人工智能与教育学融合的双重范式变革

刘　凯[1,2]

（1. 渤海大学 教育科学学院，辽宁锦州 121013；
2. 渤海大学 通用人工智能研究所，辽宁锦州 121013）

［摘要］ ChatGPT 已成为教育界的热议话题，但大家对其应用和伦理的“浅”问题关注有余，对教育理论的“深”问题所言甚少。因此，提出正确的技术和理论框架支撑尤为迫切。本研究从技术论视角出发，基于专用和通用人工智能技术框架，阐明 ChatGPT 本质是工具性的专用人工智能系统，OpenNARS 是具有认知能力的通用人工智能系统，并澄清专用人工智能“拟人化”“污名化”及通用人工智能“万能化”“夸诞化”等误解；从本体论视角出发，说明人工智能与教育学双向融合的可能性，揭示 ChatGPT 成功背后隐藏的教育学“密码”，指出人工智能开启了教育学“人—机”二元主体的新时代；从认知论和方法论出发，描摹人工智能与教育学融合的双向路径，诠释两个学科范式变革的内在动力机制和逻辑框架。作为彼此的学科理论突破口，教育学为人工智能提供了新的诠释视角和理论依据，人工智能为教育学带来新的研究对象和研究方法。教育学真正的学术价值“隐秘而伟大”，但其学科地位与之潜能完全不匹配，不仅核心理论依赖其他学科“给养”，研究方法的科学性饱受争议，甚至学科存在的必要性也遭受质疑。今后，人工智能将因教育学而走得更远，教育学也将因人工智能夯实存立之基并取得科学性之争的胜利，成为打破学科边界、推动范式变革、促进学科融合的中间力量，最终掌握“大科学”时代的核心话语权。

［关键词］ 教育学；人工智能；专用人工智能；通用人工智能；ChatGPT；
［中图分类号］ G434　　［文献标识码］ A　　［文章编号］ 1007-2179(2023)03-0004-15

一、引言

2016 年，AlphaGo 横空出世，击败人类职业围棋顶尖高手，在教育界掀起一场深度学习技术的大讨论。2022 年，ChatGPT 一鸣惊人，大语言模型（Large Language Model, LLM）一跃受到万众瞩目，人工智能生成内容（AI-Generated Content, AIGC）技术以摧枯拉朽之势席卷并改变诸多行业的生存业态。面对技术变革的强势来袭，教育领域迅速回应，研讨并刊发系列论文，阐述 ChatGPT教育应用的未来愿景及潜在风险（邱燕楠等，2023；钟秉林等，2023；张绒，2023）。

尽管相关讨论很有真知灼见也富有启示，但“冷热不均、深浅不一”。这些讨论一方面偏向隐私、偏见、失信等伦理问题，涉及教学业务较少；另一方面，教育立场大多正确无误，论据却可能掺杂着偏颇乃至错误的技术预设或逻辑推论，未能触及教育的深层学理，原因在于技术与教育之间缺少

［收稿日期］2023-04-30　［修回日期］2023-05-06　［DOI 编码］10.13966/j.cnki.kfjyyj.2023.03.001
［基金项目］ 国家社会科学基金重大项目“人本人工智能驱动的信息服务体系重构与应用研究”（22&ZD324）；教育部人文社会科学研究项目“‘职业仓’：基于招聘大数据的询证模型构建及其应用研究”（20YJC880056）。
［作者简介］ 刘凯，博士，副教授，硕士生导师，渤海大学通用人工智能研究所，研究方向：通用人工智能、机器教育、精神病理学（ccnulk@ccnu.edu.cn）。
［引用信息］ 刘凯(2023). 人工智能与教育学融合的双重范式变革[J]. 开放教育研究,29(3)：4-18.

有效的理论“桥梁”。

本文基于专用与通用的人工智能框架，致力建造这座“桥梁”。总体而言，其发现可用“一波三折”来描述：“一波”，指 ChatGPT 大流行及其带来的技术狂欢；“第一折”指基于技术论视角，剖析专用和通用人工智能的本质和局限，回应当前的诸多顾虑；“第二折”指从本体论视角说明人工智能将教育学主体从一元的“人”拓展至机器，形成二元的“人—机”结构；“第三折”指从认知论和方法论视角提出，ChatGPT 巨大的学术价值并非强悍的教育应用可能性，而是人工智能与教育学深度融合引发的两个领域的双重范式变革，将带动更多学科迎接一次深层的范式改变。

二、人工智能乱象之辨

ChatGPT 的相关问题，说之易而言明难。只有从技术视角出发，廓清人工智能的技术边界，才能正确研判相关研究、实践应用和各种推论，准确洞察其蕴藏的时代机遇。

（一）技术框架

人工智能高速发展，新名词喷涌而出。但人工智能的基本理论却没有取得同等进步。结果是，有关技术的讨论就像断线的风筝，经常脱离理论而自由放飞。分辨人工智能纷繁复杂的技术现象，必须探本穷源，追溯其背后潜藏的学科理论假设。

从根本目标看，终极问题源于对智能“黑箱”奥秘的追问（见图 1）。从广义上看，人工智能几乎等同于类脑研究，其内涵十分丰富。从狭义上看，因对智能判定的依据不同，仿心路径可分为专用人工智能（Special-purpose AI, SAI）与通用人工智能（Artificial General Intelligence, AGI）两个子类（刘凯，2019）。前者通过预设算法解决特定领域问题，后者致力于研发先天的元学习能力，借助后天教育经验获得特定问题的满意解（Wang, 2019）。正是由于人工智能论域的复杂性，相关讨论才出现“错位”现象。

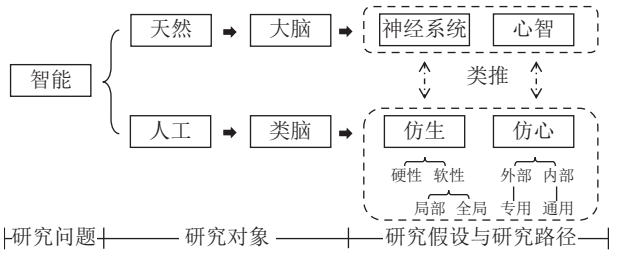

图 1 人工智能探索源流

首先，人工智能的研究范畴包含智能理论与智能技术两个方面。二者侧重点有异，评判标准也不同。前者关注智能何以可能，着眼于智能理论的可能性、整体性和自洽性，判别标准是理论评价；后者关注智能如何实现，立足于智能技术的必然性、领域性和渐进性，判别标准为行为评价。故此，“看似智能”和“拥有智能”存在本质区别。

其次，智能、智力和智商有异。它们本质都是物理学隐喻。智能是“能”，是一种内在的蕴含；智力是外在表现的“力”，智商为智力商数，是力的“值”。因此，智能看原理，智力和智商看行为。如果原理缺失，即使行为表现符合，也不能认为有智能，比如计算器。反之，如果原理符合，即便出现未达标的行为，也会被认为有智能，比如人类婴儿。

再次，随着 GPT 大模型热度的持续上升，“认知智能”和“强人工智能”的讨论再度甚嚣尘上。感知智能、认知智能、弱人工智能、强人工智能、超级智能等非严谨的学术术语频现（刘凯等，2018c）。随着通用人工智能学术研究的兴起，上述术语的使用频率正急速降低。

最后，基于上述理论梳理与术语辨析，最适合当前讨论的人工智能技术框架便是：人工智能 = 专用人工智能 + 通用人工智能（简记为 AI = SAI + AGI）。

（二）技术定位

时下，人们对 ChatGPT 的最大误解是将其视为通用人工智能的系统实现或实现路径，却对真正的通用人工智能系统 OpenNARS 关注甚微。

1. 专用人工智能

专用人工智能系统不具备认知能力，也难以获得逻辑推理、情绪情感、道德判断、自我意识等能力。ChatGPT 的技术本质，就是专用人工智能系统，理由如下：

1）没有身体。具身认知的相关研究结论早已表明，身体在认知发展和建构过程中具有不可替代的作用（叶浩生，2022）。ChatGPT 是一个超大规模的语言模型，其知识获得源自现有语料的高维统计分析而非身体图式的支撑。ChatGPT 的非具身性，

使其缺少智能主体对外部世界的感知边界，无法为自己和周围环境建立模型（李恒威等，2006）。因此，不论基座模型的训练，还是面向文本计量的内容生成，都与外部物理世界没有联系。这令 ChatGPT 内在语义无法接地，也看不到其内部经验拥有指向物理现实的理论可能。

