机的硅芯片)。这显然是机器诗人小冰制作团队所持的观点。

假如我们承认小冰的诗作具有创造性,“她”的写作过程显示出了某种程度的“拟主体性”,那么这种“主体性”显然也是一种不同于人类诗人的“类主体性”,这种“类主体性”表现为基于大数据的“情感算法”。但从具身视角来看,基于功能主义的小冰是一种数字化的虚拟存在,“她”无法感知人类社会在真实世界中的“悲欢离合”,也无法“领会”受众对于“她”的情感性反馈,这说明小冰可以模仿人类的符号表达形式,却难以像人类一样“理解”故事和

① [美]约翰·塞尔著,杨音莱译:《心、脑与科学》,第28页,上海:上海译文出版社,1991。

② 小冰:《阳光失了玻璃窗·序》,第1-2页,北京:北京联合出版公司,2017。

③ 李建会,赵小军,符征:《计算主义及其理论难题研究》,第82页,北京:中国社会科学出版社,2016。

“体验”情感。

第三,“机器写作”的“元语言”(metalanguage)困境。无论是人脑还是机器,都可视为拥有一套或多套“元语言”构成的认知系统。符号学家赵毅衡认为,“元语言”是“符码的集合”。^①所谓“符码”(code),此处强调的是符号信息编码(encoding)与解码(decoding)的规则。^②对于基于神经网络的机器智能而言,编码可以理解为程序员预先输入的一套规则,而解码则是经过深度学习和反复迭代,最终输出的另一套规则,二者并非线性的对应关系,而是共同构成了一种非线性的复杂动力系统,这往往表现为神经网络算法中的“黑箱效应”,即设计者也无法得知算法是如何推导出结论的。这种对于编码规则的“僭越”被一些人工智能设计师视为AI创意的表征。比如,对于陈楸帆与AI合作完成的科幻小说《恐惧机器》这部作品,其编程设计者王咏刚曾坦言AI写出的很多词句让他自己也惊叹不已。^③那么,这种超出人类设计者预期和意料之外的“解码”是否是机器智能的创意呢?

要回答这一问题,就必须重新审视“机器元语言”与“人类元语言”二者的区别和联系。对于人机交互式写作而言,实际上存在两套不同的元语言。人类设计者的“编码”构成一套元语言,而机器运算的“解码”构成另一套元语言,前者的依据是自然语言符号体系,而后者完全是基于数学模型模拟出来的机器语言符号系统。后者是以前者为基础,不可直接还原成为前者,但又不能脱离前者而独立存在,因而二者既相互关联又不可直接“通约”。

不过,从计算主义的视角来看,人脑犹如一台计算系统,所以当神经网络算法的复杂程度接近或超过人脑之时,它的整体认知能力就能逼近或超越人脑,甚至能够涌现出某种完全独立于人类的“自组织”形态。按照这种强人工智能的思路,所谓的文学灵感不过是一套尚未破解的高级算法。一旦人脑被机器完全破解,机器元语言就可以摆脱人类元语言而自发存在。但德国人工智能科学家克劳斯·美因茨(Klaus Mainzer)认为,即便出现拥有超强计算能力的量子计算机,它仍然会受制于“哥德尔不完备性定理”,因为无论是有机体还是机器都必须遵

从基本的逻辑和数学法则。^④

计算主义视角下的数字计算机的根本目标是要构建一个可与人心媲美的自洽而完备的形式主义系统,但这违背了“哥德尔不完全性定理”。“哥德尔不完备性定理”对于人工智能写作的本体论意义在于,建立在形式主义体系之上的智能写作机器不可能演化成为一种如人心一般完全自发的“自组织”形态,它总会存在不自洽或不完备之处,而最终它依然需要借助人类来协助解决。

这意味着,当机器诗人或机器作家在作品中出现“差错”之时,人工智能专家可从技术层面不断优化算法来最大限度降低出错率,但仍然不可能完全依靠机器自身来“纠错”。换言之,无论机器智能所打造的解码“元语言”体系如何完美,它仍然无法摆脱人类所设定的编码“元语言”这一“第一推动力”而单独存在。因此,严格意义上来说,机器智能所构建的“解码”元语言是基于人类设计师的“编码”元语言基础之上的“类元语言”。所以,机器智能所表现出的那些超出设计者意料之外的“创意”,实际上不能视为机器创意,而是人类利用机器智能所衍生出的创意产品,本质上仍然属于人类创意的变体。

