

# 人工智能之于教育决策的主体性 信任危机及其规避

孙立会

(中央民族大学教育学院,北京 100081)

**[摘要]** 人工智能为教学注入新生力量的同时也引发了诸多反思。算法偏见、隐私侵犯及情感忽视触发了智能教育中人机信任危机,进而导致决策自主权虚置。人工智能教学决策中应凸显“人”之主体性,建立由教师信心主导的决策模式能帮助教师挣脱技术中心的拜物主义泥沼,实现决策独立立场,并释放错误归因的心理枷锁,坚守教学主体责任,同时发挥教学的感通性,降低智能情感依赖。班杜拉三元交互决定论为教师决策效能的影响要素解构提供了分析框架,教师智能认知误区和职业危机感主导了信心机制的建立,自身行为的成败及替代经验连接着教学的信心与行为。人一机—智能情境因素各主体的能量转移影响着系统的和谐交互。由此,从技术主体、关系维系与机制保障三层面设计教学决策中“人”之主体性的实现路径,建立解释与理解双向发展的智能教育体系,着力塑造人一机—情境协同共生的智能教学互信生态,并不断完善智能决策伦理的法律法规制度建设,重启人工智能教育信任机制。

**[关键词]** 人工智能;教学决策;信任危机;主体自信

**[中图分类号]** G434 **[文献标志码]** A

**[作者简介]** 孙立会(1985—),男,吉林白城人。副教授,博士,主要从事人工智能教育研究。E-mail:sunlh777@163.com。

## 一、引言

人工智能的发展历经数次摇曳与停摆,但众多研究者对图灵预设的机器智能的追求从未停滞。2016年,AlphaGo击败李世石重启了人工智能热潮。现今,人工智能所涵盖的深度神经网络、自主计算、机器感知等关键算法使得机器具有了超强的“类人”智能性,由此也引发了人类对人工智能应用中人的价值理性与机器的工具性之间主体地位的深刻反思,其根本则是人工智能与人之主体之间的“信任”问题。伴随智能教学应用的普及,智能教学决策中的信任危机也受到研究者普遍关注,如何在智能教学中精准决策、规避风险,以发挥人工智能技术在教育应用中的最佳教学效果成为技术研发者和教学实施者们的一致追求。目前研究多将智能教学应用中的各类问题归咎于数据算法,以至于不断追求突破技术限度,包装美化智能机器决策可信度“外衣”,却普遍忽视了“人”在教学

决策中的决定性作用,尤其是教学主体“人”在决策中独立于智能机器之外的自我信心这一心理过程要素的作用。因此,探索教师在智能教学决策中自信心形成的心理机制是未来人工智能教学应用发展的核心议题。

## 二、问题导向:人工智能教育应用中信任危机的触发

人工智能赋能教学的同时,也引发了对其应用后果的深刻反思。正如斯蒂芬·威廉·霍金(Stephen William Hawking)所言:创造人工智能的成功将是人类历史上最大的事件。不幸的是,也可能是最后一次,除非我们学会如何规避风险。人与人工智能间的交互效果由“信任”介导,应用过程中出现的算法问题却触发了教学决策的信任危机。

### (一)技术危机:算法偏见导致教学决策失衡

人工智能以数据、算力和算法为三大基本要素,

算法是系统运行的核心遵循。算法如今以多种形式影响着人类生活的方方面面,并在教学实践中实现了广泛应用,如辍学预警系统、作业自动化评分系统、招生数据采集系统等。伴随人工智能技术在教育中应用的不断深入,系统中由于算法偏见导致教学决策失衡的问题引发了普遍重视。算法偏见是指在算法开发和应用中,开发设计人员将自身和社会偏见集成到数据采集和处理过程中,进而产生了对特定群体的歧视和偏见性的预测或推论<sup>[1]</sup>。教育人工智能系统的算法偏见设计体现种族、国别和性别偏见等诸多方面。有研究表明,毕业预测模型算法面对不同种族学生的预测效果存在差异,该模型对白人学生的误报率较低,而对拉丁裔学生的误报率较高<sup>[2]</sup>;自动化评分系统相较于教师人工评分更具高效性,但却被发现其对白人和非裔美国学生的评分更加准确<sup>[3]</sup>;性别引发的偏见同样存在于人工智能教学系统中,如有研究报告了用于预测学生课程通过率的算法系统对男生的预测效果较差<sup>[4]</sup>。无论互联网络时代的“信息茧房”还是当今智能浸润时代的“过滤泡”,无不体现算法对人认识真实完整世界的认知“窄化”,算法需求取向的智能推送是对“人”之主体性价值选择权的严重挑战。智能系统的算法发展如何,其都是数据结果的计算、排除和呈现方式的上位规划,更是技术构建者价值观的客观反映,“人”在与智能系统算法交互中的“任其发展”与“自我放弃”造成了算法偏见,也由此成为触发人工智能教育系统决策信任危机的动因之一。