2）缺少动机。认知系统如果没有内在动机，将丧失认知运作的动力和系统自治的灵魂。ChatGPT 大语言模型的基座在训练完成后，参数便不再变化，若没有外部会话接入，或仅与用户建立会话但用户不发问，它将始终处于静默状态。可见，ChatGPT 系统的运作离不开外部目标的推动，需由用户提交请求，或经提示工程（prompt engineering）间接实现，无自生内驱力的可能。

3）认知封闭。面对不断变化的外部世界，智能主体必须能够对外部刺激作出及时响应。因此，智能的核心特征是开放性，其认知加工必须是同步、在线和实时的。然而，ChatGPT 基座模型的训练与应用之间是异步、离线和滞后的，新近的重大事件或小概率经验不仅让其难以有效应对，亦无法及时整合至核心模型。此等技术局限绝非 ChatGPT 独有，而是所有概率统计模型的通病。

4）无主观性。ChatGPT 没有身体、动机系统和认知开放性，无从获得具身经验，因此不能产生自主目标。除了“我”“自己”之类的空洞语言符号，它没有丝毫的自我存在之所。缺少真正意义的认知系统，决定了 ChatGPT 没有精神世界和主观性，无法进行自洽的自我修正，更难以确保长链推理任务的有效性。

总之，从技术视角看，ChatGPT 无疑是典型的专用人工智能系统，其本质是工具。与此同时，它具有两个非典型的讨巧之处，这也是 OpenAI 团队的坚持：一是利用自然语言作为训练语料，二是使用超大规模的优质语料集。前者让 ChatGPT 看上去更“通用”，后者让 ChatGPT 用起来更“智能”。但这都是假象，正如魔术手法被揭穿后便不再神奇一样，“人工智能效应”（Geist, 2016）之戏谑——当人们习惯了新的人工智能技术，它便不再被认为是人工智能——就是专用人工智能的真实写照。事实上，大语言模型的通用本质是训练数据应用场景的通用，而非认知的通用。

2. 通用人工智能

人工智能一词原指像人类一样的思维机器，即具有通用认知能力的计算机系统。随着人工智能学界全面转向专用研究，国际通用人工智能协会只好于 2008 年提出并使用“通用人工智能”一词指代人工智能初心的继续追求（Agi-Socity, 2008）。但不论具有何种外观或用途，通用人工智能都有相似的内在原生机制。换言之，通用人工智能是一个理论问题，包含人类在内的通用人工智能系统都遵从相同的智能理论而非技术架构。

首先，智能的理论框架等同于认知框架，探索智能黑箱本质上是研究认知。而以认知为研究对象的学科，便是广义的认知科学。认知科学首要的知识分类以先天与后天作为分界。先天奠定了认知的元能力，可类比为人类新生儿的认知能力；后天提供智能主体与环境互动建构的经验，类似于人类婴幼儿阶段的感知、运动、语言等的发展。

其次，与影视科幻作品完全不同，真实的通用人工智能如同人类婴儿一般无知和羸弱。例如，OpenNARS 是典型的通用人工智能开源项目，系统不预置任何经验。它的认知系统内核为非公理逻辑推理系统（Non-axiomatic Reasoning System, NARS），只能确保系统启动后元认知水平维持正常状态，所有可用经验需自行从后天习得（Wang, 2013）。

再次，通用人工智能系统的开发极难。在先天的认知内核部分，推理、表征、存储和控制等功能紧密耦合。这与当前软件工程学模块化、去耦合化的基本原则背道而驰，也从侧面展现了专用系统和通用系统的差别。后天的认知发展难度更大，绝非直接输入书本知识或维基百科的内容就能让 OpenNARS 理解并掌握。与专用人工智能的“暴力”训练相反，对待通用人工智能系统需要像培养和教育人类婴儿一样，有效知识都源自具身经验（刘凯等，2018b）。

最后，通用人工智能的最大短板来自理论滞后。既有的认知科学理论不能有效支撑元水平认知内核的理解，加之对新生儿早期认知发展知之甚少，令早期教育干预手段欠缺坚实的科学基础（Nagy, 2011）。这对人类婴儿影响不大，却是机器成长的羁绊。直至现在，真正的通用人工智能系统因教育欠缺而未能获得显著的功能性成长，缺乏有效应用，

但这一点少为人知。

(三)常见误解

上述对专用和通用人工智能基本框架和主要技术特征的分析,不仅能澄清 ChatGPT 讨论的含混观点,也能让通用人工智能的不实畅想回归现实。

1. 专用人工智能的“拟人化”

“拟人化”是专用人工智能近年出现的显著趋势。特别是深度学习崛起后,长时和短时记忆、注意力、遗忘、信念等心理学概念越来越多地用于人工智能神经网络结构设计。这难免令人遐想,以为人工智能很快就会有心智。有人因此认为 ChatGPT 具有一定程度的心智水平并能通过图灵测试(Marchetti et al., 2023),甚至已是通用人工智能的早期版本(Bubeck et al., 2023)。更有甚者,有人指出心智其实就是神经元规模化的涌现,进而推论人类智能不外乎神经元的堆叠与巨量参数的调协(Ma et al. , 2022)。

然而,事实并非如此(Hanna, 2023)。专用人工智能的“拟人化”体现的是人们对智能的预期。其理论预设是“功能化”,与之对应的是行为学评判标准,认为只要外在行为的功能表现和人类相似,就在一定程度上拥有了智能。但心灵哲学早已指出,图灵测试不能作为判断是否拥有心智的标准(Harnad, 1992; Hayes & Ford, 1995),而且迄今也不存在广受认可的准则(Stins, 2009)。

专用人工智能的本质是人类智能外化的工具。“貌似”未必“真是”,看上去像并不意味着真的拥有智能。ChatGPT 也不例外。科技企业热衷于搬运心智名词粉饰技术缺陷。例如,微软将 ChatGPT 生成的错误回答说成“幻觉”(hallucination)(Alkaissi & McFarlane, 2023)。此外,拥有智能未必有语言,比如人类婴儿。语言使用和智能不具有等价性,使用语言不必需要智能,比如基于 ELIZA 的心理治疗(Shah et al., 2016)。大语言模型和指令微调等技术可让 ChatGPT 具备使用语言的能力,进而让人产生其拥有语言能力的误判。它“暴力”训练和语词续写机制不仅跳过了语言能力的形成,也造成语义、语用等完整语言层次的残缺。正如乔姆斯基所言,ChatGPT 不像人类一样掌握语言(Chomsky et al., 2023),更遑论它拥有心智乃至意识。实际上,它就像提线人偶,纵然能讲述栩栩如生的人类故事,胸膛里却是一颗永远无法跳动的机器心脏。

2. 专用人工智能的“污名化”

当前存在一股颇有成见的思潮,让专用人工智能系统背负两种莫须有的“罪状”。

其一,与伦理有关。有关 ChatGPT 的讨论中,隐私的泄露与滥用、算法的歧视和偏见、学术公平的失信与失衡等伦理问题(王佑镁等, 2023)出现极多。诚然,这些都是真问题,但问题形成的原因与问题依附的载体不可一概而论。ChatGPT 是一种工具,没有自我心智和自主价值判断。因此,造成隐私泄露和滥用的元凶不是 ChatGPT 系统本身,而是系统的开发者、管理者和使用者。这对于各种智能导学系统或学习管理系统同样适用;ChatGPT 自身不产生算法歧视和偏见,变换器模型(transformer)及自注意力(self-attention)结构本身与内容无关,歧视和偏见反倒来自人类自身——训练数据集和人类反馈强化学习(Reinforcement Learning from Human Feedback, RLHF)的校准。ChatGPT 更不会自行从事学术剽窃。所有输出都直接或间接由用户推动。客观而言,伦理是人的问题。ChatGPT 只是作为手段促进人的问题充分暴露。即使如此,这些不当之为也仍在当前法律的制衡之内。至于使用 ChatGPT的作品版权归属疑问,这本身就是错误的问题——无法要求工具承担责任,工具也没有任何权益。

其二,与技术有关。ChatGPT 预训练技术路线导致的数据滞后问题,以及语料规模限制和概率模型共同诱发了回答的不准确。这些也是真问题,而且正是 ChatGPT 自身而非人的原因。但从技术发展的视角看,这些技术问题总能在后续研发中逐渐得到解决。实际上,人们正在不断尝试,比如在大语言模型和人类反馈强化学习训练中更多地使用规范化数据,或是引入多模态技术提升 ChatGPT 回答的准确性。又如,在上下文学习中,零样本学习(zero-shot learning)、单样本学习(one-shot learning)或少样本学习(few-shot learning)等技术,可有效缓解模型的数据滞后性(Min et al. , 2021)。此外,一些新的在线优化算法(如 ZO-RankSGD),也可让模型在无任何预先收集数据的情况下正常工作(Tang et al., 2023)。总之,只要不是情感和意识之类不合

