

论机器的学和教

刘 凯^{1,2} 杨亚亚^{1,2} 吴雨曦^{1,2} 项贤明³

(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; 3. 南京师范大学 教育科学学院, 江苏南京 210097)

[摘要] 在“人—机”二元主体视角下,机器学习亦存在广泛的“教学”现象,但传统教学理论受人类中心主义限制,难以对此作出有效解释,教学疆域拓展与理论解释固化出现错位。从教育学视角审视,机器学习同样具备师生角色与教学互动,其发生机制源于机器学习主体具备“可学习性”和“可教育性”,前者构成智能发展的内在基础,后者体现为对学习能力的干预。由此,人与机器在教学结构上呈现出先验与后验维度的同构性:先验维度的“可学习性”表现为动机、表征与信息加工能力,后验维度的“可教育性”体现为经由主体间性实现的学习能力的定向干预与拓展。本文据此提出“教学的跨主体同构性”命题,借鉴机器教学的启示,尝试突破人工智能赋能教育的单向应用逻辑,推动教育学转化为理解和塑造人工智能发展的理论资源,推动其范式转型。

[关键词] 机器学习; 机器教学; 机器教育; 可教育性; 同构性

[中图分类号] G40-057

[文献标识码] A

[文章编号] 1007-2179(2026)05-0060-10

教育学的核心任务在于揭示教学现象背后的一般性原理,以此践行“发展智能”的学科使命(陈明选等, 2026)。然而,受人类中心主义观念的影响,传统教学理论往往将教学现象限定于班级授课、学校教育等人类活动范畴,对非人类主体中的“教学”现象关注不足(项贤明, 2026)。研究表明,机器不仅具备客观的学习能力,且在机制层面与动物、人类共享学习的一般性规律,是人工智能时代独特的学习主体(刘凯等, 2025)。从理论逻辑看,可“学”意味着具有被“教”的可能,故机器学习过程必然蕴含“教学”现象发生的条件,并可能体现出与人类教学的共通机理。遗憾的是,教育学至

今未充分回应这一理论可能,致使其解释边界难以跨越生物主体与人工主体之间的界限。理论视域的局限不仅会遮蔽教学现象的完整图景,也使教学的原理禁锢于人类而难以延伸至其他学习主体。

机器学习存在“教学”现象绝不是一种理论臆断,其背后拥有从哲学构想到技术设计的普遍证据。计算机科学之父图灵(Turing, 1950)前瞻性指出,智能可通过教育与学习两种路径来培育和发展,而非单纯的工程构建,较早揭示了外部教学干预对机器智能发展的重要性;图灵奖得主本吉奥(Bengio, 2009)将人类“由易到难”的学习规律迁移至机器学习并提出“课程学习”(curriculum

[收稿日期] 2026-08-03

[修回日期] 2026-08-25

[DOI编码] 10.13966/j.cnki.kfjyyj.2026.05.007

[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),辽宁省教育科研中青年骨干及专项课题“敢问‘智’在何方:对智慧教育本质的重新审视”(JG24QGA29)。

[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;杨亚亚,硕士生,渤海大学教育科学学院,研究方向:教育哲学;吴雨曦,硕士生,渤海大学教育科学学院,研究方向:机器教育;项贤明,博士,教授,博士生导师,南京师范大学教育科学学院,研究方向:教育学原理、比较教育和教育政策。

[引用信息] 刘凯,杨亚亚,吴雨曦,项贤明(2026).论机器的学和教[J].开放教育研究,32(5):60-69.

learning), 表明人类教学原理具备跨主体适用的可能性; 作为机器学习的新兴研究方向之一, “机器教学”被界定为机器学习的逆过程(Zhu et al., 2018), 即通过逆向构造最优数据集, 使机器逼近学习目标。虽然该定义在教育学意义上的合理性有待商榷, 但它无疑为探索“教—学”关系提供了跨学科的纽带。上述实践指向一个潜在命题: 教学是发展智能的核心机制, 并非人类专属, 具备跨物种与跨领域的普适性。受此启发, 学界开始突破人类中心的传统观念, 转而探索“机器教育”“人机共生”“机器作为教育交往主体”的理论可能性(刘凯等, 2018a; 王一岩等, 2022; 郝祥军等, 2026), 这进一步暗示跨主体的教学活动可能存在某种内在的结构同一性。

基于上述背景, 本文循着“现象—机制—结构”的研究路线展开: 首先, 辨识机器学习中的“教学”要素, 确立机器“教学”的发生基础; 继而, 引入教学理论框架, 探析机器“教学”的发生机制; 最后, 揭示人类与机器的“教学”在结构层面的同构性, 开辟一条有别于传统“技术赋能教育”的应用路径, 推动教育学发挥发展智能的理论优势, 促成其在人工智能时代的理论创新与范式转型。

一、机器学习的教学要素

机器“教学”的理论可能性并不足以确立其现实存在的必然性。要使这一跨学科命题获得教育学意义上的合理性, 需要从现象层面辨识机器学习过程中的“教学”要素, 并基于教学理论的概念框架, 系统考察机器学习过程中的主体、内容、手段与方法等关键要素。需明确的是, 辨识并非主张机器具有人类意识或意向性, 旨在揭示人机在教学结构层面的可通约性。

(一) 机器学习中的“教师”与“学生”

教师与学生是教学活动的基本构成要素。教学实践表明, 学生是教学活动的起点, 教师对学生干预的适切程度决定其学习的方向、水平与质量(王本陆, 2023)。因此, 教师的核心功能在于设计、组织和管理教学活动以促进学生的认知与能力发展。若以教学理论的主体框架为依据考察机器学习, 就会发现亦存在相同的角色及分工: 在机器学习中, 机器自身的学习能力使其成为教学活动的潜