### (二)伦理危机:隐私侵犯引发道德决策风险

个性化学习、自动化评估及监督管理系统为教育教学带来科学性和便利性的同时,也引发了人工智能教学应用的伦理和道德省思。智能教学系统将师生置于马克·波斯特(Mark Poster)所说的“超级全景监狱”之下<sup>[5]</sup>,一切的学习和教学行为活动数据都变得开放透明而无所遁形。教育数据的开放共享是打破信息孤岛、数据垄断等问题壁垒,助推精准决策的基础要素,但数据开放随之带来的便是个人信息过度暴露造成的隐私侵犯等问题。学生和教师隐私保护问题是目前围绕中小学人工智能教学应用的伦理隐患之一。目前,智慧校园建设中的人脸识别技术服务于宿舍、图书馆和课堂出勤等日常教学生活管理,精准记录学生的学习时长和活动时间;更有智慧课堂行为管理系统分析学生学习行为(阅读、举手、书写、起立等)以及情绪状态(高兴、伤心、愤怒)以判断和分析学生的学习模式<sup>[6]</sup>。这些数据使用得当可以便利教学,但如若未能得到有效的管理和规范的准则约束便会对学生

隐私的侵犯,而学校普遍默认学习数据使用的知情同意权,并且也多忽视考虑教学监控和跟踪中学生表现失真和不安情绪等问题。由此看来,隐私侵犯问题的实质是人工智能过度“入侵”学习者个人空间而使其产生担心领域“失守”的心理状态,而这种“想使用但又怕受到伤害”的心理势必会埋下人与人工智能之间“怀疑的种子”,造成智能教学决策的信任危机。

### (三)情感危机:情感交互忽视造成认知决策偏差

人工智能教育应用以其对教学与管理中数据的全面挖掘和客观分析而被视为科学教学决策的保障,但“冰冷”的数字和机器是否能代表教学本质委实需要长期的实践考量,人工智能在教学中的应用多表现为帮助师生完成“功能性”的任务,而忽视了师生“情感”领域的信息,尤其对智能环境中师生间情感交互问题的关注。情绪和情感是人对客观外界事物态度的主观体验和相应的行为反应,表明了主体需要与客观外界事物间的关系,由个体主观体验、面部、姿态和语调表情等外部表现及呼吸、心跳、血液循环和皮肤电反应等生理唤醒作为其具象化体现。智能教学机器遵循上述原则设计算法通过捕捉学习者生理和行为信息以解释学习者实时的情感体验。如利用神经网络模型捕捉学生面部生理表情以分析学生课堂中的情绪反应,并与教师相应时段的教学内容相联系以对教学效果作出研判与改进<sup>[7]</sup>。近来兴起的情感计算研究则试图创建一种能感知、识别和理解人的情感,并能针对人的情感作出智能、灵敏、友好反应的计算系统<sup>[8]</sup>。单向度的情感识别和分析缺乏对教学主体交互的关注,情绪可以被捕捉,但情感却难以被计算。或许这正如“莫拉维克悖论”所喻:对人工智能系统而言,高级推理需要的计算量反而少于低级无意识认知。教学活动作为一种鲜活的生命体验尽显无意识的“细节之美”,师生互动中所表现出的关怀、尊重、信任、同理和共情等真实的情感流露是否能通过系统“计算”出来目前还未十分明确,这也导致了人机教学决策中的信任危机。

## 三、视角转向:智能教学决策中“人”之主体自信的建立何以必要

机器“智能化”精度的不断升级同时也助推教学中人向“机械化”执行者的角色转变。究其根本,人工智能教学决策中信任危机的产生源于人对人工智能的信任度和人工智能可信度的错位匹配。人对人工智能的信任随着智能体在教学决策中的功能性表现而动态改变,从而使教师产生对智能辅助决策的信任程度

的预判,并受此影响可能产生错误归因的恶性循环。智能教育应用的“虚幻”迷局中,教师与机器的交互关系是其本质所在,信任作为一个中观概念,一个方向连接“人”的微观心理活动,另一个方向连接宏观智能环境,使得教学决策的责任左右摇摆,而要发挥人工智能教学的优势则必须将责任天平摆向“人”,人要学会在“纠缠”的人机关系中独善其身,发挥人之独立体信心之源的决策,以此彰显教学决策中教师“人之为人”的必要性和价值所在。