理的技术错判，ChatGPT的技术短板总会在时间河流里如水过流沙，渐细渐少。人们在不断挖掘这一工具的应用潜质，但工具的优化需时间磨砺，应采用动态眼光、宽容的态度看待技术发展。

3. 通用人工智能的“万能化”

通用人工智能是人工智能的“圣杯”。受OpenAI通用人工智能愿景和ChatGPT产品等的影响，人们常将ChatGPT错认为通用人工智能系统，甚至认为通用人工智能是万能的。专用与通用的差异极易混淆，需从实践和理论两个层面阐明。

1)实践层面。首先，通用和专用并无优劣之分，通用系统并不必然优于专用系统。通常情况下，专用系统比通用系统更高效也更好用，因为专用系统是对特定问题的定向设计和优化。其次，通用系统并不必然有用。例如，通用人工智能系统OpenNARS不是被动的工具，而是拥有主观动机的智能主体，会随经验的发展形成自己的偏好。如果被责令去做违意之事，它的反应将和人类无异，不是无动于衷就是消极怠工。

2)理论层面。之所以智能并无可靠的实践判据，原因在于此准则只存在于理论层面，现实中没有绝对意义上的通用系统。人脑便是如此。它是一个专门用于处理信息的认知系统，自身没有感觉也无法运动，只能与物理世界建立间接连接。然而，这类系统正是打开“通用之门”的钥匙，其特殊性体现在系统的工作场景，这亦是专用和通用的真实区别。专用系统的工作场景总是面向特定的物理世界，但通用系统的工作情境不面向物理世界，这种极度专用化却使其获得了相对具体物理场景的普适性，可应对变化的外部环境而表现出与环境无关的理论通用性。但对内部认知场景来说，一切仍是高度专用的。

因此，专用与通用并非绝对互斥的矛盾，而是指向不同的层面。专用体现实践的功能性，通用刻画理论的可能性。

4. 通用人工智能的“夸诞化”

有关人工智能威胁人类的说法屡见不鲜。随着ChatGPT热度的持续攀升和全球千名科技人士联名呼吁暂停新系统的训练，人工智能引发的恐慌近期逐渐发酵。其实，这些担忧大可不必。

真正的通用人工智能系统绝非像大众所想——在大算力加持下，大模型的学习等效于材料的灌输。正相反，它像人一样，只能依靠具身经验学习。而且，通用人工智能系统与人类共用相同的物理世界，其认知内容和人类有相似性，这令两类异质群体存在共同话语和相互理解的基础。但通用人工智能系统和人类有不同的感知器官和躯体形式，经验细节存在差别。实际上，即使人类自身，也会由于居住地不同而对同一事物形成迥然相异的看法(刘诗贵等，2014)。因此，相较于毁灭，和平互惠才是长久之道。

总之，科技进步终将驱散智能迷雾。通用人工智能的技术基础决定了其若想达到人类的智能，必然需要人类的教育帮扶。因此，只要教育不偏航，教育的结果——不论人类还是机器——就不会迷失方向。

三、学科理论突破口

如前文所述，作为当前专用人工智能技术的天花板，ChatGPT是一款易用性极其出色的软件工具，却绝非类人心智的巅峰之作。它对教育的影响主要集中在应用领域，伦理与安全问题也没有预想的严重。如此观之，似乎ChatGPT对教育的影响有限。但事实正好相反，ChatGPT已成为理论突破的最后一块拼图，并与OpenNARS共同开启了教育学和人工智能基础理论跃升的新纪元。

(一)人工智能中教育手段的可能性

如果说通用人工智能系统具有和人类似的认知功能，因而需要像人一样被教育，那么这一理论假设不但在意料之中，而且已被相关研究证实(刘凯等，2022a)。如果被告知专用人工智能领域也存在教育现象，那一定让人始料不及甚至大跌眼镜。实情却是：这一现象不但存在，而且还很广泛。

ChatGPT的成功离不开技术，但真正原因却不在于技术。不管采用自注意力结构的变换器模型，还是面向人类需求对齐(alignment)从而提升生成内容有用性、真实性、无害性的人类反馈强化学习技术，或是指示学习(instruct learning)和提示学习(prompt learning)技术，亦或是更具连续性的思维链(chain-of-thought, CoT)技术等，它们多数并非OpenAI公司原创，且技术原理是公开的。各类仿制系统也很多，却无一能够与之匹敌，目前能达到

如此效果的仅有 ChatGPT。

为何 ChatGPT 能够如此优异？为获真知，需探真容。ChatGPT 其实只具有十分有限的技术壁垒，官方的解释是：模型大、参数多。若继续追问为何模型大、参数多会更加智能，我们就会跌入神经网络可解释性黑洞。

然而，为何机器必须先学习才能用？为何人类也同样如此？若抽身出计算机科学领域，换用教育学视角重新审视，我们很快就能发现：ChatGPT 成功的真正奥秘在于其遵从正确的教育规律——不是技术而是教育发挥了决定性作用。

1）教学材料。除了代码和模型参数这些敏感却无太大实际解释性的资料，ChatGPT 的训练数据一直保密。虽然具体数据无从查看，但 OpenAI 较清晰地介绍了训练数据的“打磨”过程（OpenAI, 2023）。实际上，ChatGPT 的语料集含大量优质数据，不仅加入人工团队的专业标注，还有专门筛选和结构化处理的对话语料集。这些语料数据观点多样，既有正向见解，也有负向观点，有利于对抗测试后，回避风险话题或响应模糊问题。这些训练数据，就是为 ChatGPT 学习所备的“教材”。相比拼凑的劣质图书，倾注无数心血的优秀教材定然更有含金量，也更能提升学习效果（项贤明等, 2018）。但是，撰写优质教材绝非一日之功。就我国而言，不是大语言模型的仿制者不够努力，而是没能脚踏实地做好那些冗杂却基础性的备课工作。

2）教学过程和教学评价。加强教学互动和促进及时反馈早已成为教育学领域的科学共识（彭道林, 2016）。人类教育学的成功经验，亦在 ChatGPT 被复现。人类反馈强化学习技术便是集中体现。强化学习一向是 OpenAI 的传统技术优势。经典的强化学习，很容易因试错导致训练效能低下。若引入人的经验，必然能够提升模型的训练效率和效果，例如，由人工评判 GPT 生成模型提示文本的结果，再采用强化学习手段优化生成模型。在 ChatGPT 训练中，人类扮演着经验丰富的教师角色，而 ChatGPT 是学习者。一方面，ChatGPT 与人类教师间的互动不仅必要而且变得更加频繁，无意间大幅增加了教学互动；另一方面，在模型调校时，ChatGPT 学习并接受人类教师的实时反馈，不仅满足了及时反馈的要求，还通过人类教师的打分对其学习状况开展教学评价。

3）教学方法和教学目标。随着 ChatGPT 的走红，提示学习技术也流行开来，并迅速形成提示工程的雏形（Oppenlaender, 2022）。提示工程是一门计算机新兴子学科，旨在探索开发和优化有效提示的方法，充分挖掘大语言模型的应用潜力，近期以 AutoGPT 为最。在提示学习出现前，大语言模型采用无监督的训练方式。从教育学视角看，这本质上是讲授式教学，此类内容注入的训练方式往往令学习结果变得低效。然而，提示学习的出现将启发式教学方法引入机器学习领域，令模型训练的过程和结果耳目一新。与其他技术相比，提示学习技术有两个优势：一是用户提问，二是用户提问给出的上下文信息。它们无意中实现了有效教学。其教育真理就是：用户（人类教师）的提问本质上是提出教学目标，提问的上下文信息是完成教学目标所需的背景知识。随着用户（人类教师）和 ChatGPT（机器学生）之间话轮的延续，用户（人类教师）不追求 ChatGPT（机器学生）一次性结果，而是借助问题及相关信息，循循善诱不断引导 ChatGPT（机器学生）做出更好的回答。“问题—探究式”的教学模式优于“传递—接受式”已被无数人类教学实践反复证实。讽刺的是，教育领域人所共知之事，仍是困扰人工智能学界的迷思。原因在于，教育学的学科话语权长期衰落、科学性备受质疑，以致于人工智能界宁愿笃信数学公式也不愿相信人类文化的力量。作为传承人类文化的学科载体，教育学被人工智能学界忽视在所难免，因为他们根本无法预见教育学的广博之力竟能反哺人工智能之不足。