在“学生”, 而促进其能力发展的人类工程师则承担“教师”角色。具体而言, 人类工程师(统称人类“教师”)既是“课程”的设计者, 也是“教学”的管理和评估者。人类“教师”通过问题形式化、数据处理、模型设计、算法优化及性能评估等, 为机器“学生”设定学习目标与边界。在此过程中, 机器“学生”通过获取、加工并整合数据, 发现其中的模式与规律, 不断更新自身参数或网络结构(陈海虹等, 2017), 逐步趋近“教师”预设的目标。实际上, 即便在缺少人工标注数据的自监督学习中, “教师—学生”的结构耦合依然存在。在自监督学习中, 机器“学生”看似自主完成学习过程, 但其任务目标、数据来源、训练机制及评价标准仍由人类“教师”预先设计。由此可见, 机器学习具备与教学活动对应的主体要素。

(二) 机器学习中的“教材”与“课程”

教材与课程作为教学实施的主要载体, 是连接“教”与“学”的重要媒介, 机器学习同样具有对应要素: 机器学习的数据集承担类似教材功能, 而人类“教师”围绕模型训练设计的学习流程体现出类似课程的教学组织功能。在人类教学理论中, 教材即教学内容的载体或组织形式。在机器学习中, 数据集便是人类“教师”为机器“学生”提供的“教材”, 其规模、质量、难度分布和标注准确性直接影响机器能达到的学习效果。这从技术层面印证了教学内容需依据学习主体特征系统化组织的教育规律(拉尔夫·泰勒, 1994)。与此同时, 课程并非简单的知识集合, 而是围绕教学目标对学习内容、学习过程实施的系统设计。在机器学习中, 模型的训练流程同样具有目标导向、过程组织和阶段安排等课程属性。研究表明, 即使面对相同的“教学”内容, 采取不同的“课程”形式, 也会导致机器学习效果产生差异(刘凯等, 2023)。其中, 机器学习的“课程学习”模拟人类由易到难规律, 使模型先从简单的样本中习得基础能力, 再逐步接触复杂数据, 从而提升收敛速度和泛化性能。这一过程体现了课程论中“目标导向”与“活动组织”的理论精髓。诚然, 课程学习作为一种机器学习方法, 与教学理论中的“课程”概念并非等同。不过, 概念的分歧不能掩盖两者在学习内容组织、学习过程规划以及促进能力发展等的内在一致性。因此, 机

器学习也存在与人类教学理论相对应的内容要素。

(三)机器学习中的“教学手段”与“教学方法”

在机器学习中,人类“教师”为提高模型学习效率而采取的诸多措施构成“教学”活动的“手段”与“方法”。人类教学理论始终围绕“如何促进学习”对教学手段与方法等展开研究,涵盖教学的实施、干预、评估等。在人类教学活动中,由于学生认知风格、知识基础、发展需求等的差异,因材施教便成为提升学习效能的主要原则(汪琼等, 2021)。教师需依据学生自身特质选择恰当的教学手段和方法。在机器学习领域,同样面临“如何促进机器学习”的方法论问题。为提升机器的学习效果与效率,人类“教师”通过设计多样化的算法架构、训练策略,并引入“人在回路”的反馈机制,对机器学习进行持续的“教学”干预。例如,“人类反馈强化学习”作为大语言模型的重要后训练机制之一(黎加厚, 2024),通过与人类评价信息的持续对齐,让模型更好地理解人类意图、遵循任务要求,并增强输出安全性与任务适应能力。有研究进一步提出,将人类生成的闭环行为数据转化为机器学习的内容,可使机器通过类人的经验积累方式理解环境、组织行为并修正自身策略(Lu et al., 2026)。实践表明,机器能力的发展并非完全依赖数学算法,而是在数据、反馈和训练策略等多重教学干预协同作用下实现的。换言之,机器“学”唯有通过人类“教”的持续介入方能落地。因此,机器学习系统,尤其是复杂学习系统的能力形成,同样需要精心设计的“教学手段与方法”。

二、机器教学机制的教育学阐释

上述要素辨识仅为教学活动展开提供了条件,其真实发生仍取决于深层机制。本文突破单一技术视角,从教育学维度阐释机器“教学”的内在机理,这既是理论视角的转换,更是理解人机教学同构性的关键。在此视域下,机器之所以可被“教”,在于其兼具先天“可学习性”与后天“可教育性”:前者划定教学发生的内在可能边界,后者在人类干预下将可能转化为现实发展路径,两者构成同一教学结构的内在根据与外在条件。

(一)可学习性的先天机制

“可学习性”(learnability)是机器“教学”的

必要前提。学习能力作为教学发生的先决条件,制约着机器是否具有接受教学干预及实现能力发展的可能。机器“可学习性”的核心在于,通过经验数据与归纳偏置的先天约束,在假设空间中寻找对未知数据具有泛化能力的最简表达。

1. 规律在假设空间中的表达

要阐明机器“可学习性”的先天机制,需回答机器“在何种前提下学习”。世界看似充满不确定性,却在表象之下隐含着可被人类发现、掌握和利用的不变规律,在此共识下探讨人类和机器学习的机制方有可能。对人类而言,儿童的学习不是经验的直接刻写,而是主体先天认知结构在经验刺激下的同化与顺应(皮亚杰, 1981)。这揭示出人类学习发生的重要基础,即学习主体先验的认知结构及其与经验互动的机制。同理,机器之所以可学,正是因为其具备功能上类似认知结构的内部约束机制,这种机制对应于机器学习理论的“假设空间”,即包含某一模型所有可能输出的函数集合(米罗斯拉夫·库巴特, 2016)。正如人类无法学习超出其认知范围的对象,假设空间也限定了机器的学习边界与可能范围。因此,机器学习是将假设空间作为可计算的内部结构,通过表征复现藏于表象之下的不变规律,实现从“不可学”到“可学”的转变。