综合以上三点来看,当前的“机器写作”在语言表达、语义理解和元语言三个层面遭遇的困境都指向机器是否具有独立的主体性这一根本症结。从强人工智能来看,主体性可以是脱离肉身和特定环境之上的某种抽象的认知功能。比如AI小冰,她会作诗、画画、唱歌,拥有无数的技能,但她仅仅存在于数字化世界之中。对于那些同样执着于虚拟世界的受众而言,她就是一个与真实主体具有等价身份的“拟主体”。但

① 赵毅衡:《符号学:原理与推演》,第227页,南京:南京大学出版社,2011。

② 胡易容,赵毅衡编:《符号学-传媒学词典》,第72页,南京:南京大学出版社,2012。

③ 南方都市报公众微信号:《人工智能已进入小说的创作》,2019-01-27, https://www.sohu.com/a/291702622_161795

④ Klaus Mainzer. Artificial Intelligence-When do Machines Take Over? Springer, 2020, pp.216-217.

从反计算主义的观点来看,由于机器诗人或机器作家是一种去历史化、去社会化、去身体化的“算法式存在”,它无法像真实的人类一样感知和体验生命,也无法通过日常生活实践获得常识和经验,因此这种“机器写作”只是将文学写作降格成了一种由高级算法所搭建的符号组合游戏,所以它仅仅只是一种交互性的写作工具,不具备任何主体性。

三、“机器写作”的发展进路探究

强人工智能继承了西方“无身认知”的传统观念,认为世界的本质可以还原为脱离于肉身的算法规则,所以人的心智可表征为一套可计算性的功能或状态。由此可以推论,文学创作这类高级认知活动同样可以还原为可操作化的精密算法。按照计算主义的思路,“机器写作”的未来必将走向一种无须依托于肉身的“数字化存在”。^①不过,由于这种数字化的虚拟作者缺乏如人类一样的真实肉身,它往往被视为与主体可直接感知的现实世界相对立的“虚拟客体”。

现代虚拟技术所带来的高度“拟真化”与“沉浸感”不仅在打破传统的虚实界限,也在重塑数字化世界中的主客关系。如果我们承认在算法技术达到一定高度之时,数字化机器作者同样可以具有类似人类一样的拟主体性,那么就等于默认了数字世界的真实性。这一观点的思想基础是虚拟实在论。虚拟实在论的核心主张认为虚拟世界不仅是描述现实世界的技术手段,而且是与其平行的数字化存在。按照虚拟实在论的说法,人工智能机器作者所构建的数字化文学世界同样可以等价于人类作家通过符号化所呈现的心智世界。尽管当前的机器写作离不开人类的协助,但计算主义者坚信,随着技术的不断迭代,在现实世界中被看作“虚拟客体”的机器作者有朝一日终将会在“赛博空间”(Cyberspace)中成为具有自主意识的“数字化主体”。

按照这种演进路线,机器作家或诗人就如同人工智能所创造的虚拟偶像,他们无需像人类作者一样被肉身所“累”,而是可以在虚拟世界中自由变换,拥有无数“化身”。随着超级算法的升级换代,甚至可能出现数字世界中的莎士

比亚与曹雪芹。不过,在传统的一元实在论者看来,基于物理世界的赛博空间可能如同数学中的“极限值”一样无限逼近现实世界,但它终究是一种“拟像”,而不可能完全替代真实世界。

美国哲学家普特南(Hilary Putnam)的思想实验“钵中之脑”正说明了“拟像世界”与真实世界的不可通约的特性。“钵中之脑”假设一个人被切除了身体,他的大脑被盛放在营养液中,并与计算机相连接,计算机可以模拟大脑的信号向他传递“刺激性”的所有信息,让他以为自己仍然生活在一个真实的世界,甚至能“体验”到人类所能感知到的一切。普特南将“钵中之脑”类比为智能机器。而人工智能机器写作所追求的“数字化主体”与“钵中之脑”极为相似,二者都符合“无身认知”的基本纲领,即心智可以独立于身体而存在。