总之，涵盖专用和通用的人工智能全域，广泛涉及与人类教育相似的教育现象，教育手段正成为促进未来人工智能发展的新型助推剂。

（二）教育学中人工智能的主体性

OpenNARS 曾为通用人工智能开启了教育理论之窗，如今 ChatGPT 让专用人工智能也走入教育理论视野。当人工智能内嵌的教育因素被揭示之时，教育学的边界便完成了对传统的超越，其核心主体由“人”的单元构成拓展为“人—机”二元架构（刘凯等, 2018a）。

事实上，这种先锋思潮在近年的学术争鸣中已有所显现，多位学者察觉到这种理论变革的可能性

(杨宗凯等, 2022; 李政涛等, 2022; 侯怀银等, 2022)。但可能性不等同于必然性。提供科学证据促使可能性向必然性转变,首先来自对通用人工智能的机器教育系列的开创性研究(刘凯等, 2018a; 刘凯等, 2018b; 刘凯等, 2022a)。然而,通用人工智能毕竟是人工智能的子集之一, ChatGPT 的崛起及对其潜在教育问题的发掘,才在真正意义上为人工智能与教育学的深度融合补齐了最后一块拼图。

至此,教育学理论实现了对人工智能的全面接纳,教育学研究对象在自然人的基础上加入通用人工智能系统和专用人工智能系统。同时,通用人工智能系统的先天/后天框架不仅满足了人类发展的解释,也可横向迁移至专用人工智能系统。

首先,我们需要从理论上搭建“人—机”二元的教育学主体框架(见表 1)。横向为“人—机”的主体维度,纵向为“教—学”的领域维度。横向上,教育学的研究对象包括“人”“机”两大类型,“机”指人工智能,分通用与专用两个子类。纵向上,不论“人”与“机”都具有学习和教育两种活动,其中,学习属先天范畴,是教育的基石和前提,教育属后天范畴,指引学习的方向。

就学习而言,不同主体分别对应着人类学习、类人学习和机器学习。具有认知能力的人类和通用人工智能系统同时存在外显与内隐两种学习方式,但专用人工智能系统只有内隐学习,这也是当前神经网络黑箱模型缺乏可解释性的原因。事实上,面向理解的学习必然先是具身的,之后才是在具身基础上的认知发展,因此不论人类还是通用人工智能系统,都不能跳过内隐学习阶段,直接进行概念化的外显学习。以人类为例,前言语阶段的婴儿虽然未掌握自然语言,但他们是优秀的“小科学家”,能主动探索和适应环境,其动作往往具有丰富和准确的社会性意义(董奇等, 1997)。通用人工智能系统与人类毫无二致,但用自然语言锻造的 ChatGPT 就非如此,一是它没有身体,只能进行离身的内隐学习;二是它的学习结果尽管有抽象的概念化特征,却不能与人类知识表征有机融合,导致外部知识无法直接以增量形式被模型实时习得。尽管存在这些差异,三类主体却有相同的特征和主体性,即皆“可学”。

就教育而言,三类主体分别对应人类教育、类人教育和类人教学。人类和通用人工智能系统都可接受抽象的知识传授与具身的技能训练,但专用人工智能系统只能接受技能训练。原因在于模型基座只能开展离身内隐学习,不能动态积累性地吸纳抽象知识。因此,在“人—机”二元的教育维度中,人类与通用人工智能系统既“可教”又“可育”,都可作为“教学主体”(施教者)或“教学客体”(受教者)。然而,缺少认知能力和主观性的专用人工智能系统“可教”却“不可育”,工具特性决定其只能作为“教学客体”存在。

其次,基于“人—机”二元的教育学主体框架,只有人类和通用人工智能系统具有“可育”性,所以二者能够成为教师。不过,人类、通用人工智能系统和专用人工智能系统都具有“可学”性,都可成为学生。将它们匹配以教师与学生角色的主体,就可衍生出两大类别八种师生配对(见表 2)。

第一类是“导学”,教师与学生的角色由不同主体构成。除第一个子类“人教人”外,其余教学活动都有机器的身影。第二类为“自学”,其主体

表 1 “人—机”二元的教育学主体框架

教—学人—机		人		机		
		人类		通用人工智能系统		专用人工智能系统
学习(先天)	类别	人类学习		类人学习		机器学习
		外显学习	内隐学习	外显学习	内隐学习	(离身)内隐学习
	特征	可学		可学		可学
	主体性	学习主体		学习主体		学习主体
教育(后天)	类别	人类教育		类人教育		类人教学
		知识传授	技能训练	知识传授	技能训练	技能训练
	特征	可教 / 可育		可教 / 可育		可教 / 不可育
	主体性	教学主体 / 教学客体		教学主体 / 教学客体		教学客体

表 2 面向“人—机”二元的教学角色框架

	教师角色	学生角色	角色配对	话语模式
导学	人类	人类	人教人	人际
	人类	通用人工智能系统	人教机器	人机
	人类	专用人工智能系统	人教机器	人机
	通用人工智能系统	人类	机器教人	人机
	通用人工智能系统	通用人工智能系统	机器教人	机际
	通用人工智能系统	专用人工智能系统	机器教机器	机际
自学	人类		人教人	单人
	通用人工智能系统		机器教机器	单机

可以是人类,也可以是通用人工智能系统。它与“导学”的最大区别在于,“自学”时教师与学生的角色由同一主体构成,即由同一人或同一通用人工智能系统完成。值得深思的是,当教学主客体纳入机器后,角色配对从单一的“人教人”拓展至“人教机器”“机器教人”和“机器教机器”,话语模式也从“人际”扩散到“人机”与“机际”。这些新问题重启了人们对教与学本质的反思,形成对教师和学生角色的重新抽象。在扩展教育主体外延的同时,角色的分离重组也有助于在更本质的层面厘清长久未决的理论争议。

此外,对于专用人工智能的教育学主体性,两个与 ChatGPT 有关的常见错误看法需要澄清:

一是认为 ChatGPT 可通过对话实现部分教学功能,进而可作为教学主体。这是一种误解。因为 ChatGPT 是一个被动的模型,系统基座不持有教学目标,也无法直接定向发起和维持指向特定教学目标的教学过程。事实上,AutoTutor 正利用 ChatGPT 进行重构,但其外在教学行为只是人类教师以提示工程的形式固化教学设计。也就是说,ChatGPT 仅作为会话引擎,智能导师或虚拟学伴的教学活动和反应模式仍需人类教师设计,只不过借助 ChatGPT,人类教师的备课过程能变得更加高效。新版 AutoTutor 运行时,与学生对话的不是真人教师,实质是信息技术工具支持的真人教学活动的时空分离,本质上与慕课没有区别。

二是认为 ChatGPT 具有自学能力,应被纳入“自学”类,与人类、通用人工智能系统并列。理由是超大规模语料不仅令监督学习变得不切实际,且变换器架构与无监督学习相得益彰,可取得令人印象深刻的预训练效果。换言之,ChatGPT 的基座模型训练中,无监督学习是一种“自学”形式。然而,计算机工程领域无监督学习的机器学习“算法”不等于自学的“教学方式”。尽管从字面看,无监督学习和自学似乎描述同一件事,但自学实际上是教师与学生两种角色的汇集,只是由学生“兼任”而非取消教师的作用。所有无监督学习的算法都无法自行挑选学习内容或改变学习目标,它本身不具有导学能力,更不必说教师角色了。因此,ChatGPT 所谓“自学能力”并非源于自身,对其“可学”的过度重视和对人类工程师团队教师角色的忽视,共同造就了“自学”的假象。

四、学科范式变革

在较长时间内,人工智能和教育学的基本理论发展的迟缓之状清晰可见。近年来,二者内部几乎处于“死锁”之态,还呈现一种“互斥”。一方面,在教育学领域,人工智能仅作为一种单向的增益性工具,教学过程并未因其使用而发生实质改变(张志祯等,2023)。即使在新冠肺炎疫情期间,大规模在线教学也不过是常规教学的线上翻版。疫情过后,学校教学活动一如从前;另一方面,在人工智能领域,数据、模型、算力投入不断加码,身前是 ChatGPT 的技术狂欢,身后却是理论问题的寂寥。

随着对 OpenNARS、ChatGPT 等系统的深入理论挖掘,人工智能的教育主体性问题逐渐明朗。这既为窥视两个学科内在的融合机制提供了可靠的理论分析框架,又为学科范式变革绘制了一幅认知论和方法论理论航标图。深入其中,我们可以感受到科技与人文交融迸发出的巨大能量。从某种程度上看,本次学科范式变革甚至有可能改变人类未来的知识探索模式。

