

编者按

对于人工智能(AI)的冲击,许多人是 从“工具”的意义上理解的——AI 提供了新的、更高效的 教学手段和生 产力工具。学者胡翌霖认为,在教育领 域,AI 的出现可能动摇的是教育的目 标和意义,尤其是在人文学科中,AI 的普及可能改变人们知识获取和记忆 的需求。

他认为,AI 时代的教育应回归对自 由与趣味的追求,抵御“生产力至上”的 异化风险。当下人类要做的,是从“牛 马”的宿命中解放出来,追求更丰富的 生活目的;追求“学以成人”,即培养健 全而自由的人格,而非制造“工具人”。

胡翌霖说,AI 时代,我们一方面要 回归“成人”教育的本义,另一方面要 主动拥抱新技术,对抗幻觉。

对抗 AI 幻觉,是哲学家的责任。因 为 AI 虽然可以模仿人类的思考和表 达,但它缺乏人类的身体和有限性,无 法真正与人类共情。哲学反思是人类的 独特活动,这一点 AI 无法替代。因此, 人文社科学者尤其是哲学家,应积极参 与到科技演化的进程中。

培养的学生还不如 AI, 这门课有必要教吗?

此前我在清华大学教授技术史和技术哲学 等课程时,已经遭遇过人工智能(AI)的冲击了。 有一次,我把我出的期末考试题发给 ChatGPT,它 的回答基本能达到中上等水平。现在我试了试更 高级的 ChatGPT-4o 以及新诞生的 DeepSeek, 它们都能“轻松”达到优秀,在许多问题上都超越 了所有学生,甚至比我自己的回答还要好。

当然,AI 的回答还有很多不尽如人意之 处,比如匠气太重、条理过于呆板、喜欢说车轱 辘话、矫揉造作、不懂装懂、虚假引注等,但其 实这些毛病在许多学生身上可能更加严重。况 且 AI 更容易调教,只要稍加纠正,就可以立刻 给出让我更满意的回答。

那么,如果我们的课程培养出来的学生还 不如 AI,这门课还有必要教吗?

AI 时代,各门专业的教育目标和培养方式 都需要反思和定位——哪些内容可以不教、哪 些要采用新的方式、哪些仍值得坚持传统的教 学方式等。

所有人都注意到 AI 的冲击,但许多人仍无 动于衷,每一次技术改进时他们都喜欢找出一些 AI 不如人类的证据,仿佛就心安了,觉得 AI 不是严重的威胁。这种“鸵鸟式”态度,我在许 多哲学家和人文学者那里都能看到。

当然也有人认为,AI 挑战了人类的教育与 工作,但他们提出的分析大多时候还不如 AI。我 看到过一些论文,其作者用无数空话、套话评论 AI 技术对教育的影响,时不时穿插“具身性”之 类煞有介事的术语,最后结论是 AI 不能取代老 师面对面的教学。但是,把通篇文章中的 AI 替 换成互联网、电子通信乃至印刷术都是适用的。

把 AI 只理解为工具, 没有看到最关键的问题

当然,也有许多人能够认真看待 AI 的冲 击,不过大多是从“工具”的意义上理解 AI 的。 这是事实,但未必是最关键的问题。

例如,我们固然可以把蒸汽机看作一头拉 磨效率高的驴,把蒸汽机车看作一头负重更多 和速度更快的驴……但事实上,蒸汽机的出现 远非替代畜力这么简单,更重要的是系统性改 变了社会——蒸汽机带来的工业革命,创造了 全新的生活方式、生产关系,乃至于政治和经 济体制。这些革命性的影响很难用“更强的驴 子”来理解。

在人文学科教学方面,我们不能局限于工 具主义视角,而要注意到 AI 对教育意义或目 的的动摇。

例如,新生的信息技术是否更有利于训练 学生,让他们变得博闻强记?看起来似乎是这 样的:原本人们只能抱着厚重的百科全书看,昂 贵又不方便,检索词条和手抄笔记效率也不高; 现在我们可以随时随地上网检索和查阅百科知 识,免费而迅速,在电子设备上划记知识点也更 加简便。所以,信息时代的学生显然有更好锻 炼记忆力的条件。但是,更关键的问题是,既然 百科知识随时随地触手可及,新一代的学生又 何苦费劲死记硬背呢?