从心身关系角度而言,“钵中之脑”揭示了虚拟认识论存在的一个根本性难题:“钵中之脑”没有根植于现实世界的真实肉身,所以它无法跨越虚拟世界与物理世界。“数字化生存”所面临的尴尬依然在于缺乏可以打通虚实两个世界的肉身。对于人工智能诗人而言,它们永远无法“感悟”文学对于真实世界的意义。因为它所想所说所感受的一切都属于数字化世界中的一部分,它无法跳出数字化世界来化解虚拟世界与实在世界的认知鸿沟。这个难题依然是“无身认知”导致的症结。

但是,极端的虚拟认识论者并不认同“钵中之脑”,在他们看来,由智能机器构筑的“虚拟王国”(Virtual Kingdom)才是整个人类通往永恒的“应许之地”,罗伯特·杰拉齐(Robert M. Geraci)称其为“启示录人工智能”(Apocalyptic AI),“启示录 AI 期望构建一个机械化的未来,在此,人类可以将自己的心智上传到机器并在完美的虚拟身体之中享受虚拟现实的天堂。”^②从这个角度来看,“数字化”的“无身存在”本身就代表了比实在世界更高级的“真实存在”。

① 张怡、酆全民、陈敬全:《虚拟认识论》,第78-79页,上海:学林出版社,2003。

② Robert M. Geraci. "Apocalyptic AI: Religion and the Promise of Artificial Intelligence", *Journal of the American Academy of Religion*(76):1,2008,p.140.

他们认为,机器写作必然沿着“心身分离”的进路持续进化。可以设想,在“虚拟王国”之中,人类读者与机器作者完全是一种超越身体之外的心智交流,虚拟主体代替了物理主体,肉身存在成了一种多余和过时的存在物,自然也就不存在虚实世界的“跨界难题”。

类似于“启示录人工智能”的观点可谓的强人工智能的极端表述,它可被归为“心智上传假说”(The Mind-Upload Hypothesis),该假说的理论根基完全是建立在“无身认知”的基础之上,与此相对立的主张则是“具身心智理论”(Embodied Mind Theory)。^①具身主义思潮与现象学密切相关,梅洛-庞蒂的知觉现象学、海德格尔的存在现象学、唐·尹德的技术现象学都承认“身体”在认知活动中的特殊地位。美国学者德雷福斯(Hubert Dreyfus)是较早运用现象学反对强人工智能的代表人物,他认为强人工智能忽视了身体与经验世界的相互作用,存在难以克服的缺陷。如他所说:“一个躯体化了的主体,才能生存在这个世界上而不必把万物都形式化当作无穷尽的任务。”^②这种具身认知的路径给机器写作的重要启发是必须要考虑到身体与心智的相互作用,而不能一味“去身体化”。^③

根据“具身认知”的主张,机器写作必须抛弃“心身分离”的思路,重新回归“心身合一”。那么,如何能达成这个目标呢?这个问题需要从可应用层面和可探索层面两个视角来分析。从可应用层面来看,构建“人机协作”创作模式具有现实可行性。基于“具身认知”的“人机协作”强调的是让人类主体作为智能机器写作过程中不可或缺的一环,至于在何种程度上参与则需要依据智能技术发展的水平和创作题材的需求来决定。比如,2016年,一位导演与纽约大学的人工智能研究者合作开发了一款人工智能系统,该系统在“消化”一些已有的经典剧本之后,“创作”出了一部名为《太阳之春》(Sunspring)的剧本,该剧本讲述的是一个关于爱与绝望的黑暗科幻故事,有媒体评论它是“古怪的有趣,不可思议的动人”。^④这并非是个案,而是反映出了一种人机交互写作的新趋势。在2019年,作家陈楸帆与人工智能合作完成的科幻小说集《人生算法》也采用了类似的“人

机协作”创作模式。实际上,假如没有设计者所预设的算法规则和挑选的创作素材,人工智能系统将无法独立生成完整的情节或故事。因为只有基于肉身的人类才具有穿梭于实在世界和虚拟世界的“跨层思维”,所以纯数字化的智能写作系统离不开人类这一“第一推动力”。

不过,必须承认,一旦智能写作系统经过持续的算法迭代和优化之后,其生成的故事品质和效率都将会远远超出人类设计者,并形成一种积极的“正反馈”效应。如此来看,未来高级“机器写作”的发展方向是如何平衡“算法思维”的利弊,而不是替代人类作家。因此,可以展望,基于“人机协作”的文学世界不是人类作家与机器作家“一决高下”,而是二者如何通过机器这一技术工具构筑“自然思维+算法思维”的新型“文学生态链”。比较而言,“自然思维”的不足是个体的经验受限,且主观性强,难以标准化,而算法思维的缺陷则表现在机器过于强调逻辑和推理,缺乏真实的情感体验。若能实现“人机协作”,则有助于形成互补共生的人机交互式写作。