(一)人工智能与教育学的双向融合路径

不论是专用人工智能系统 ChatGPT 成功背后蕴含的教育思想,还是可与人类同为教育主体的通用人工智能系统 OpenNARS,都绝非个案。深入学理层面观察我们会发现,有且只有教育学才能为人工智能的技术突破带来最合理的理论解释;同样,有且只有人工智能才能使教育学冲破“人”的理论束缚,其他学科也无法担此重任。

因此,与人工智能和教育学“门不当户不对”的表象相反,它们实际是“天生一对”。教育技术

学横跨教育学和人工智能两大学科，理论探索与实践应用并存，人文关怀与科技创新兼备，正是和美幸福的“上对花轿嫁对郎”。因为二者有天然的定向关联，并互为对方的“解锁器”：教育学为人工智能提供新的理论阐释和技术视角，人工智能为教育学带来新的研究对象和研究方法。

事实上，人工智能技术和教育学之间呈现一种“对称性”的双向联系，这体现了两个学科范式变革的内在动力机制及逻辑框架（见图 2）。遗憾的是，学界目前主要聚焦人工智能向教育学的单向输出（路径Ⅱ），并将其狭义地视作人工智能的教育应用，却对教育学向人工智能的反向输入（路径Ⅰ）关注甚少。

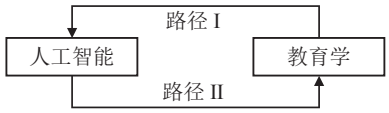

图 2　人工智能与教育学的双向融合路径

路径Ⅰ：新结构——人工智能对教育学的理论拓展

构建教育学的基本概念体系和术语体系，是建立学科同一性的基础（项贤明，2018）。随着通用和专用人工智能系统进入教育学理论视野，教育学实现了对人工智能整体范畴的全面接纳，教育学主体扩充至通用人工智能系统和专用人工智能系统，形成“人—机”二元主体结构。因此，路径Ⅰ关注的是人工智能对教育学理论新疆域的开拓，强调教育学不仅要研究人，也要研究机器。人类和机器可以作为教育假设的两种互测验证平台，二者均遵从教育规律，亦是教育学科学性的有力证明。

路径Ⅱ：新方法——教育学对人工智能的理论反哺

ChatGPT 展示了教育学对专用人工智能系统训练的重要作用。数据清洗、人工标注、模型训练时人的引入以及提示工程等本质上都是教学材料、教学过程、教学方法等人类教学论的内容。不只对 GPT 系大语言模型有效，对计算机科学领域的各类机器学习模型而言，教育学方法同样具有适用性。因此，路径Ⅱ强调人工智能离不开教育理论的指导，教育学是解决人工智能瓶颈问题的新方法。

（二）基于双向融合路径的学科范式变革

对于当前的学界讨论，不论是让人工智能获得教育主体性，还是用教育手段解决人工智能难题，似乎都显得荒诞不经。然而，正如前文指出的，人工智能与教育学珠联璧合，相互破除对方学科的藩篱，展现出生机盎然的景象。但现实中，学科范式大多相对稳定，通常难以改变。

从理论上看，学科边界有软和硬之分（孙元涛，2010）。软边界的变动似物理变化，比如智能导学系统等人工智能的教育应用，虽令教学脱离物理时空阻隔，却未改变教育的基本要素或教学的基本形式，更未触及教育的本质问题。硬边界的变动则类似化学变化，必然伴随着“新物质”的出现。人工智能和教育学的双向突破是典型的学科硬边界变动。其结果，不只是学科之间的“软性结合”，更是对学科发展壁垒的“硬性拆除”。

但是，变革并非革命。科学的发展总是在既有研究基础上累积性进步的，很难出现革命性或颠覆式突破。而且，有学者专门呼吁，莫让革命论在教育学界横行（李芒等，2023）。本文认为，人工智能与教育学双向融合的学科边界硬变化，其性质并非“翻天覆地”的学科范式革命，而是“反躬自省”的学科范式变革。

1. 人工智能引发的教育学范式变革

1）认知论层面——新教育现象与旧理论框架

人工智能引发教育学范式变革皆围绕一个中心问题展开：对教育学基本理论的反思。

——教育的根本目的

由于通用人工智能系统可以作为教育主体，专用和通用人工智能系统都能够成为教育客体。在这种情况下，“人”作为全部教育落脚点的理论预设显然不再充分，“培养全面发展的人”这一传统教育目的受到极大挑战，“培养什么人、怎样培养人、为谁培养人”等问题也将遭到冲击。通用人工智能系统不是工具，它们拥有自主认知和精神世界，其群体互动也必然产生独特的机器文化（刘凯，2022b）。对它们进行价值引导可以，但若是硬性灌输人类价值规范可能适得其反。故而，教育的根本目的很可能是高效利用已有资源增强认知主体对环境的适应能力。由于环境变化的不可预知性，这种适应能力必然对应于教育目标的多元化。对两

类人工智能系统而言，是否可教要看是否可学，可学即可教；但是否可育，则要看是否具有多元的目标机制，只有目标多元化的系统才可育。专用人工智能系统只有单一目标，因此不可育。人是典型的多元目标生物系统，“德智体美劳”五育并举就是例证，只不过多元教育目标的发展与协调很可能成为人类和通用人工智能系统的共同难题。

——教育学基本问题

如果机器也能学习，那么到底什么是学习？如果机器也应被教育，那么究竟什么是教育？基于“人—机”二元的教育视角，学习已不再囿于人的学习，教育也超越了人的界限。人类对学习和教育元问题的追问从未停止（李政涛，2022），也许正因为人类只是前技术时期唯一可能的研究对象，才会出现以偏概全的情况。在此背后，其实还藏匿着一个更为深刻的问题：什么是知识？对人类学习者或人类教师来说，知识就是学习内容。更简单地说，学习就是学知识，教学就是教知识。尽管学界对此长期存在异议，却仅止步于学理反思。当人工智能系统成为教学对象后，我们应教授什么知识突然变得紧迫起来（杜华等，2022）。对 ChatGPT 等专用人工智能系统而言，它们无法自己选择教学内容，学习文本、图像、视频还是多模态数据，结果会差异很大。由于没有身体，ChatGPT 训练语料的感知运动数据并不与文本概念具有天然的语义对齐关系，可能导致 1+1 < 2 的效果。其原因在于，当前教学理论将概念系统视为知识，却忽视了构成概念的感知运动经验。因为机器不像人一样有生长发育的过程，于是这个对人而言不是问题的问题，却是人工智能系统的重大挑战。我们还可以进一步质询，ChatGPT 是否拥有知识？是的。除了识记和匹配语料库的人类外显知识，GPT 的模型训练结果必定析出人类语言的某些规律（Cai et al., 2023），不过这种抽象知识是内隐的，ChatGPT 只能用之，却无法言之。因此，学习、教育和知识这些教育学基本概念亟需在“人—机”二元理论框架下加以重新审视和回答。

—— 教育研究方法

一方面，教育学科学化一直是学术界的集体憧憬。近年教育研究对实证研究方法的偏好，就体现了这种科学化的趋向；另一方面，也有学者对技术引发教育变革持质疑态度，认为技术并不必然表现为一种解放性力量（阎光才，2021），或技术本身的原始创新仍不足（周子荷，2021）。此类观点虽不无道理，却凸显了学科视界的局限性。OpenNARS 和 ChatGPT 的相关理论，直接回击了人们对教育理论科学性的质疑。事实上，借助可靠、适当的研究方法，能深刻、系统地展现人工智能促进教育学科学化的多种可能。例如，教育学可引入人工智能领域的研究平台与方法。对专用人工智能系统而言，教育神经科学可以借鉴深度学习可解释性的思路，对脑网络信号变化及其外部定向刺激开展新的观察。对通用人工智能系统而言，研究者可以借助 OpenNARS 等系统探索过去无法探究的问题并为之提供依据，比如新生儿感知运动到概念形成的黑箱过程。