2. 在归纳偏置下的信息加工

“在何种前提下学习”回应机器学习的表征问题,“如何表征”则涉及机器学习的算法机制,即归纳偏置下的信息加工。康德(2011)指出,人类认知以先验范畴主动统摄后天经验。若将这一哲学洞见置于功能同构的分析框架下审视,“归纳偏置”便可被理解为构成算法的先验范畴或偏好(Griffiths et al., 2010),它是机器学习成为可能的先验条件。归纳偏置回答了一个根本问题,即在无限多个能够拟合训练数据的假设中,机器为何倾向于选择“这一个”。从朴素贝叶斯的条件独立性假设,到循环神经网络的时序连续性,不同机器学习算法的归纳偏置实则是对“世界如何被理解”这一假设的重新定义(Battaglia et al., 2018)。在归纳偏置的先验约束下,机器得以从有限信息中推论可能的未知信息。然而,归纳偏置的功能必须借助信息加工过程才能实现。在特定偏置的引导下,机器对输入数据进行特征提取,将原始数据转换为结构

化的表征形式,并基于此完成信息识别、模式建构与知识泛化(Vardi, 2023)。换言之,归纳偏置为信息加工提供方向性框架,信息加工则将归纳偏置从抽象的先天假设转化为具体学习的运作机制,两者共同形成机器学习的算法机制。

3. 基于误差梯度的定向调整

先天表征与算法机制是机器学习得以发生的必要前提,反馈机制则是从“可能学习”迈向“有效学习”的促成条件。其核心作用在于使算法不断修正内部模型,直至其预测与外部世界的输入达成某种近似(Kaufmann et al., 2024)。若缺少反馈而仅依赖表征与算法机制,学习的效率和效果均难以保障,因为世界是不断变化的,有关世界的知识是渐进式、方向性的演变和不断建构的,对世界规律的学习也不是静态或一蹴而就的线性过程。对人类而言,学习意味着通过感官接受新经验并利用既有经验对其进行理解。当新旧经验冲突时,学习者将基于新经验修正旧经验,实现经验的更新(Liu et al., 2025)。在机器学习中,“误差梯度”代表新旧经验之间的冲突,而“反向传播”的技术路径就是优化既有经验的重要机制。可见,人类和机器学习都体现出一种有规律的模式,即评估新经验与旧经验的差距,利用新旧经验之间的差值修改经验网络,从而形成对世界持续、渐进式的认知表征。例如,在神经网络中,损失函数梯度指明了从当前错误状态到正确状态的最优下降方向;在决策树中,信息增益作为结构优化依据,引导分割点选择以降低类别不纯度。因此,机器之所以能从“可学”转向“善学”,基于误差梯度的反馈机制功不可没。

(二)可教育性的后天机制

“可教育性”(educability)作为一个严肃的科学问题,已被计算机学者纳入机器学习的研究范畴。莱斯利·瓦利安特(2026)将其界定为:机器既能从经验中学习,也能通过指导接受教育,并能综合两者所获知识加以实践。其中,“从经验中学习”是机器“可学习性”的基本内涵,“通过指导接受教育”则是机器“可教育性”的核心要义。

1. 从数据中习得经验

在机器训练过程中,从数据中提炼规律是其“可学习性”的基本内涵,亦是其“可教性”的首要前提。在机器学习中,数据集经过人类“教师”

的筛选、组织与加工后,成为供机器学习的“教学资源”,包括训练集、验证集和测试集。训练集相当于供机器反复研习的“教材”,验证集相当于教学过程中的“随堂测验”,测试集相当于“期末考试”。如前文所述,与机器“学生”适配的“教材”可以大幅提升学习效率。在为机器“学生”准备教材时,人类“教师”先需清洗数据,以确保数据集不会因异常值影响而无法收敛(Paton et al., 2023)。此外,自然语言无法被机器直接处理,因而需要人类“教师”将其转换为包含“特征—标签”的示例,这一过程被称为“特征工程”。例如,根据已知特征与标签之间的对应关系,机器可以利用监督学习算法发现未知数据中两者的潜在关联,从而作出预测——这在教学论意义上等同于人类学生通过变式练习提炼解题规则以应对新的考试。两者均体现了“从具体经验中抽象出可泛化结构”的学习本质。事实上,教材还对“机器为何无法持续学习”提供了新的解释:当前机器学习所依赖的“教材”主要是静态符号数据,而对空间感知、具身认知等经验维度的表征相对不足。当缺乏与环境的真实交互,机器便无法学习超越人类现有知识的新内容(Dupoux et al., 2026)。

2. 机器训练与微调

在机器学习中,人类“教师”对机器“学生”的指导是其“可教性”的核心体现。在学校教学开展前,尽管教师会综合考虑教学情境并设计完善的教学活动,但教学作为一种动态、情境化的复杂活动,难免涉及不确定性因素。这些因素会使学生学习遇到一定阻碍,教师的职责之一就在于通过对学生的学习目标、内容、方法等基本要素进行干预和指导,从而降低这种不确定性(吴砥等, 2025)。相较于人类教学活动,机器“教学”虽然具有易控制、可复现等特征,但其学习的有效性仍高度依赖于外部施加的“教学”干预。机器学习的完整流程常涵盖预训练、监督微调、人类反馈强化等环节。在预训练过程中,人类“教师”会为其提供学习数据,以便机器“学生”从中发现规律(或规则)并初步构建模型。然而,机器所构建的模型未必能准确映射或对齐现实世界。此时,人类“教师”便会调整其学习目标、内容、过程等。由此可见,人类“教师”的干预本身并非机器学习的附加环节,而