当前的“人机协作”的创作模式仍然处于“外化”的初始阶段。从可探索层面来看,其进一步的演进方向则是迈向“人机合一”的“内化”阶段,而后一阶段则要借助人机接口和脑机结合技术的飞跃才可能实现。“人机合一”的发展态势与“启示录人工智能”表面上十分相似,但二者具有本质上的差异。“启示录人工智能”是彻底否认身体与现实世界的关联,并试图将人类的主体性“让渡”给机器所构建的“虚拟主体性”。而“人机合一”是基于具身认知的

① Massimiliano Lorenzo Cappuccio. "Mind-upload. The Ultimate Challenge to the Embodied Mind Theory", *Phenomenology and the Cognitive Sciences*(16), 2017, pp.426-427.

② [美] 休伯特·雷德福斯著,宁春岩译,马希文校:《计算机不能做什么:人工智能的极限》,第262页,北京:生活·读书·新知三联书店,1986。

③ [智]F. 瓦雷拉、[加]E. 汤普森、[美]E. 罗施著,李恒威、李恒熙、王球、于霞译:《具身心智:认知科学和人类经验》,第261-262页,杭州:浙江大学出版社,2010。

④ Jacob Turner. *Robot Rules: Regulating Artificial Intelligence*, Palgrave Macmillan, 2019, p.27.

思路,反对极端的工具理性对于肉身的消解,强调包括智能机器在内的媒介形态只是身体的延伸和强化,而不是否认人类在人机交互关系中的主体性。因此,不管“数字化存在”发展到何等完美的境界,它依然需要依托根植于真实世界中的人类主体才能获得存在的价值。所谓“文学即人学”,人的价值诉求和伦理困境往往超出算法和表征的范畴,它对于个体而言往往是具身化的、情景化的和实践性的价值理性,它属于不可还原为数学模型的生命体验。

综上所述,根据“无身认知”和“具身认知”这两种思路,机器写作的发展进路也可能向两种不同的走向演化。当前的人工智能写作依托的思想根基是基于“无身认知”的强人工智能,它主张智能写作机器无需依托实在世界,而可以通过超级算法构建一种基于“虚拟世界”的“智能体”。在具身认知看来,这种缺乏身体感知能力的数字化“智能体”不可能真正理解文学对于人类世界的独特意义,因为文学创作必须基于真实的生命体验。因此,从“心身合一”的思路来看,构建“人机协作”的交互式写作或许是在可预见的时间范围内具有可行性的方案。若从远景来看,机器智能或许会无限接近人类作家的“创作状态”,但很难从质的层面突破人类心智的“临界值”,唯有将基于肉身的人类心智与机器算法有机结合,才有可能在根本上使人工智能文学创作主体最终超越基于“无身认知”的“算法式存在”。

结 语

当前基于人工智能算法的文学写作与强人工智能所主张的“心身分离观”密切相关。由于忽视了身体在认知活动中的作用,它会遭遇到一系列难以跨越的认知难题。此类难题极大的限制了“机器写作”的发展前景,使其无法仅仅通过优化算法来有效处理复杂的虚构型文体。要避免“无身认知”的困境,就不能只是将“机器写作”视为一种脱离肉身的“数字化存在”,而是要立足于人的主体性来讨论文学虚构与真实的意义。

从具身的角度来看,文学写作是基于主体生命之上的情感体验,因而有必要重新认识肉身体验对于文学写作的内在价值,若抽空了人类创作者在文学创作中应有的位置,势必会将文学写作引向一场主体缺位的“符号组合游戏”。从未来发展趋势来看,人工智能写作将成为作家大脑的媒介延伸或融入主体身体的一部分,但它本质上是服务于人类心智的表达和传播工具,所以机器智能的演化进路绝不是为了“吞噬”人类作家,而是要积极营造具有互补共生特质的人机交互式写作生态链。

【作者简介】黄文虎,文学博士,华侨大学新闻与传播学院讲师,硕士生导师,主要研究方向有智能传播与传播符号学。