### (一)从场依赖到场独立:挣脱拜“物”主义的泥沼

化解人工智能教学决策危机,教师应当实现从智能体依赖型决策到依靠主体信心决策方式的转向,即对教学效果的预判不应由智能机器的工具理性决定,而教师需要具备的是:无论智能机器预测效果的好坏,都能有信心、有能力从人的“价值理性”的向善维度把控教学全局,作出精准的抉择与判断。正如赫尔曼·威特金(Herman A. Witkin)所提出对“场独立”和“场依存”两类认知方式的区分。场依存型倾向于以外界参照作为内部信息加工依据,场独立型学习者的内部信息加工较少受外界参照物的影响。同样,“场依存”的教学决策者在智能教育情境中普遍存在。人类模拟复制人的思考方式而创造了人工智能,而在应用中却深陷于其影响“无法自拔”,在自身不自知的情况下认知和思维方式受到“钳制”,以至于形成拜“物”主义倾向。教师在智能环境中的教学决策判断亦是如此,研究显示,教师对人工智能的信任会随着智能机器的性能动态演变,即人工智能系统表现好,教师对其信任无变化,但如若其表现不好,则对其信任程度急剧下降,并其信任程度会随着智能机器的表现下降而持续下降<sup>[9]</sup>,这种“以物悲喜”的信任决策心理机制将会严重制约教师自身的教学判断。由此看来,挣脱智能机器之“物”的治理泥沼,以教师自身信心为目标,建立场独立性教学决策机制势在必行。

### (二)从责任外推到责任归位:释放错误归因的心理枷锁

技术辅助课堂教学的责任归属问题从未像人工智能在教学中的应用这般饱受热议,试想传统的信息技术辅助课堂教学,教师几乎不会将课堂中的教学问题归咎于教学课件和媒体资源,但这种责任归因方式对于人工智能教学应用则不再成立。面对教学智能体,教师似乎更加容易心安理得地将教学责任外推。安德烈亚斯·马蒂亚斯(Andreas Matthias)指出,这可能是由于人工智能系统的使用者无法预测其运行后果,甚至无法干涉其运行<sup>[10]</sup>。诚然,人工智能系统

的“类人”设计原则以及训练数据的分布不同使得其行为更具自主性和不可解释性,更有研究者声称人工智能系统的自主性已足以达到将其视为与人类同等程度的道德主体<sup>[11]</sup>,但这种放弃自身责任主体地位的想法无疑是可怕的,更加容易使教师在决策中迷失自我,产生不当责任归因进而影响后续的教学动机与行为。如研究表明,教师如接受人工智能的建议得到积极反馈,则教师对人工智能的信任增强,自信心没有变化;拒绝人工智能得到积极反馈,则对人工智能信任下降,自信心提升,但教师对人工智能的信任程度并不会影响教师接受或拒绝人工智能的决定,真正具有决定意义的则是教师在决策中自身的信心<sup>[9]</sup>。借用伯纳德·韦纳(B. Weiner)归因理论中的“三维度模型”,人工智能教学决策中教师主体自信心作为内部可控不稳定性因素更加有利于教师教学决策。

### (三)从情感依赖到情感独立:厚植“人之为人”的感通性

教师的个人情感是建立对人工智能信任的原初认知,目前大多关于人工智能信任的研究集中于从认知方式入手提高使用者的信任程度,而对情感因素在信任决策中的作用机理少有探讨。人类对类人机器人的依恋问题是影响人机交互系统中信任建立的关键。情感心理学认为,人在责任和压力环境下会本能产生“转嫁”心理,使得人们更可能将自主和智能特性“投射”于机器之上,在其内心建立依恋纽带,进而产生“同行安全感”和“责任共担感”,以缓解和减轻教学焦虑<sup>[12]</sup>。人对人工智能的过度情感依赖问题是以智能机器行为导向决策的主要表现形式,同时过度的情感依赖容易导致“讨好型人格”的形成。于全景式的教学监控之下,师生难免产生迎合心态,粉饰自身行为以符合既定算法标准,甚至养成“表演型人格”<sup>[13]</sup>,使得教师在教学决策中“迷失”自我。人工智能研究和发展即使再进一步深入,其根本原则依旧是数据、算法上的择优,而非情感、道德的选择,人机交互中人得以凸显出来的依旧是人的心性特点<sup>[14]</sup>。因此,亟须建立以教师主体自信为导向的决策心理机制,实现从人对智能体情感依赖到情感独立的转向,发挥“人之为人”的感通性,在教学决策中以教师个人信心为中心连接和互动,以体现人工智能辅助教学中人的“切身参与性”和“实践主体性”。