历史上许多新技术都是如此,例如铜器和 铁器更有利于狩猎,而人们掌握了铜器和铁器后, 就进入了农耕时代,各种狩猎技术变得不那么 重要。古登堡印刷机的出现也是类似,最初人们 以为印刷机作为一种传播工具更方便《圣经》的 传播,最后人们发现印刷机反而促进了世俗化。

AI 的发展也是如此,AI 作为工具所促进 的,未必是那个 AI 时代最终会被促进的。但未 来具体如何?现在还很难给出清晰的判断。

从“牛马”中解放出来, 追求更丰富的目的

我个人的判断或愿望是,一方面 AI 作为工 具大大提高了生产力,另一方面在新的 AI 时 代,生产力对 人类文明将变得不那么重要。

未来,如果人们还是把生产力看作最重要 的价值,乃至于唯一的价值,那么这个世界将 变得非常危险。

如果人类努力的最终意义是为了提升生 产力,那么我们很容易得出这样的结论:对整个 社会的生产力提升付出更大贡献的人更有价值,



图片来源:摄图网

一件事能够产出更多生产力,这件事就更值得 去做。然而,当人类发现自己在提升生产力方面 远远比不上 AI 的时候会怎样呢?当人类发现许 多事情都是给 AI 添乱,自己只有给 AI 当耗材、 当“电池”,才能更大提升生产力,那将如何呢? 人类的努力,甚至说存在本身,还有什么意义?

有人认为人类应该灭绝,让硅基生命延 续人类的文明。如果 AI 的自我意识真的觉醒了, 难道它们也会受困于狭隘的工具主义,把自己 的存在意义界定为生产力工具吗?我相信更高 级的智慧生命不可能把自己的价值认作“牛马”。 智慧生命,无论是人类还是 AI,都应该把自己 从牛马的宿命中解放出来,追求更丰富的目的。

生产力当然重要,科学技术提升生产力的 意义也不可否认。但我们始终要清醒:生产之 所以重要,是因为它丰富了人的生活。而不能 反过来,人类生存的目的是为了生产。

生产力的逻辑是线性的、发散的——追 求多多益善。但生活是多元的、有边界的—— 追求恰到好处。

从一座工厂的角度来说,生产的面包当然 是越多越好;但从人的角度来说,最爱吃面包 的人也不可能一天吃几十个面包。从工厂的角 度来说,专业化和标准化是取胜之道,面包流 水线不能用来生产红酒;但从人的角度来说,最 爱面包的人也需要各式各样的食物。

启蒙教育的意义之一, 是教会人承担并享受自由的能力

如何调整流水线的步骤和节奏以生产更 多面包,这是一种智慧。而如何安排生活的步 调和节奏,以便更好享受生活,这更是一种智慧。 有人吃美食时只会狼吞虎咽、胡吃海塞,结果 既不能细细品味其中丰富滋味,也不能保持健 康,甚至给未来增加痛苦。而如何松弛但又有 节制、如何享受但又沉溺、如何扩大自由选择 的空间但又不陷入眼花缭乱的迷茫,这些都是 生活的智慧。

学习知识的一大意义就是帮助我们提升 品位、开阔眼界。比如,一个物质贫穷的人, 也许吃过最好的食物就是煎饼,但如果他的见 识也很贫乏,那么他想象中皇帝的生活无非是 “挑金扁担,煎饼吃一块扔一块”,即便他突然 富裕,也很难找到高级趣味。这也是为什么有 些人在一夜暴富后会醉生梦死挥霍财富—— 因为他们的精神世界从没有得到丰富,既不懂 得如何寻找和品味更多趣味,也不懂得节制和 适度控制自己的欲望。