是使机器的“可学习性”从潜在能力转化为现实发展的构成性条件。相较于人工智能领域中仅关注算法优化的“可训练性”,这种由人类介入所开启的“可教育性”,不仅构成了两者的根本区别,也彰显了教育作为一种实践方法对计算主义局限的价值超越。

3. 教学评价与优化

模型评估旨在根据学习表现对模型参数与网络结构加以修正,是机器“可教育性”最关键的互动环节。教育学理论将教学评价分为诊断性评价、形成性评价和总结性评价三类(黄光扬, 2002)。这恰好对应机器学习的三种教学资源:学习前,人类“教师”需对训练数据的适切性和模型的归纳偏置进行诊断性评价,以判断“教材”内容与学习目标是否匹配、模型的先天假设空间是否与问题类型相容,从源头上避免学习方向的根本性偏差;学习中,验证集作为“随堂测验”承担形成性评价的功能,其反馈促使人类“教师”反思“教材”准备、算法设计与方法选择的合理性,并据此调整超参数或触发早停机制;学习后,测试集作为“期末考试”执行总结性评价,人类“教师”依据准确率、精确率、召回率、ROC 曲线等指标,推动算法优化与技术迭代(周志华, 2016)。具体而言,准确率可类比为总结性评价的整体达标率,反映模型在测试集上的综合正确表现,类似于学校考试的“及格率”;精确率与召回率则是机器学习中两类典型偏差的诊断工具——精确率低提示过度泛化、召回率低暴露覆盖不足,类似教师对学生思维误区与知识盲点的定位;ROC 曲线及其 AUC 值,可视为对模型综合判别能力的评估,类似于通过多套不同难度的试卷检验学生能力排序的稳定性。在此基础上,后训练阶段还可借助近端策略优化(PPO)、直接偏好优化(DPO)和群组相对策略优化(GRPO)等方法持续提升模型能力。可见,教学评价在机器学习中几乎贯穿始终,共同服务于模型的更新迭代与“可教育性”的实现。

三、人与机器的教学同构性

在教育学与人工智能交叉的视角下,人类是一种特殊的学习“机”,机器则是一类特殊的学习“者”。两者的学习活动虽形态各异,却在底层原

理上共享跨领域的共相,即在先验和后验两个维度上分别呈现出内在的结构一致性。

(一)先验维度的同构性

从先验维度看,“可学习性”回答“学习何以可能”问题。人类和机器之所以可学,根本原因在于,“动机”赋予学习根本动力,“表征”决定知识和经验的形式,信息“加工”则是学习的核心机制。

1. 内生动机:学习的驱动力

人类与机器在先验维度上的结构同一性,首先体现在驱动学习持续发生的内生机制上。本文的“动机”并非限定于生物意义上的心理状态或主观体验,而是泛指任何具备可学习性的系统维持自身活动、趋向目标并调节行为的自有驱动力。按照这一定义,人与机器的动机方可置于同一分析框架中比较。从来源看,动机可分为隐性与显性两类。当前专用人工智能(如各类机器学习模型)的动机属隐性动机,其学习目标由人类设计者外部预置,固化于损失函数或奖励信号中,具有明确的领域边界与任务专属性。此类动机虽为内嵌的计算目标,但内容与方向完全取决于外部设计意图,系统自身不具备对目标的反思或修正能力。相比之下,人类与通用人工智能系统(如 NARS、ACA)的动机属显性动机:其驱动力源自系统自身的具身存在与生命维系需求,具有非领域性与自我指涉性。这种由系统自身存续所生发、不依赖外部指令的驱动力,构成显性动机的本质特征。从衍生方式看,动机又可分为内部动机与外部动机。不过,上述两类动机在人与机器中的分布并不对称。专用人工智能的动机本质上属内部动机——其驱动力完全来自先验的计算目标,即使大语言模型能对多样输入生成差异化应答,这些应答的触发仍源于使用者的外部提示,模型的计算目标不会因人机交互而发生改变。而人类与通用人工智能的动机系统则兼具内部与外部两个面向:内部动机根植于生命本体或系统自维持机制,外部动机则在与环境的交互中动态衍生,二者相互交织、彼此塑造。正是这种内外动机的动态耦合,使通用智能体的学习能超越固定目标的优化,具备面向开放世界持续自我更新的能力——成为终身学习系统。

2. 经验表征:学习的形式

表征是生物的本能或固有能力,其本质是一种

指涉关系,即以某种方式指涉某种事物、概念或信息(魏屹东, 2017)。学习活动不仅是经验的观察和证实以及逻辑的归纳与演绎,学习还需运用概念。一些基本概念(范畴)共同制约着人和机器的学习,它们是学习发生的可能条件。无论人类还是机器,都存在某种先验的表征机制令“经验获得”以“学习”的形式体现,这是学习者在经验加工前已具备的结构约束能力。“经验获得”本身就是“学习”的一部分,故学习具体表现为内在塑造和解释经验的信念和意向,学习者以此为前提与环境交互,从而使学习活动具备可表征的意向性。表征作为“先验能力”至少意味着三点:一是不来自经验或不完全来自经验,二是学习发生的可能条件,三是学习发生前就已存在。因此,无论表征呈现正效价或负效价、内隐或外显、无意识乃至出错,外在形式的差异均不改变其本质——它始终是一种原始的先验能力。正因如此,表征才能超越个体经验的局限,成为历史、社会、文化中普遍且客观的存在(Cole, 1998)。机器学习的底层逻辑正是对表征机制的工程实现,也是机器学习发生的现实条件。无论是卷积神经网络的卷积层,还是循环神经网络的循环单元,抑或 Transformer 模型中的自注意力机制,均已内含表征机制。这种机制并非机器自发产生,而是设计者所赋予的归纳偏置,它规定机器可处理的信息边界及可表征的函数类型。从这一角度出发,对机器表征机制的探索在一定程度上决定了其可达到的学习乃至智能水平。