## 四、要素解构:智能教学决策中“人”之主体自信发生机制

智能教学环境中,教师信心主导的独立教学决策

在行为上体现为教师个体对其是否有能力完成某一行为所进行的判断,同样也是个体自我效能的一种表现。自我效能现象作为班杜拉社会学习理论中三元交互决定论的经典内容,正是社会学习理论对人的主体因素及作用机制的有效彰显与对人性及因果决定模型的理智把握。三元交互决定论中对环境、行为与人的主体因素对自我效能的影响为我们提供了剖析智能教学决策中人之主体自信发生机制的解构框架。

### (一)个人主体因素的主导:认知误区和职业危机感

影响智能教学中教师决策信心之“人”的主体因素指教师在使用智能技术辅助教学过程中的个体认知及情感的内部心理变化,而教师对人工智能的认知和情感体验对其把握教学节奏,实现精准教学预判具有决定性作用。教师对课堂教学中智能系统科学性理性认知是其决策自信的基本前提,目前部分教师对人工智能及教学应用存在认知误区,影响了其是否接受利用人工智能系统开展教学,这种认知误区主要表现在对人工智能的概念和功能层面的误解。人工智能的表层意蕴为其披上了“类人智力”的外衣,并使其拥有了“超级计算”的属性,这同时也在心理为其树立了“强大且神秘”的形象,从而可能会致使教师产生“过度崇拜”的盲目信任或“恐惧担忧”的拒绝使用两类极端反应。人工智能教育应用并没有如此高端神秘,作为教学资源、辅助管理、批改作业、智能答疑等功能性程序系统,早已渗透至学习的各个层面。因此,排除教学者对人工智能的认知障碍是教师信心建立的基本前提,但无论人工智能发展程度几何,使用者应警惕的永远是其智能的“灵魂”,而非机器的“外壳”。

教师于智能环境中的情感状态也制约其决策自信心的建立<sup>[15]</sup>。人工智能技术“先入为主”、高效精准的特征使教师在与之交互中也不断萌生对自身教学地位的隐忧,从而产生未来被人工智能所取代的职业危机感。危机感产生于教师内心“失踪”的安全感,安全感的缺位恰是教师自身不自信的重要表现。究其根本,笔者认为,此境遇是人工智能教育应用场景中教师主体的“错位比较”造成,以人的“有限性”去对抗人工智能的“无限性”,从而放弃了教育中对人主动权的“自守”。如与人类教学执行者相比,人工智能系统的快速存储和海量资源传递以及重复机械性教学行为操作和“不知疲倦”的监督管理等功能确实具有不可比拟的优越性。但教师于教育语境中触及心灵善美的追逐与求真探索的深切体验之“道”的地位是目前智能应用之“器”难以企及的,即从教师心灵层面生

发的对教学对象的影响是其永不湮灭、恒久不断的使命所在。由此看来,即使生理意义上的人存在被彻底还原的可能,但哲学意义上人类的心灵不可能被彻底还原,秉承着人工智能的教学应用应当只是“代具”而非“代替”<sup>[16]</sup>,在教学中“借力”而非“依赖”人工智能是自我决策信心建立的心理基础。

### (二)行为效果因素的连接:自身行为的成败及替代经验

三元交互决定论中的行为因素指的是个体在环境因素和主体人因素的作用下进行的活动,分为实际行为和潜在行为。实际行为一般指个体亲身参与的活动及产生的延迟影响效应;潜在行为可理解为个体心理及学习的行为准备及受先验逻辑和榜样示范作用主导的心理倾向,将其投射到影响智能环境中教师决策信心的影响因素可分为教师主体与人工智能交互活动的成败经验及观察榜样行为的替代性经验。教师个体在人机交互教学中个人活动行为的成败经验这一效能信息源对其自身决策信心建立的影响最大。换言之,成功经验会提升效能期望,反复的失败经验则会降低个体的效能期望。由是,教师与人工智能教学系统的交互积极性“初体验”十分重要。与此同时,教师对人工智能教学失败经验的正确归因也是决策信心建立的有效保障。积极的归因方式应将成功归因于教师内部因素,增加教师智能教学中的自我价值认同;同时善于将失败归于外因,减少责任负担感,达到信心自我保护机制建立的目的。