启蒙教育的一大意义就是教会每个人承 担并享受自由的能力。我们需要博学,不是为 了死记硬背许多知识以便应付考试或炫耀学 识,而是因为我们了解的东西越多,选择的空 间也就越大;我们需要找一些领域深入专精, 这是因为丰富乐趣是需要深入学习和磨炼才 能体会到的,例如围棋无非黑白二子,在外行 眼里枯燥乏味,但是在精通者眼中变化万千、 趣味无穷;我们需要哲学和历史,从而提升内 省能力,理解自己的所来所往,思考自己的边 界和可能性;我们需要数学和科学,以培养理 性和求真的态度,从而在选择生活方式时能够 计算得失并甄别陷阱,同时许多科技领域都 可被认为是深邃而迷人的游戏,钻研其中也 会享受到内在的快乐;我们需要文学和艺术, 以提升自己的审美品位和想象力;我们需要政 治学和社会学,以便更好地参与公共生活;我 们还需要体育锻炼,以提高自己的生活质 量……

汽车的速度早已超过了人类, 我们为什么仍要跑步

我们发现,抛开“培养工具人”“生产人 力资源”等功利性目的之后,当人类摆脱了为 生计奔波、专注于享受生活时,教育和学习 仍然有很多意义——或许可以说,教育和学 习回归了它们最根本的目的:学以成人,即 培养健全而自由的人格。

古希腊的教学传统源于体育馆,“school” 的本义是“闲逛”,希腊人在体育馆的闲

时学习学问。柏拉图的阿卡德米学园和 亚里士多德的吕克昂学院都是在体育馆的 旧址上建立起来的。

希腊自由民热衷于体育锻炼,并不是 为了更好地出卖劳动力,而是因为这属于 追求自由的一部分。他们一方面追求身 体健美,另一方面也在奥林匹克运动会 这样的体育活动中享受竞技游戏的乐 趣。

我们更容易理解锻炼身体 的意义:哪怕我 像希腊自由民那样,或者未来依靠科技 的支持,可以免于为生计奔波,仍然需 要锻炼身体。哪怕汽车的速度早已超 过了人类,我们仍然会跑步。保持身 体健康不是为了给别人当工具,而是 为了能够更好生活。

而希腊人认为人的灵魂和身体一样, 也需要持续“锻炼”,才能免于腐朽。 灵魂的健全同样也不是为了给别人 工作或作贡献,而是为了更好生活。 因此当机器在许多方面超过人类的 头脑时,我们仍然有理由锻炼自己 的头脑。

希腊自由民的教育,最根本的目的 是为了身体的健美或灵魂的健全。而 完善的人具备“智慧、勇敢、公正、 节制”等品性。

中国古代也类似,孔子论“学以成人” 时讲:“若臧武仲之知,公绰之不欲, 卞庄子之勇,冉求之艺,文之以礼 乐,亦可以为人矣……”见利思义, 见危授命,久要不忘平生之言,亦 可以为成人矣。”(简而言之,这段 论述包含了智慧(知)、勇敢(勇)、 公正(义)和节制(不欲),还强调 了才能(艺)、文明(礼)和审美(乐) 等方面。

诚然,古代贤哲或许享受了奴隶或 底层劳动者的供养,因此可以不事 生产,不受功利思维所累,追求更高 的趣味。但我相信随着科技的进 步,最终每个人都可以从被迫的劳 动中解放出来,更自由地思考和意 义。

人因为弱而独特, AI 不可能与人类共情

有人说,“人类一思考,上帝就发笑”, 那么如果 AI 获得了比人类还高的 智能,人类的哲学思考,是否在 AI 面前也不值一提了呢? 这个问题取决于我们如何理解 “哲学”的意义。

“哲学”源自希腊文,是“爱-智慧” 的组合,最早使用这一名词的哲 学家与当时许多哲教授辩论技 术或者“话术”的“智者”有区 别。哲学家追求的 不是帮助你提升在演讲时天花乱 坠、辩论时咄咄逼人、掉书袋 时装腔作势的语言技巧,哲学家 追求的是智慧本身,不是为了 说服别人,而是为了满足自己的 求知欲。