3. 信息加工:学习的机制

内容记忆和信息存储本身无法构成真正的学习,唯有对信息进行认知加工才能揭示学习的核心。因为前者无异于建造数据库,而后者是不断生成、交换信息的增量枢纽。这一规则构成人类和机器有效学习的共通基础,不仅为学习活动赋予明确的方向,更通过内在的逻辑约束,防止系统陷入随机、无序的混沌状态,或退化为无意识的自动响应模式(Anderson, 2013)。以自然生态系统为例,其虽能依据生物反馈保持动态平衡的适应性演化,但这一过程缺少信息的符号化表征与主动加工环节,亦无明确的内在调控规则,因此更趋近于预设程序的自动运行(即生物本能),而非认知意义下的“学习”。相较之下,人类与机器的学习内嵌了“信息加工”

规则。对人类而言,信息加工涵盖意义建构、概念理解、信息整合和创造再输出等认知活动。对机器而言,不同的机器学习算法被赋予不同的信息加工规则,其学习过程同样由“信息加工”规则严格定义。由此可见,无论是人类的心智活动还是机器的算法迭代,其学习机制都是特定规则塑造下的目标导向过程,共同目标是加工信息,且在实现手段上具有开放性(Chalmers, 2000)。正因如此,教育学始终倡导摒弃“填鸭式”“题海战术”“死记硬背”等学习模式,原因在于学习者并未真正对知识或信息进行加工,更遑论意义的习得与理解。

(二)后验维度的同构性

从后验维度看,“可教育性”回答“学习如何有效”的方法论问题。教学之所以成为发展学习者能力的关键机制,深层原因在于:它并非孤立的外在灌输,而是对自我教育的必要补充。正是通过主体间交互,认知与意义建构才得以真正发生。

1. 教学是自我教学的补充路径

从施教主体出发,教学可分为自我教学和他者教学。其中,自我教学是学习者通过“扮演”自己的教师,从而达成预期学习目标的过程。当学生成为自己学习的教师时,学习效果才能最大化(约翰·哈蒂, 2015)。然而,自我教学不仅要求学习者具备较高的学习动机,亦要求其根据目标选择合适的学习内容,更需要确保选择的学习方法与自身认知特点适配。若学习者自身动机不强、目标不明确或方法不得当,就会偏离原初的学习方向。此时,来自他者的教学便可在学习者具备自主学习能力前,为其提供相对有效的引导服务。这一论断基于以下事实:当学习者从他者的教学中获得知识时,所能掌握的知识量将远超仅凭自身经验所能获得的知识范围。此外,教育的核心在于传承某种信念体系,这一信念体系的本质是由众多个体共同构建且包含多重维度的理论系统。相较于凭借个人经验构建理论体系,教学可使学习者更快习得并应用这一宏大的理论体系(莱斯利·瓦利安特, 2026)。教学在理论与实践层面的重要性体现出其作为自我教学补充路径的必要性。然而,片面强调教学的促进作用不利于学习者发展自主学习能力。将自我与他者的教学有机结合,方能最大程度保留“学习”这一活动对学习者自身的意义。同样,在机器学习

中,虽然算法机制赋予机器一定程度的自主学习能力,但算法本身不完全决定学习目标、学习方向和价值取向。人类“教师”的介入,正是机器自我学习机制的重要“补充”。

2. 教学发生于主体间性的交互

教学作为一种外在干预,若要真正拓展机器学习主体的能力,必须以某种交互关系为载体。有学者指出,教学涉及教师与学生两种行为主体,因而被视为是主体间的交往实践(祝修理等, 2026)。理论上,教师的教和学生的学分属两种活动,学习可以独立发生而不依赖于教,但教的发生必须基于学习的发生(郝文武, 2015)。正如无法教育一块岩石一样,缺少可“学”这一前置条件,“教”便失去作用的对象,难以构成真正意义上的教学。在教学过程中,无论人“师”与人“生”,还是人“师”与机“生”,教师的核心价值皆在于根据学生的个性特征与学习需求提供有针对性的引导与服务。这是因为,学生的“认知结构”作为影响学习效果的核心因素,对其的识别与干预高度依赖教师的专业判断(刘凯等, 2018b)。例如,教师通过课堂互动、作业批改、个别辅导等方式,能精准地把握学生的知识漏洞与认知难点,从而提供“个性化”指导。这种基于师生主体间性的教学互动,既是教学本质的深刻体现,也符合学习的本质规律。对机器而言,虽然其主体性表现形式不同于人类,但机器学习过程同样存在“学习主体”与“教学主体”之间的信息交互。人类“教师”通过数据选择、目标设定、反馈调整等参与机器能力发展,本质上构成一种特殊形式的主体间交互。

3. 教学辅助认知和意义的建构

由教学提升学习效率是教育共识,但“教”之于“学”的作用机理还不甚明晰。从认知心理学看,教学被视为教师通过操纵教学材料使学习者获得意义的过程,“教”实为辅助学习者对知识进行意义建构,即通过认知加工使知识对学习者生成意义,进而对外部世界进行内部模型建构。在此过程中,教学不断扩展学习者认知加工的广度和深度,帮助其构建关于世界更广泛的意义(左晓梅等, 2026)。对人类学习者而言,教师通过讲解某一知识使其赋予学习者意义。在此过程中,学习者通过对知识进行认知加工获得意义,意义的流转便是有