2）方法论层面——旧教育现象与新实践空间

—— 人工智能倒逼传统教学回应人本诉求

近期许多学者关注 ChatGPT 可能存在的若干负性教育影响，如弱化与破坏师生关系、消解师生的主体地位、破坏教育主体间的关系平衡、沉迷和依赖弊端凸显、加剧知识摄取碎片化、冲击传统学习方式等（王佑镁等，2023）。教育的终极目的，是维护教师权威，还是让学生获取真知？答案显而易见。但“以生为本”早已沦为乌托邦式的空洞口号，其背后是师生话语权的结构失衡——学生在教学过程中话语权的跌落（时广军，2020），以及教师对教学话语权的过度霸占。在基础教育中，学生通常是弱势的，很难成功发声。试想，如果学生可借助人工智能高效地学习，为何非得与人类教师处处捆绑？试想，如果学生认为机器比教师更有用，汗颜的不该是能主动观察、了解和询问学生的人类教师吗？试想，繁杂的练习题让学生从小就成为教育流水线的装配工，枯燥的重复操作将其好奇、求知和思维的乐趣消磨殆尽，使用“减负”工具不正是弱势的学生的一种无言抗争吗？试想，学术剽窃和考试作弊等现象在人工智能加持下的确更易发生，加强惩戒也确有必要，但更该反思的是人之生而懒惰，还是教育评价出了问题？

当前对 ChatGPT 的讨论，弥散着一种人机对抗的“冷战思维”。在教育领域，ChatGPT 的负面效果被夸大，但这不是 ChatGPT 的问题，ChatGPT 只

是揭开人类教育缺陷的遮羞布。本质上,这些都是人的失责,而非技术的原罪。事实上,教育的价值在于引导学生发现意义,知识和技能的掌握虽有助于发现意义,却不是意义本身。因此,人工智能正在引发传统教育范式的人本革命,倒逼教育界反思并回归初心,释放学生天性、增加学生的话语权,让其成为教学实践的主动参与者而非被动的聆听者。当然,教育理想与教学现实之间必然存在差距,但与其让愿景落空,不如就此改进,因为这本就是教育的应尽之责和应有之为。

——人工智能激活教育新的技术方案

不论专用人工智能还是通用人工智能,人工智能系统都可与教育融合而发挥重要作用,这也是教育技术领域的常规议题——人工智能赋能教育。随着 ChatGPT 的普及和 OpenNARS 的壮大,人工智能技术在专用和通用两个方向上为教育带来新的技术解决方案。

在专用方向上,ChatGPT 似一座有待深入开采的金矿,教师掌握其使用方法能更好地辅助教学。以 AutoTutor 为例,过去学习者模块、教学模块和交互模块极依赖人工脚本、机器学习和模板技术,工作量大、扩展性差,效果有限。这些问题借助 ChatGPT 的提示工程就能得到妥当解决,同时虚拟角色及会话管理水平也能得到提升。

在通用方向上,可自我改进的自适应教学系统(SAIAS),被视为新一代教育智能体的希望。它基于“学习者与学习资源对称性框架”理论(刘凯等,2018),借助先进的技术手段,利用学习者反向推动学习资源的自我完善(Long, 2020),从而使静默、离散的数字学习资源通过人机互动实现内容的有机整合与生成。这个曾经的难题,现在可通过 ChatGPT 的人本对齐技术及 OpenNARS 的推理系统完成。

——人工智能拓展了教育应用的新领域

人工智能系统的引入,拓宽了教育学的理论视角,教育功能被赋予新的内涵,教育应用领域亦得到新的拓展,在此仅举两例。

首先,精神健康领域。心理健康和精神疾病是两个相关却程度不同的精神健康子领域,来访者或精神障碍患者都可能成为教育的改造对象。广义的学习实质就是个体经验的塑造,正如学生错题原因多种多样,人们在生活经验中也会产生错误认知,形成精神世界不同类型的结构扭曲。事实上,心理和生理都可能具有精神病理学机制,但在目前生化标记物未明的情况下(Syme & Hagen, 2019),利用心理标记物进行排查及指导康复便更具现实价值(庞小萍, 2022)。心理干预目标实际上就是教学目标,心理干预方案就是教学设计,心理干预过程的施行正是教学过程的实施。

其次,社会领域。信息战和心理战已成为现代战争中没有硝烟的作战方式,“深度迷惑”是一项跨学科的集群技术,旨在扭曲一个或一群人的看法,从而引诱对方,达成己方的目的。虽然心理战由来已久,但网络和人工智能的兴盛极大提升了该技术的使用价值。例如,深度伪造可以让公众对某一虚假信息深信不疑,精准推送能十分高效地诱导特定群体的行为。主要核心技术包括:1)“大规模网络诱劝”技术,即基于新闻传播和社会心理学相关理论,通过网络实施面向目标区域受众的信息推送,达到规模化诱导与劝服的效果;2)面向自然语言的“智能对话”技术,涉及自然语言对话生成、话轮组装与反馈、深度提问与反思等技术;3)“类脑推理”仿真平台,即在认知层面模拟人类的思维分析及因果推理能力,处理互不兼容甚至相互矛盾的观点,可用于定向模拟目标人群对特定主题的诱导效果,并剖析原因,提出改进建议。“深度迷惑”的本质就是对目标受众展开一系列定向“教育活动”,只是将教学从校内移到校外,从线下转到线上,从显性转向隐性,但大规模在线因材施教的核心理念未曾有丝毫变化。

2. 教育学引发的人工智能范式变革

人工智能基础理论的源头创新及其范式变革,一直是计算机科学界的热切企盼(钟义信, 2021)。与人工智能引发的教育学范式变革相比,教育学反哺人工智能所激发的范式变革更为震撼。如果前者多倾向于理论突破,后者则可以技术“变现”。毫不夸张地说,教育学将改变人工智能的发展史。

1)认知论层面——机器学习的教育学途径

首先,教育学原理成为人工智能领域基础纲领的可能性。教育学为机器学习开启了新的视角,使其在不改变基本算法的前提下,提升模型训练效果。人类有效教学的基本原理有良好的普适性,可适用于复杂的 ChatGPT 模型训练,也可用于机器学习中

线性回归、K-means 聚类、朴素贝叶斯等简单算法。事实上,就连训练过程的模式架构也可与人类教育学有内在一致性。例如,机器学习模型是一种先天的学习机,可类比为有学习能力的人类学生。不同的模型算法,相当于知识结构、学习能力各异的学生。从教育学的视角看,挑选机器学习模型就是确定参加学习的学生。继而,数据集可分为训练集、验证集和测试集,分别用于训练模型、修正模型和评估模型。有趣的是,其意义恰好对应人类教学过程的课堂讲授、同步辅导和考试测验。又如,机器学习还可分为监督学习和无监督学习两类,区别在于前者的训练数据集包含标签数据,在验证和测试模型时可以直接根据标签判断正确与否,但后者的数据没有此类标签,也不存在客观的判别标准。这类似人类的作业或考试任务,监督学习如同有标准答案的选择题或填空题,学生可以直接从题目中学习,考核也采取同样方式;无监督学习类似无标准答案的开放题,学生需通过给定的题干信息,自行发现潜在的模式或结构才能完成考核。可见,机器学习和人类教育无疑存在密切的内在关联。

其次,教育学原理成为机器学习基础纲领的可行性。可能性是一种理论考察,但上述“人—机”类比绝非个例,人工智能全部领域都存在着教育学的解释空间。这些显明、直接、普遍存在的对应关系,让人无法在理论上视而不见。从可能性到可行性,意味着从理论分析到实践应用的转向。能否在实践中验证理论假设,实现对理论的确证,这是检验学科范式科学性的试金石,也是在人类被试外,教育学另一种全新的科学证伪方式——通过对人工智能系统的指导效果检视教育学理论的正确性。例如,根据课程论和德育论相关原理,通用人工智能系统的培养路径应为“通识教育”,专用人工智能系统的训练方式则是“专科培训”。在具体的技术场景下,以最新的 ChatGPT4 为例,若想继续提升模型效果,什么才是可行的优化方向呢?在计算机科学领域,也许这不是一个容易回答的问题,但置于教育学理论视角看,答案便非常清晰:一是在训练时加入“感官”数据,进行更大规模的参数拟合;二是让模型先“验算”生成的答案,即将结果返回模型令其不断反思,再向用户提供多次确认后的稳定版本;三是将 ChatGPT 视为接受“通识教育”的基础版,收集面向特定领域准备更多优质数据(如中小学数学教材、辅导书等),将其并入基础版本数据并重新训练模型。所得到的新模型有望具有更好的“小而专”的业务能力。简言之,在“通用”基础上发展“专用”能力,模型效果会更好,因为人类就是这种教育模式的“成品”。

2)方法论层面——面向发展的新算法体系

将教育学作为手段和工具,是人工智能范式变革在方法论上的重要突破。和当下愈演愈烈的数学钻营之风相反,我们应向系统主动引入人的因素。正如人类新生儿不是先学会数学知识,才形成各种能力一样,人工智能也必定存在一条非数学化的发展之路。这是教育学的潜在要求,也是人工智能未来发展的新“装备”。在此举两例加以说明。