榜样行为的替代经验也是影响教师决策信心建立的因素之一。人行为的效能期望大多来源于观察他人的替代性经验,即教师观摩和借鉴示范者的成功经验,以帮助建立成功把控人工智能教学及独立决策的信心,而这一行为结果与教师群体间的经验交流方式密切相关。教师在“经验迁移”的过程中需审慎的关键环节即注意将自身真实的教学情境与榜样示范的问题模式及情境框架进行迁移比较。同时,替代性经验帮助决策时还应充分考量两方面因素的干扰效应:一则是“易得性心理偏差”的干预,即人们总是倾向于作出容易的决定而非有价值的决定,这与该经验相关的认知提取的难易程度相关,如教师在教学观摩中的感官体验与参与程度决定了其是否会认同并复制这一教学行为,同时观摩活动与自身的联系程度及熟悉程度也会影响教师对人工智能的选择和使用;二则应防范的是观察替代性经验的成功与否而非切身体验可能会使教师对人工智能教学应用产生“先入为主”的心理体验,教师可能从内心分化出抵触和排斥

人工智能的意志情感,并主观弱化与人工智能之间的连接,使得教师和人工智能产生非教育性疏离感,从而使教师产生对技术教学应用的消极心理体验和被无用感、无价值感等低落心境所支配的被动技术应用行为<sup>[7]</sup>,这同样会限制教师主体决策自信的发生与建立。因此,替代性经验的观察学习应谨记个体并非从经验行为中学习,而是通过对经验的反思性建设而学习。

### (三)智能环境因素关系转移:人、机、环境三者交互能量的“此消彼长”

三元交互决定论指出,人同时是环境的产物和营造者<sup>[8]</sup>。主体人作为环境的营造者,通过自身的选择定义环境,并通过自身的行为来改变环境。行为、人和环境因素相互连接并相互作用以产生影响,“交互”体现了事物之间的相互作用,而“决定”则表现为“事物间交互影响的产物”。由此,笔者认为在智能教育情境中,“环境”应当被看作是一种人机与周围环境“关系”的具体体现,这种关系则由教学生态中三者之间的交互活动所决定,不同情境的交互体现了三者关系网络中能量的“转化和迁移”。教师如何有效地适应和处理不同人工智能教学情境关系网络中各主体间能量的“此消彼长”是其决策信心发生和建立的关键,即教师主体在智能教学环境中应找准自身定位,处理好与学生和人工智能两类“环境”之间的关系。技术“入驻”教育引发了教学的主体性功能之“辩”,而智能技术因其“智化”特征,在处理其关系中还应审慎考量伦理责任之“思”。教师与智能技术的关系处理是其主体决策自信建立的逻辑起点。人工智能教育环境中,教师只有坚守本心,杜绝教学自主决策权的“放逐”与“让渡”,才是规避风险有效决策的肯綮之所在。

师生关系问题同样也是影响教师决策自信建立的主要因素之一。智能教学环境中阻碍教师决策自信机制建立的师生关系问题呈现出“松散和紧密”两极分化的状态。“紧密”一面表现为过度“让渡”,即智能的加入使教师角色可有可无。如学生通过人工智能自适应学习系统开展自主学习,并基于对自身知识储备、学习风格和认知发展的反馈及时调整自身状态,并且智能系统也通过“学习”实现对资源更加精准的匹配与推荐。此学习形式从目标确定、内容选取及学习评价过程全程都无需教师参与,教师教学耳濡目染的言行示范和潜移默化的支持教导退出了教学舞台,从而造成了教学中师生的“角色失位”<sup>[9]</sup>。“紧密”的另一面则表现为过度“控制”,智能语言、图像和行为识别能够采集并分辨学生的各类学习状态,这意味着学生学习的一言一行与变化发展都在教师的严密掌控

之下,使师生演化为绝对的“施控者”和“受控者”,使教学中师生交互的分寸和界限逐渐模糊,更甚者弱化了教育教学的价值。因此,教师在智能教学中有效决策和信心建立的“环境”保障是良好师生关系的建立与维系,这种良好关系建立的前提应重点关注教学中师生的情感交流与表达,挖掘教学中师生间有“温度”的互动关系,加强师生主体间的对话交流,以营造和谐共生的师生关系环境。

## 五、信任重启:智能教学决策“人”之主体自信建设机制

智能教学环境需要培育以自身信心为主导的优秀决策者,坚守教学的教育性为基本准绳,凸显决策中“人”的主体性作用;应从技术共生的视角切入,建立“解释与理解”双向发展的智能技术体系,并考量人、机、环境三者和谐互生的关系网络,不断完善智能决策机制的外部建制保障,实现顶层规划与整体布局,重启人机交互中的信任机制。