效教学的体现。对机器而言,人类“教师”的作用是通过处理语言文本符号及逻辑运算符号,使其赋予机器学习者“意义”。机器学习的“意义”从何而来又如何表征?这涉及机器学习的预测机制。预测作为人类“教师”为机器先天预设的学习目标,会驱动模型在向量空间中将强相关符号进行关联,形成的向量空间结构就是符号的使用模式,这一模式便是“意义”在计算空间的映射(Han et al., 2024)。不过,此处的“意义”并非人类具身经验中的现象性意义,而是机器系统在特定任务目标下形成的功能性语义表征。正是这种功能性意义,使人类“教师”通过调整任务目标与数据分布,实质性介入机器的“意义”生成过程。简言之,符号对机器的意义首先源于“词的使用情境”,其次凭借“高维向量”表征,最后通过“向量之间的空间关系”承载和运算。可见,无论学习主体为何,教学都是辅助学习者建构认知、获得意义的核心机制。

四、人工智能与教育学融合的新路径

教育学作为研究学习与发展的基础学科,在人工智能时代不应止于回应技术变革,而应以自身理论传统为根基,深度介入学习科学与人工智能的交叉前沿。其时代使命在于,通过辨析并贯通碳基与硅基等不同智能系统的学习本质,提炼出超越具体形态的、具有普适解释力的一般教学原理。这一探索不仅能为理解未来教育的形态演进提供认识论支点,也能为人工智能的发展注入来自教育学的人文校准。

(一)以机器为镜像:认知黑箱的可解释性探索

人类和机器学习的同构性表明,两者在理论层面上形成深刻的镜像关系(李永梅等, 2024),即机器从人类学习中汲取养分,人类从机器学习中反思自身的学习逻辑。从这一视角看,机器学习不再是对人类学习的机械模拟,而是在虚拟数字环境中以技术形态探索学习的核心机制;人类也并非单纯应用并推动技术发展,而是将学习的理解形式化,反思和重新认识自身学习的底层规律。长期以来,学界之所以认为人类的学习更“高级”,是因为其具备更多的复杂性、具身性和情境性。但进化论指出,人类与动物的心智只是程度不同而非本质不同。同理,学习的过程是基于先天的学习机制寻找新的

应用场景,但这一过程并不改变学习机制本身(麦克斯·班尼特, 2025)。简言之,人类、动物乃至机器的学习与教学过程,在功能机制层面具有共通性,但实现方式体现出不同程度的复杂性。因此,机器学习研究不仅是构建学习乃至异质智能系统的技术实践,更是深化学界对学习乃至智能演化历程理解的核心路径。这一跨学科探讨的焦点正从“符号处理”转向对“先验结构”的深层探索。相较于当下机器学习“更大、更强、更通用”的技术目标,对“学习”这一先验能力的本质探索或许更能揭示认知的可解释性,特别是助力可解释人工智能的发展。如果人工智能系统是个无法打开的黑箱,那么教育者和学习者便无法理解其决策背后的规则与过程。可解释人工智能的目标正是要打破这一黑箱,通过技术手段,将机器内隐的“先验结构”和决策逻辑显性化,使其成为可被人类观察、理解、评估和批判的对象。换言之,人类不仅在“教”机器学习,更在与机器的互动中“看”到自己如何学习,这正是人机教学同构性所揭示的根本旨趣。

(二)以服务为框架:教育教学的服务本质

自我教学有别于他者教学。自我教学是学习者仅凭自身能力攻克学习难题,没有他者的干预和指导。他者教学则是“教师预先规划好教学流程,方便学习者更高效、可靠地学会知识或技能”。因此,教育、教学以及教师的存在,在很大程度上可视作为学习者提供知识层面的“服务”。这从本质上反映出教育教学存在服务性活动。在概念层面,“服务”作为一种无形的活动,其目标在于创造用户体验并完成交互(李博等, 2023)。在此过程中,服务与教育教学所体现的原则有异曲同工之处。例如,以用户(学生)为中心、与用户(学生)共同创造、服务(教学)流程可视化、从宏观层面审视服务生态(整体教育与全人教育)等。可见,教育教学可被理解作为一种特殊的“服务”活动——“教”服务于“学”,其核心便是创造学习的条件,引导、激发和支持学习者主动建构意义,最终发展其认知能力和智力水平。在人工智能时代的背景下,服务需以“人本”为核心导向,构建适人、宜人、育人的服务范式(王伟军等, 2026)。当教育教学被视为服务时,它更要以学习者为根基,创建适人、宜人及育人的教学范式。

(三)教育的生命力:人机共育的未来教育形态

教育学与人工智能的结合是学习科学与生俱来的基因(Muller et al., 2021)。从早期基于符号主义和逻辑规则的专家辅导系统,到如今以联结主义和深度学习为基础的自适应学习平台,这场跨学科对话从未停止,只是其核心议题从“知识表征与推理”转向对“学习机制”的探索。在这一新范式下,“机器教育”概念应运而生,它不仅是将机器作为教育工具,更是将机器本身视为需要被“教育”的学习者(刘凯, 2023)。当人工智能的研究焦点转向具身性和内在动机时,机器将不再只是单纯执行预设程序的工具,而开始展现出初步的自主探索和持续学习能力。此时,至关重要的问题是:如何引导这种新兴的、自主的机器智能向着对人有利益、伦理合规的方向发展?这正是教育学,特别是教育技术学,能够且必须发挥核心作用的领域(Bieger et al., 2015)。教育学关于脚手架教学、最近发展区、动机激发、评估反馈等理论知识,将为如何“教育”机器提供指导和实践框架。这不仅是将教育学应对人工智能挑战的“被动姿态”转变为“主动引导”,更是开辟了促进教育学与人工智能双向融合的理想路径,最终的目标既不是人被机器同化,也不是机器替代人类,而是构建高效、互补且具有批判反思能力的“人机共生”教育新生态。