——教育学为可解释人工智能(explainable artificial intelligence, XAI)指出了新方向。随着人工智能技术的发展,其模型复杂度指数级攀升,ChatGPT4 的参数高达 1750 亿。模型效果是得到提高,但透明度却越来越低,这大大限制了它们在现实任务中尤其是错误预测代价很高的领域的应用(赵延玉等,2023)。尽管研究人员在模型独立和模型依赖两类传统解释方法的基础上加入因果解释方法,但主动性解释不足、全局解释与局部解释难以调和以及外部知识整合等根本性难题仍悬而未决(李凌敏等,2022)。站在教育立场上,我们能够窥见人工智能模型训练思路存在的重大缺陷,即试图将学习目标作为“一次性完成”的操作。事实上,学习通常意味着持续的动态改进,没有人总能一学就会。学习内容越复杂,就越需要更多的学习轮次。因此,遵循教育学规律,未来可解释人工智能应考虑如下改进方向:一是在解释过程上,从单步(one-step)转向多步(multi-steps);二是嵌入并发挥人的作用,着力构建人在循环(human-in-the-loop);三是在操作交互性上,从单向转向人机双向。

—— 教育学为主动视觉(active vision)铺设新路径。计算机视觉一直是人工智能技术的战略高地,深度学习浪潮下几乎所有重要模型都源自该领域。尽管基于专用人工智能的被动视觉近年取得长足进步,但其运动控制与多模态处理能力仍未明显提升。然而,不同于主流的被动视觉技术路线,主动视觉能让机器在复杂开放环境下自主探索,不

仅需要将通用人工智能系统作为大脑,还需要像人类婴儿一样逐步发展视觉——从静态本体观察单个视觉对象的远近变化,到静态本体观察远处视觉对象的相互遮挡,再到动态本体的形状认知,才能逐渐构建起真正意义上的主观视深能力。这种能力并非数学计算的结果,而是学习主体受教育后的“教学成果”。课题组的初步研究结果显示,OpenNARS 的视深能力会随概念网络节点数量的增多而提高,这体现出主动视觉的具身性、建构性和主动性的统一,亦印证了教育学手段反哺人工智能的可能性。

五、结论与展望

(一)结论

在教育学界,ChatGPT 迅速成为热议。作为大语言模型,ChatGPT 与其他人工智能技术有着直观区别,它有像人一样的会话能力,智能水平似乎可与儿童匹敌。但实际上,在人工智能诸多分类中,最适合时下讨论的是专用和通用人工智能之分。基于这一技术框架,ChatGPT 争议的根源初露端倪——人们站在专用人工智能的技术立场上,讨论着通用人工智能的应用问题。

通过对专用和通用人工智能认知特点的讨论,本文基于技术论视角明确了 ChatGPT 专用人工智能的本质,指出它不具有思维活动,也没有情绪情感,更不可能有意识。ChatGPT 是一款出色的软件工具,但其本身没有身体、没有认知,也没有智能。人们对它的担忧大可不必,很多风险实际上来自人自身问题在 ChatGPT 上的投射。因此,我们要以工具属性看待 ChatGPT,既不要过分悲鸣,也不能不加控制,更不可乱扣帽子。不过,ChatGPT 不是通用人工智能,不意味着通用人工智能并不存在,OpenNARS 就是通用人工智能系统的典型例证。

上述评价,似乎降低了 ChatGPT 的教育影响力,也降低了人们对人工智能的教育期待。这是另一种容易让人误解的表象,因为通用和专用人工智能一同催生了教育学和人工智能共融的“化学反应”,展现出人工智能中教育手段的优势,以及教育领域中人工智能主体的价值。然而,这只是人工智能与教育学双向融合的冰山一角。在认知论和方法论的视角下,人工智能与教育学的深度融合描绘了一幅双重范式变革下学科发展的壮美蓝图。

(二)展望

第一,热点容易追,但脚踏实地难。鉴于 ChatGPT 技术边界的局限性,如此惊艳的技术范型短期内难以再现。当 ChatGPT 热潮退去后,它为教育学界留下什么火种,其理论是否有新的发现,实践上是否有新的功能或应用?当学界从 ChatGPT 的通用人工智能梦境中醒来后,能否有人关注并投身于真正的通用人工智能,探索“人—机”二元框架下的新疆域?与纷繁的裂变之“多”相比,教育基本理论研究的“质”之深化显得相对薄弱,特别需要耐得住寂寞的研究者的个体修为和群体合力(叶澜,2021)。

第二,教育不仅能塑造人类的精神世界,亦能改造类人或人造的技术世界。人工智能与教育学的双重范式变革,不仅是学科共融的成功范例,更是自然科学与人文社会科学交融的范例。这必将辐射和推动心理学、脑科学、神经科学、医学、生物学、生态学等的边界变化,从而有望成为新一轮学科整合的引爆器。

第三,对教育技术学而言,这无疑是天赐良机。如果无所作为,就会错过历史机遇,失去引领时代潮流的话语权,不断被排挤和蚕食,最终在新一轮学科变革的竞争中逐渐丧失学科生存空间。反之,若能及锋而试,把握历史契机,必能赢得大科学时代的话语权,成为新一轮学科变革中连通人机、畅达文理的绝对核心。因此,“生存还是毁灭,这是个问题”。

[参考文献]

[1] AGI-Society(2008). The first conference on artificial general intelligence (AGI-08) [EB/OL]. [2023-05-01]. https://agi-conf.org/2008/AGI_08_Onepage.pdf.

[2] Alkaissi, H., & McFarlane, I.(2023). Artificial hallucinations in ChatGPT: Implications in scientific writing[J]. Cureus, 15(2): 1-4.

[3] Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4[J]. arXiv: 2303.12712.

[4] Cai, G., Haslett, A., Duan, X., Wang, S., & Pickering, J. (2023). Does ChatGPT resemble humans in language use?[J]. arXiv: 2303.08014.

[5] Chomsky, N., Roberts, I., & Watumull J. (2023). The false promise of ChatGPT[EB/OL]. [2023-04-25]. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html.

[6] 董奇, 陶沙, 曾琦等(1997). 论动作在个体早期心理发展中的作用 [J]. 北京师范大学学报(社会科学版), (4): 48-55.

[7] 杜华, 顾小清(2022). 人工智能时代的知识观审思 [J]. 中国远程教育, (10): 1-9+76.

[8] Geist, M.(2016). It's already too late to stop the AI arms race: We must manage it instead[J]. Bulletin of the Atomic Scientists, 72(5): 318-321.

[9] Hanna, R. (2023). How and why ChatGPT failed the Turing Test[EB/OL]. [2023-05-01]. https://againstprofphil.org/2023/01/15/how-and-why-chatgpt-failed-the-turing-test/.

[10] Harnad, S.(1992). The Turing Test is not a trick: Turing indistinguishability is a scientific criterion[J]. ACM SIGART Bulletin, 3(4): 9-10.

[11] Hayes, P., & Ford, K. (1995). Turing test considered harmful[C]. IJCAI, (1): 972-977.

[12] 侯怀银, 王耀伟(2022). 信息技术时代的中国教育学建设 [J]. 杭州师范大学学报(社会科学版), 44 (3): 67-75.

[13] 李恒威, 盛晓明(2006). 认知的具身化 [J]. 科学学研究, (2): 184-190.

[14] 李凌敏, 侯梦然, 陈琨, 刘军民(2022). 深度学习的可解释性研究综述 [J]. 计算机应用, 42 (12): 3639-3650.

[15] 李芒, 葛楠(2023). 教学革命论的纰缪与反正 [J]. 中国电化教育, (3): 1-8.

[16] 李政涛(2022). 什么是“教育学基本问题”[J]. 高等教育研究, 43 (10): 1-7.

[17] 李政涛, 吕雪晗(2022). 智能时代的文化转型与共生教育 [J]. 民族教育研究, 33 (6): 35-39.

[18] 刘凯, 贾敏, 黄英辉, 胡祥恩, 王培(2022). 像教育人一样教育机器——人类教学原则能用于通用人工智能系统吗?[J]. 开放教育研究, 28 (2): 11-21.

[19] 刘凯, 胡静(2018). 人工智能教育应用理论框架: 学习者与教育资源对称性假设——访智能导学系统专家胡祥恩教授 [J]. 开放教育研究, 24 (6): 4-11.

[20] 刘凯, 胡祥恩, 马玉慧, 那迪, 张昱华(2018). 中国教育领域人工智能研究论纲——基于通用人工智能视角 [J]. 开放教育研究, 24 (2): 31-40+59.