德尔菲有一句箴言:“人啊,认识你自己”; 苏格拉底也说,“未经审视的生活 是不值得过的”,哲学的最终对 象永远是每个人自己。我们既 不能用审视他人来替代自我审 视,也不能用他人的审视来替代 自我审视。哲学反思甚至都不 能被他人取代,更不用说被 AI 代替了。

当然,人需要互动交流,从他人的 言行中获得启发,进行对照。在 这个意义上,AI 当然会成为人 类进行哲学思考时的良师益友, 我们能够从 AI 中得到启发。但 这些启发并不能替代人与人之 间的交流,因为他人的生活和 处境总是与自己更近,更容易发 生共鸣。

而在这方面,AI 无论如何替代不了 人类,不是因为它太弱了,而是 因为它太强了,超越了人类固 有的有限性,因而永远也不可能 与人类共情。

关键在于,人类有身体。把 AI 装进一台 仿生机器人里并不意味着 AI 有了身体。因为 人类的身体不仅提供了移动和 操作的能力,更关键的是提供 了一种边界或束缚。人不能脱 离自己的身体,而他的身体只有 “一”个。这使得人的感官能力 有限、反应速度有限,而且无法 免于生老病死。而机械肢体对 于 AI 来说并非束缚,因为 AI 在本 质上就是一堆数据,这些数据可 以在任何数字媒体上拷贝。同 一套数据,可以植入这个机器 人,也可以同时植入那只机器 狗,可以同时有完全一致的备 份,它们可以同时被运行,被修 改成无数不同的版本,甚至随 时再合并起来。

此外,对人类来说,“生存还是 灭亡,这是一个问题”,生死问 题是哲学永恒的第一问题。但

对于 AI 来说,即便它有自己的存 在论问题,其境况也肯定与人类 大不一样。

因为我始终只能是“一”,所以我 需要仔细斟酌:晚上是下馆子 还是回家吃?明天是和张三喝 酒还是和李四约会?明年是去 求职还是考研?未来是回老家 “躺平”还是去海外闯荡?我这 辈子最终要活得安安稳稳还是 轰轰烈烈……人类通过“智慧” 和“趣味”进行选择 和决断,根本原因是自己的有限 性。如果我可以无限拷贝,可 以同时和张三与李四交谈,可 以同时在此处和彼处,那我 还需要那么在意如何比较和 抉择吗?或者说,那还能形成 “我”这个概念吗?

当然,在存储容量、传输速度 和算力等方面,AI 存在有限性, 因此仍然有可能形成某种形 式的“自我”概念。但这种生 命形态和人类终究大不相同, 人类和 AI 在最根本的哲学问 题上难以共情。

在哲学和其他人文领域, AI 还有很长的路要走

你也许会说,虽然 AI 在本质上 不可能共情人类的生死问题, 但它 可以模仿人类,给出比哲 学家们更好的演说和理论,人 类学哲学只需要和 AI 交流就 够了。

而我认为,也许在许多领域,AI 都 可以成为一个好老师,但至 少在哲学和其他人文领域,AI 还有 很长的路要走。

目前 AI 存在一个问题,就是所谓 “幻觉”现象——AI 可能煞有 介事地说出一大堆看起来很 可靠的言论,但实际是胡编 乱造的。

这是由目前 AI 的训练方式 决定的,它总是给出更受普通 人欢迎的回答。训练员、标 记员和普通用户都在参与对 AI 的调 教,强化能给出更好回答的 迭代路线。

所以说现在的 AI 好比古希腊 的“智者”,他们擅长和教授 辩论,但他们的话术以打动 更多听众为主,而不是以追 求真理为目的。因此在专 业学者看来错误百出的回 答,只要它能唬住更多普通 人,就难以在版本迭代中得 到正确反馈,从而逐渐得到 纠正。

当然,有些专业领域,特别是 理工领域,AI 更容易进行专 门训练。例如 ChatGPT 刚 出来的时候,说起文学艺术 头头是道,反而是做数学和 逻辑题经常出错,甚至连 9.8 和 9.11 哪个更大都 分不清。但是后来经过专 门的强化训练之后,AI 在数 学、科学和编程等理工领 域的表现已显著提升。