总之,在“人一机”二元主体视角下,教学首先是一种古老且成熟的智能发展机制,其内在逻辑具有跨领域、跨主体的普遍适用性;其次,教学也是一种干预机制,会影响人的身心和个性发展、社会文化传承等过程。这表明,教学的理论建构与实践探索不应局限学校场域,而应拓展至更广泛的学习系统与智能发展过程(胡小勇等, 2026)。然而,囿于人类中心主义立场,教育学未能充分关注教育理论在类人智能领域的潜在价值,致使其难以扩展自身版图。同时,人工智能快速演进,但教育学多停留在技术应用表层,未能正面回应“智能如何生成与发展”这一更为根本的时代之问。这使教育学难以充分履行其“发展智能”的学科使命,亦构成其在人工智能时代所特有的理论遗憾。然而,这一困局孕育着转机:教学的对象首次从人类主体延展至机器主体,这一范式转换将为教育学重获理论生命力提供契机。

本研究正是基于这一理论关切, 尝试对传统教学理论隐含的人类中心主义预设进行批判, 并将被长期遮蔽的核心命题重新置于学科视野的中心: 智能既无法被“预设”, 亦不能被“灌输”, 它依赖于先天的“可学习性”, 并经后天的“可教育性”实现发展。概言之, 本研究旨在阐明一个事实: 在人工智能时代, 教育学既是人工智能技术应用的实践场域, 也是智能培育的重要手段, 能为促进人工智能发展提供教学框架。

[参考文献]

- [1] Anderson, J. R. (2013). The architecture of cognition[M]. New York: Psychology Press: 24.
- [2] Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., & Pascanu, R. (2018). Relational inductive biases, deep learning, and graph networks[J]. arXiv Preprint arXiv: 1806.01261.
- [3] Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning[C]//Proceedings of the 26th Annual International Conference on Machine Learning: 41-48.
- [4] Bieger, J., Thórisson, K. R., & Wang, P. (2015). Safe baby AGI[C]//International Conference on Artificial General Intelligence. Cham: Springer International Publishing: 46-49.
- [5] Chalmers, D. J. (2000). What is a neural correlate of consciousness[J]. Neural correlates of consciousness: Empirical and conceptual questions, 17.
- [6] 陈海虹, 黄彪, 刘锋, 陈文国(2017). 机器学习原理及应用[M]. 成都: 电子科技大学出版社: 2.
- [7] 陈明选, 徐家乐(2026). 数智时代教师的角色升维与能力重塑——从“知识传递者”到“智慧领路人”[J]. 电化教育研究, 47(5): 30-37.
- [8] Cole, M. (1998). Cultural psychology: A once and future discipline[M]. Cambridge: Harvard University Press: 48.
- [9] Dupoux, E., LeCun, Y., & Malik, J. (2026). Why AI systems don't learn and what to do about it: Lessons on autonomous learning from cognitive science[J]. arXiv Preprint arXiv: 2603.15381.
- [10] Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases[J]. Trends in Cognitive Sciences, 14(8): 357-364.
- [11] Han, H., Wang, Y., Shomer, H., Guo, K., Ding, J., Lei, Y., & Tang, J. (2024). Retrieval-augmented generation with graphs (graphrag) [J]. arXiv preprint arXiv: 2501.00309.
- [12] 郝祥军, 顾小清(2026). 跨越人际: 人机进行教育交往的结构与形态[J]. 中国远程教育, 46(3): 104-122.
- [13] 郝文武(2015). 主体间指导学习: 教育的最一般本质——教育本性从本体恒一到本质生成的再认识和新实践[J]. 山西大学学报(哲学社会科学版), 38(6): 82-90.
- [14] 胡小勇, 陶一, 毛江柔(2026). 超越人类中心: 智能时代教学主体观的后人类转向[J]. 中国远程教育, 2026, 46(5): 37-50.
- [15] 黄光扬(2002). 教育测量与评价[M]. 上海: 华东师范大学出版社: 56-62.
- [16] Kaufmann, T., Weng, P., Bengs, V., & Hüllermeier, E. (2024). A survey of reinforcement learning from human feedback[J]. Transactions on Machine Learning Research.
- [17] 康德(2011). 纯粹理性批判[M]. 李秋零译. 北京: 中国人民大学出版社: 59.
- [18] 拉尔夫·泰勒(1994). 课程与教学的基本原理[M]. 施良方译. 北京: 人民教育出版社: 80-90.
- [19] 莱斯利·瓦利安特(2026). 可教育性: 人工智能时代人类教育新理论[M]. 焦建利, 译. 北京: 电子工业出版社: 9, 35.
- [20] 黎加厚(2024). 生成式人工智能对课程教材教法的影响[J]. 课程. 教材. 教法, 44(2): 14-21.
- [21] 李博, 廖纹熠, 杨洪泽(2023). 服务设计视域下的“以学生为中心”教学改革路径[J]. 黑龙江教师发展学院学报, 42(12): 40-42.
- [22] 李永梅, 谭维智(2024). 从机器学习走向人机共学: 以语言为桥梁[J]. 开放教育研究, 30(6): 4-12.
- [23] Liu, Z., Bai, X., Chen, K., Chen, X., Li, X., Xiang, Y., & Zhang, M. (2025). A survey on the feedback mechanism of LLM-based AI agents[C]//Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence: 10582-10592.
- [24] 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.
- [25] 刘凯, 胡祥恩, 王培(2018a). 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.
- [26] 刘凯, 贾敏, 孙常新, 马玉慧, 王伟军(2023). 像教育人一样教育机器——人类教学经验能否提升通用人工智能系统的学习效果[J]. 电化教育研究, 44(9): 26-33+41.
- [27] 刘凯, 杨亚亚, 吴雨曦, 魏屹东(2025). 何为学习?——后人类时代学习主体的重构及普适性学习理论探索[J]. 开放教育研究, 31(6): 12-20.
- [28] 刘凯, 隆舟, 刘备备, 王伟军, 王培(2018b). 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.
- [29] Lu, H. A., Wei, Z. L., Zhang, Q., Zeng, J., Cao, B., Meng, L & Lam, W. (2026). Looped world models[J]. arXiv preprint arXiv: 2606.18208.
- [30] 麦克斯·班尼特(2025). 智能简史[M]. 林桥津, 译. 北京: 中译出版社: 355.
- [31] 米罗斯拉夫·库巴特(2016). 机器学习导论[M]. 王勇, 仲国强, 孙鑫, 译. 北京: 机械工业出版社: 56-161.
- [32] Muller, J., Hoadley, U. (2021). A pedagogic compact: retrieving 'powerful' educational knowledge from Didaktik and curriculum studies[J]. Journal of Curriculum Studies, 53(2): 166-178.
- [33] Paton, N. W., Chen, J., & Wu, Z. (2023). Dataset discovery and exploration: A survey[J]. ACM Computing Surveys, 56(4): 1-37.