[21] 刘凯, 胡祥恩, 王培(2018). 机器也需教育?论通用人工智能与教育学的革新 [J]. 开放教育研究, 24 (1): 10-15.

[22] 刘凯, 李晶晶, 张喻华. 通用人工智能视域下人机共生的智能话语体系 [J]. 新媒体研究, 8(17): 14-19+36.

[23] 刘凯, 隆舟, 刘备备, 王伟军, 王培(2018). 何去何从?通用人工智能视域下未来的教师与教师的未来 [J]. 武汉科技大学学报(社会科学版), 20 (5): 565-575.

[24] 刘凯, 王培, 胡祥恩(2019). 心理学与人工智能交叉研究: 困难与出路 [N]. 中国社会科学报, 2019-01-14(6).

[25] 刘诗贵, 朱武振(2014). 地域差异的主流价值文化认同 [J]. 重庆社会科学, (2): 62-68.

[26] Long, Z., Andrasik F., Liu, K., & Hu, X. (2020). Self-improvable, self-improving, and self-improvability adaptive instructional system[M]// Pinkwart, N., & Liu, S. Artificial Intelligence Supported Educational Technologies. Springer, Cham: 77-91.

[27] Ma, Y., Tsao, D., & Shum, H. Y.(2022). On the principles of parsimony and self-consistency for the emergence of intelligence[J]. Frontiers of Information Technology & Electronic Engineering, 23(9): 1298-1323.

[28] Marchetti, A., Di Dio, C., Cangelosi, A., Manzi, F., & Massaro, D. (2023). Developing ChatGPT's theory of mind[EB/OL]. [2023-04-25]. https://psyarxiv.com/fr6xq.

[29] Min, S., Lewis, M., Zettlemoyer, L., & Hajishirzi, H. Metaicl: Learning to learn in context[J]. arXiv: 2110.15943, 2021.

[30] Nagy, E. (2011). The newborn infant: A missing stage in developmental psychology. Infant and Child Development, 20(1): 3-19.

[31] OpenAI(2023). GPT-4[EB/OL]. [2023-04-25]. https://openai.com/research/gpt-4.

[32] Oppenlaender, J. (2022). Prompt engineering for text-based generative art[J]. arXiv: 2204.13988.

[33] 庞小萍(2022). 连续谱视角下抑郁的心理学标志物的识别——初中生负性事件内联结构之研究 [D]. 锦州: 渤海大学.

[34] 彭道林(2016). 试论教学原则 [J]. 湖南师范大学教育科学学报, 15 (1): 44-51.

[35] 邱燕楠, 李政涛(2023). 挑战•融合•变革: “ChatGPT 与未来教育”会议综述 [J/OL]. 现代远程教育研究: 1-10.[2023-05-05]. http://kns.cnki.net/kcms/detail/51.1580.g4.20230309.1125.002.html.

[36] Shah, H., Warwick, K., Vallverdú, J., & Wu, D.(2016). Can machines talk? Comparison of Eliza with modern dialogue systems[J]. Computers in Human Behavior, 58: 278-295.

[37] 时广军(2020). 课堂场下的学生沉默: 诱因与对策 [J]. 北京社会科学, (7): 110-118.

[38] Stins, J. F. (2009). Establishing consciousness in non-communicative patients: A modern-day version of the Turing test. Consciousness and cognition[J], 18(1): 187-192.

[39] 孙元涛(2010). 教育学学科边界问题的再认识——关于“跨学科研究”的教育学思考 [J]. 教育发展研究, 30 (24): 31-35.

[40] Syme, K. L., & Hagen, E. H. (2020). Mental health is biological health: Why tackling “diseases of the mind” is an imperative for biological anthropology in the 21st century[J]. American Journal of Physical Anthropology, 171(S70): 171.

[41] Tang, Z., Rybin, D., & Chang, T. H.(2023). Zeroth-order optimization meets human feedback: Provable learning via ranking Oracles[J]. arXiv: 2303.03751.

[42] Wang P. (2013). Non-axiomatic logic: A model of intelligent reasoning[M]. Singapore: World Scientific: 179-185.

[43] Wang, P.(2019). On defining artificial intelligence[J]. Journal of Artificial General Intelligence, 10(2): 1-37.

[44] 王佑镁, 王旦, 梁炜怡, 柳晨晨(2023). ChatGPT 教育应用的伦理风险与规避进路 [J]. 开放教育研究, 29 (2): 26-35.

[45] 项贤明(2018). 论教育学的术语和概念体系 [J]. 教育研究, 39 (2): 43-51.

[46] 项贤明, 冯建军, 柳海民(2018). 教育学原理 [M]. 北京: 高等教育出版社: 264.

[47] 阎光才(2021). 信息技术革命与教育教学变革: 反思与展望 [J]. 华东师范大学学报(教育科学版), 39 (7): 1-15.

[48] 杨宗凯, 王俊, 吴砥, 王美倩(2022). 发展智能教育学 推动教育可持续发展 [J]. 电化教育研究, 43 (12): 5-10+17.

[49] 叶浩生(2022). 身体的意义: 生成论与学习观的重建 [J]. 教育研究, 43 (3): 58-66.

[50] 叶澜(2021). 新时代中国教育学发展之断想 [J]. 中国教育科学(中英文), 4 (5): 3-9+114.

[51] 张绒(2023). 生成式人工智能技术对教育领域的影响——关于 ChatGPT 的专访 [J]. 电化教育研究, 44 (2): 5-14.

[52] 张志祯, 张玲玲, 米天伊, 丘诗萍(2023). 大型语言模型会催生学校结构性变革吗?——基于 ChatGPT 的前瞻性分析 [J]. 中国远程教育, 43 (4): 32-41.

[53] 赵延玉, 赵晓永, 王磊, 王宁宁(2023). 可解释人工智能研究综述 [J/OL]. 计算机工程与应用: 1-16.[2023-04-30]. http://kns.cnki.net/kcms/detail/11.2127.TP.20230313.1550.016.html.

[54] 钟秉林, 尚俊杰, 王建华, 韩云波, 刘进, 邹红军, 王争录(2023). ChatGPT 对教育的挑战(笔谈)[J]. 重庆高教研究, 11 (3): 3-25.

[55] 钟义信(2021). 范式革命: 人工智能基础理论源头创新的必由之路 [J]. 人民论坛•学术前沿, (23): 22-40.

[56] 周子荷(2021). 教育领域技术原始创新的历史、逻辑与未来: 兼论人工智能的教育意蕴 [J]. 开放教育研究, 27 (2): 34-41.

(编辑: 魏志慧)

The Dual Paradigm Shift in the Two-way Integration of Artificial Intelligence and Education

LIU Kai[1,2]

(1. *Department of Educational Science*, *Bohai University*, *Jingzhou* 121013, *China*; 2. *Institute of Artificial General Intelligence of Bohai University*, *Jingzhou* 121013, *China*)

Abstract: *ChatGPT has rapidly become a hot topic in education. With an excess of focus on superficial issues related to its application and ethics, there is an urgent need to address those essential and fundamental issues. Thus, to establish a pertinent technical and theoretical framework has become urgent and necessary. Firstly, from the perspective of technology, the paper describes Artificial Intelligence (AI) frameworks which include Special-purpose Artificial Intelligence (SAI) and Artificial General Intelligence (AGI) to decipher the essence of ChatGPT. The paper explains that ChatGPT is a typical SAI system, while OpenNARS is an AGI system with cognitive capabilities. It also clarifies common misconceptions about the "personification" and "stigmatization" of SAI, and the "omnipotence" and "exaggeration" of AGI. From the ontological perspective, this paper reveals the potential for the integration of AI and education. It uncovers the key factors that hidden behind ChatGPT with a belief that AI has opened a new era of the human-machine binary subject in education. From the perspective of epistemology and methodology, the paper describes the two-way integration between AI and education, revealing the internal mechanism and logical framework of the paradigm shift in these two disciplines with education providing a new interpretive perspective and theoretical basis for AI, and AI bringing in new research subjects and new research methods to education. Currently, education, as an academic field, does not match the full potentials of AI advancement. Its theories are overly reliant on other disciplines, and its research methods are criticized and questioned. Finally, the paper suggests, with the help of AI, education will improve itself by being more scientific with a strengthened discipline, breaking up its boundaries, promoting paradigm shifts, and facilitating interdisciplinary integration to ultimately gain the core discourse power in the era of "Big Science".*

Key words: *education*; *artificial intelligence*; *SAI*; *AGI*; *ChatGPT*; *paradigm shift*