### (一)技术共生:建立“解释与理解”双向发展的智能教育体系

智能教学环境中突出人之主体自信在决策中的重要作用,建立由教师主导的决策信心机制的首要前提是建立人工智能的“主动解释”和人类教师“主动理解”双向发展的教学体系。不可否认,智能体作为教师的首要互动对象,其数据基础和算法程序的不断精细化发展,对加强人机信任程度,实现更加精准的教学决策有着重要意义。技术开发者应不断推动并追求可解释性人工智能的发展。已有研究表明,信任与人工智能的可解释性呈现积极的正相关,算法的透明度所反映的基本操作规则和内部逻辑在信任机制的建立中至关重要<sup>[20]</sup>。可解释性关系建立的诉求与人工智能的算法“黑箱”效应对应,人工智能应用系统的最大难题即系统使用者甚至程序开发者都难以对系统输入和输出数据中间内部算法的工作机制有着绝对精准的把握,尤其对于无监督学习无法提供训练数据,机器通过聚类发现数据中心的相似群自行“学习”,从而使得其输出结果更具不确定性。多数教育工作者更是对人工智能的原理处于“一无所知”的状态,此过程中产生的问题很容易被归咎为算法的不透明性及机器数据自我训练的未知性。可解释性人工智能的目标便是使系统的预测和行为更易被理解,其在因果关系推理层面的突破成为破除人工智能算法“黑箱”壁垒的关键。智能系统模型大多基于相关关系展开决策分析,而没有揭示对象属性间的因果机理,因此,积极推

动发展基于反事实推理的因果表征学习关系可能是人工智能技术发展的破局之作。

人工智能“可解释性”的发展也意味着其“可理解性”的增强,人机关系的建立表现为机器向人靠拢,由“智”趋向“人”的一侧。教师也应向人工智能“主动奔赴”,积极加强对人工智能技术的学习和理解,形成人机双向接近的互动生态,以此建立对主体人决策信心的保障支持。研究显示,支持智能技术的可靠性和使用者的能力是人与人工智能信任发展的两个重要维度,其中,人对智能系统的理解成了制约两者对人工智能信任影响关系的最强干预因素<sup>[21]</sup>。课堂教学中用不用人工智能已然成为一个“假命题”,而如何在教学中更好地应用人工智能才是一个真问题。并非“逃避”人工智能就能不受其影响,教师应做的是积极做好心理与行为建设,避免陷入技术应用的心理窘境和慌乱无措感支配的被动应用中。教师在行动层面应主动了解并学习人工智能教学专业知识,积极参与人工智能科普教育类与专题培训类活动,在亲身实践和与优秀教师教学交流切磋中提升教师使用人工智能的获得感和成就感,明晰教学中人与人工智能的责任边界。

### (二)关系互生:打造人、机、情境协同共生的智能教学互信生态

教师信心为主导的教学决策模式表现为教师对教学整体的把控有信心时接受人工智能的建议,而在没有信心时则选择拒绝人工智能系统的建议,这种“心无外物”的决策方式是人之为人的“独立姿态”的集中反应,这种独立和强大并不是远离外物的“傲慢与孤僻”,而是与其周围事物的“适应与和解”,即能够恰当地处理与各主体之间的关系。人、机和问题情境构成了智能教育生态网络,建立并发挥由教学主体信心主导的决策效能的关键在于营造教师、人工智能及教学问题情境三者和谐共生的“互信生态”,建立并打造互信共同体。定义角色以实现责任共担,教师应在不同的教学情境中对人工智能的角色进行心理设定,以保障其在智能教学中调整并维持最适交互状态。人工智能的角色定义一般可划分为三种类型,从亲密程度排列依次为敌人、工具和伙伴,通过不同情境中人工智能所“扮演”的角色归于这三类中的某一类而展开分析或批判,教师不应将某一教学情境中的人工智能角色定义为固定的某一类别,教师应学会在不同的活动行为中做到适时地“转换”,应以对“敌人”的戒备状态审慎人工智能教学应用的风险和危机,以“工具”辅助视角助力教学有效开展,并以“伙伴”合作心态正