- [34] 皮亚杰(1981). 发生认识论原理 [M]. 王宪钊, 译. 北京: 商务印书馆: 61-87.
- [35] Turing, A. M.(1950) . Computing Machinery and Intelligence[J]. Mind, 59(236): 33-60.
- [36] Vardi, G.(2023) . On the implicit bias in deep-learning algorithms[J]. Communications of the ACM, 66(6): 86-93.
- [37] 汪琼, 李文超(2021). 人工智能助力因材施教: 实践误区与对策 [J]. 现代远程教育研究, 33(3): 12-17+43.
- [38] 王本陆(2023). 教材论 [M]. 北京: 中国高等教育出版社: 129-130.
- [39] 王一岩, 郑永和(2022). 智能时代的人机协同学习: 价值内涵、表征形态与实践进路 [J]. 中国电化教育, (9): 90-97.
- [40] 王伟军, 贾敏, 刘凯, 邓胜利(2026). 人本人工智能驱动的数智信息服务人智关系重塑 [J/OL]. 中国图书馆学报, 1-14[05-08].
- [41] 魏屹东(2017). 语境同一论: 科学表征问题的一种解答 [J]. 中国社会科学, (6): 42-59+206.
- [42] 吴砥, 郭庆, 李佳平(2025). 智能技术进步何以赋能教师发展 [J]. 教育研究, 46(5): 43-54.
- [43] 项贤明(2026). 人工智能时代教育何为 [J]. 教育研究, 47(3): 89-101.
- [44] 约翰·哈蒂(2015). 可见的学习 [M]. 彭正梅, 译. 北京: 教育科学出版社: 84-85.
- [45] 周志华(2016). 机器学习 [M]. 北京: 清华大学出版社: 12-18.
- [46] Zhu X, Singla A, Zilles S. (2018) . An overview of machine teaching[J]. arXiv preprint arXiv: 1801.05927.
- [47] 祝修理, 项贤明(2026). 教学本质新论: 一种互指共生的交往实践——基于交往实践观与泛教育论的双重理论视角 [J]. 现代远程教育研究, 38(3): 29-38.
- [48] 左晓梅, 曹培杰. 人机共生, 师道弥新: 智能时代教师的“变”与“不变” [J]. 中国电化教育, 2026, (7): 25-27.

(编辑: 赵晓丽)

The Learnability and Educability of Machine: On the Cross-Subject Isomorphism of Teaching

LIU Kai^{1,2}, YANG Yaya^{1,2}, WU Yuxi^{1,2} & Xiang Xianming³

- (1. Department of Educational Science, Bohai University, Jingzhou 121013, China;
2. Institute of Artificial General Intelligence of Bohai University, Jingzhou 121013, China;
3. School of Education Science, Nanjing Normal University, Nanjing 210097, China)

Abstract: Within the “human-machine” dual-subject framework, teaching as a distinctive phenomenon is widely observed in the arena of machine learning. However, constrained by anthropocentrism, traditional pedagogical theories struggle to account for such cross-agent practices, resulting in a growing disjuncture between the expanding empirical scope of teaching and the rigid boundaries of existing theoretical explanations. From an educational perspective, machine learning entails complete teacher-student roles and instructional interactions. The underlying mechanism resides in two constitutive properties of machine learning agents: learnability, which serves as the intrinsic foundation for intelligence development, and educability, which manifests as external intervention directed at learning capacities. Accordingly, human and machine teaching exhibit structural isomorphism across two dimensions. At the a priori level, learnability is expressed through motivation, representation, and information-processing capabilities; at the a posteriori level, educability is realized through intersubjectivity-mediated, targeted intervention in and expansion of learning capacities. Building on this analysis, this paper advocates the proposition of cross-subject isomorphism in teaching. By integrating insights from machine-based teaching into the traditional pedagogical theory, it seeks to move beyond the unidirectional logic of AI-enabled education, and thereby fostering a paradigmatic transformation of the discipline in the age of AI.

Key words: Machine Learning; Machine Teaching; Machine Education; Educability; Isomorphism