视教学中人与人工智能的关系,以夯实教学各环节中的责任主体地位。

互通有无以实现价值共享。实现人工智能与教学主体的利益共联,其程序设计应将教师和学生的利益置于首位,并保持与人类相近的价值观。如研究者指出,机器若想获取人类的信任,必须通过反复“沟通”了解人类的价值取向(意图和信念),并不断向人类解释其每一步决策的意图,而实现这一“双向沟通”的关键是建立心智模型以建模人类意图,简而言之,即将人类意图解释为机器能够理解的形式,以加强机器自身建设而增强人机互信。机器“揣摩”人类心智减少人机之间的认知思维差异以加强互信的实现目前尚有难度。笔者认为可转换角度,从文化的土壤中挖掘人机物互通的交流方式,以人工智能时代万物主体和谐共生的方式为其注入灵魂。中华文化历来崇尚“天人合一”的一统思想和“中庸为德”的无为之道,而反对“绝对”和“分裂”的两极化思想,引鉴中华文化中“一”之精髓以审度人工智能教学系统生态中各主体的关系互生,以整体观、系统观、和谐观和全局观映照“人”于人工智能参与的大系统整体<sup>[22]</sup>,发挥教学中人之主体性,敬畏技术,敬畏自然,努力建构互信生态。

### (三)创生共建:完善智能决策伦理的法律法规和制度建设

智能教学环境中教师信心主导的决策机制充分释放了“人”的自由意志,这一信任机制的建立同时也需外部建制的完善。人工智能在六十多年的发展中几经沉浮,由此演化出了强弱人工智能的分说。教育领域甚至各领域中多由弱人工智能主导,即利用人工程序系统解决特定教育问题,但强人工智能的实现一直是研发者和应用者的终极目标。强人工智能的技术逻辑是能够自主学习推理解决问题且具备自我意识的智能体,关于人工智能的伦理审思和法律权利的探讨也由此发轫。2021年,科技部颁发了《新一代人工智能伦理规范》,提出应充分考虑当前社会各界人工智能的隐私、偏见、启示和公平等伦理问题,并提倡善用,防止滥用,促进人工智能的健康发展。虽取得一定进步,但更加完善的社会建制尚待不断探索。

与此同时,面向强人工智能自主意识的发展目标的责任和权利主体之争也甚嚣尘上。人工智能的注入“扰乱”了原有的法理体系,亟需法律体系内部的整体反思与全面整饬,在人工智能相关法律的制定中应及早扼断法律层面对人工智能应用的完全人格、责任主体性的倾注趋势,坚守以“人”为主体性的价值审判立场。教育教学与司法审判同样都表现为人性、道德和

人的组织性交织的活动特性。教学主体在教学内容、学情分析、学习评价等活动中也应持有绝对的自主决策权,教师绝不应将其自主权“让渡”给人工智能。应

谨记:人工智能未来无论发展如何,其工具性属性不会改变,教育过程中人的主体性决策是永恒不可替代的意志行为。

#### [参考文献]

- [1] BAKER R S, HAWN A. Algorithmic bias in education [J]. *International journal of artificial intelligence in education*, 2022, 32: 1052-1092.
- [2] 李世瑾,胡艺龄,顾小清. 如何走出人工智能教育风险的困局:现象、成因及应对[J]. *电化教育研究*, 2021(7):19-25.
- [3] BRIDGEMAN B, TRAPANI C, ATTALI Y. Comparison of human and machine scoring of essays: differences by gender, ethnicity, and country[J]. *Applied measurement in education*, 2012, 25(1): 27-40
- [4] HU Q, RANGWALA H. Towards fair educational data mining: a case study on detecting at-risk students[C]//*Proceedings of The 13th International Conference on Educational Data Mining*, Jul 10-13, 2020. Morocco: EDM, 431-437.
- [5] 马克·波斯特.信息方式[M]. 范静哗,译.北京:商务印书馆,2000:17.
- [6] AKGUN S, GREENHOW C. Artificial intelligence in education: addressing ethical challenges in K-12 settings [J]. *AI and ethics*, 2022, 2: 431-440.
- [7] 孙立会,王晓倩. 人工智能之于教育的未来图景:机器行为学视角[J]. *中国电化教育*, 2022(4):48-70.
- [8] 徐振国.教育智能体的发展历程、应用现状与未来展望[J].*电化教育研究*, 2021, 42(11):20-33.
- [9] CHONG L, ZHANG G, GOUCHER-LAMBER K, et al. Human confidence in artificial intelligence and in themselves: the evolution and impact of confidence on adoption of AI advice[J]. *Computers in human behavior*, 2021, 127:1-10.
- [10] MATTHIAS A. The responsibility gap: ascribing responsibility for the actions of learning automata [J]. *Ethics and information technology*, 2004, 6(3): 175-183.
- [11] 刘博韬, 金香花. 人工智能的责任鸿沟——基于归因过程的分析[J].*道德与文明*, 2021(6):115-131.
- [12] GILLATH O, AI T, BRANICKY M, et al. Attachment and trust in artificial intelligence [J]. *Computers in human behavior*, 2021, 115:1-10.
- [13] 冯锐, 孙佳晶 孙发勤. 人工智能在教育应用中的伦理风险与理性抉择[J].*远程教育杂志*, 2020, 38(3):47-54.
- [14] 湛衡. 从“人禽之辨”到“人机之辨”——人工智能时代教育的成“人”属性[J].*全球教育展望*, 2021, 50(10):35-46.
- [15] 刘清堂,李小娟,谢魁,等. 多模态学习分析实证研究的发展与展望[J]. *电化教育研究*, 2022, 43(1):71-78.
- [16] 张务农.人工智能时代教学主体的辨识与伦理反思[J].*教育研究*, 2022, 43(1), 81-90.
- [17] 李芒, 申静洁.论教师教学的信息技术疏离感[J].*中国电化教育*, 2020(5):62-68.
- [18] BANDURA A, CERVONE D. Differential engagement in self-reactive influences in cognitively-based motivation [J]. *Organizational behavior and human decision processes*, 1986, 38(1):92-113.
- [19] 赵磊磊,陈祥梅,杜心月.人工智能时代师生关系构建:现实挑战与应然转向[J].*教育理论与实践*, 2021, 41(31):36-41.
- [20] COECKLBERGH M. Artificial intelligence, responsibility attribution, and a relational justification of explain ability [J]. *Science and engineering ethics*, 2020, 26:2051-2068.
- [21] BALFE N, SHARPLES S, WILSON J. Understanding is key: an analysis of factors pertaining to trust in a real-world automation system[J]. *Hum factors*, 2018, 60(4):477-495.
- [22] 王萍萍.人工智能时代机器人的伦理关怀探析——以《老子》“善”论为视角[J].*自然辩证法研究*, 2021, 37(5):54-59.

## Artificial Intelligence in Educational Decision-making Subjectivity Trust Crisis and Its Avoidance

SUN Lihui

(School of Education, Minzu University of China, Beijing 100081)

(下转第43页)

the knowing of artificial intelligence and the substitution of artificial intelligence for the process of human knowing will limit human subjectivity. (3) Understanding human subjectivity in terms of non-knowing can provide a reasonable solution to the dispute over the theory of teaching subject, which in turn can develop the theory of teaching subject, and can also provide a thorough dialectical materialism perspective for the establishment of the theory of teaching subject in the era of artificial intelligence.

[Keywords] Not-knowing; Teaching Subject; Artificial Intelligence; Subject Philosophy; Philosophy of Educational Technology

---

(上接第 27 页)

[Abstract] While artificial intelligence (AI) has brought new energy into teaching and learning, it has also caused a lot of rethinking. Algorithm bias, privacy invasion and emotional neglect have triggered man-machine trust crisis in intelligent education, which leads to a lack of decision-making autonomy. The subjectivity of "human" should be highlighted in the decision-making of artificial intelligence teaching. The establishment of a decision-making model dominated by teachers' confidence can help teachers break away from the quagmire of technology-centered materialism, realize the independent field of decision-making, release the psychological shackles of false attribution, so as to uphold the main responsibility of teaching subject. At the same time, it gives play to the perceptual nature of teaching and learning, and reduces the emotional dependence of intelligence. Bandura's triadic interactive determinism provides an analytical framework for deconstructing the factors influencing the effectiveness of teachers' decision making. Teachers' intelligent cognitive misconceptions and professional crisis dominate the establishment of confidence mechanism, and the success or failure of their own behaviors and alternative experience connect the confidence and behavior in teaching. The energy transfer of each subject of man-machine-intelligence contextual factors affects the harmonious interaction of the system. Therefore, the realization path of "human" subjectivity in teaching decision-making is designed from three levels of technical subject, relationship maintenance and mechanism guarantee, an intelligent education system with two-way development of interpretation and understanding is established, and a mutual trust ecology of intelligent teaching and learning in which human-computer-context coexist is shaped. Meanwhile, the legal and institutional construction of the ethics of intelligent decision-making is continuously improved, and the trust mechanism of artificial intelligence education is restarted.

[Keywords] Artificial Intelligence; Teaching Decision-making; Trust Crisis; Subject Confidence