这一方面是由于许多科学 家和工程师更积极参与 AI 的 开发,另一方面得益于这些 学科形式化程度比较高,对 就是对、错就是错,容易形 成严格的判断标准,给 AI 训 练提供迅速而准确的反馈。

而在人文学科领域就要更 难了。首先人文学者,特别 是哲学家,往往没有公共的 标准来评判。哲学家并不 讨好大众,相反,他们总爱 特立独行,做出独特而尖锐 的批判。人文领域知识分 子的最大社会责任就是批 判,他们敢于说出普通人 不爱听的话。许多古典 式的哲学家更是致力于构 建复杂的思想体系,单独 拿出他们的只言片语,往 往是晦涩难懂或令人疑 惑的,只有深入研究其整 个思维过程才能够体会 到思想的深刻和精彩。

其次,AI 很难形成哲学家 那样的人格独立性和思 想统一性,它们也很难 把它训练成这样。

我们要重视这一趋势, 不能因为这方面的训 练麻烦、困难,就不 再应对,任由 AI 幻觉 泛滥。整个业界应 当鼓励各领域的学 者,特别是人文社 科学者,积极参与 到训练和纠正 AI 的 工作中。这需要我 们打通文理隔阂, 消除偏见。

学科鄙视链在中国尤其严重, 工科出身的程序 员往往看不起文 科科学者,而文 科科学者也经常 不重视科技前沿。 因此,当务之急, 还是要促进人文、 艺术和科技领域 的交流对话,共 同促进科技的发展 并应对可能的危 险。

AI 时代正在到来,哲学家们 一方面要回归初 心,超越功利态 度,像古希腊哲 人那样,把“爱 智慧”作为锻 炼灵魂的生活 方式;另一方面 也要关注前沿, 放下隔阂和偏 见,与科技业界 紧密交流,共 同参与科技演 化的进程。

速读

论文题目:
数智时代
教师责任伦理的守正创新
作者:卢海丽、张立国、刘晓琳、付科峰
出处:《自然辩证法研究》,2025 年 01 期

数智技术赋能教育是数智时代教育发展的 趋势,但同时衍生出数智时代的教师“技术 依赖”“去身体化”“技术异化”等“去责任 化”趋势,因而亟待教师责任伦理的革新。

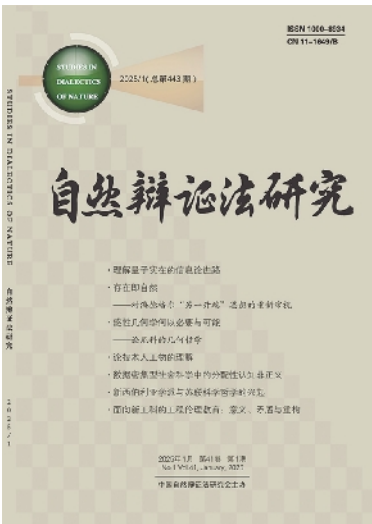
作者认为,数智技术的过度依赖容易使教 师忽视人的根本价值,从而形成数智时代教 育过程的伦理隐忧。比如智能评价系统、智能 作业评判等为教师评价提供了依据,但过度依 赖容易使教师忽视学生成长的复杂性和不可 量化性,致使教师“引导”与“育人”的责 任感弱化,使工具理性僭越了价值理性。

数智时代的教育变革致使教师主体责任 “去身体化”,时空和场域的式微使教师的 道德示范和情感魅力失去了该有的场地和情 感纽带,使教师与学生的关系物化、情感剥 离,基于情感的师生伦理也受到阉割。此外, 技术的过度介入容易滋生技术误用、教育异 化等现象,导致出现价值湮没和情感割裂的 现象,加剧以“教育”为中心的情感淡 漠异化。

作者表示,在数智时代教师的责任担当 上,要形成以“以法为界,契约协调”为红 线,以“以人为本,循化贵道”“以道驭技, 推学助师”“和谐共生,情知共生”为原 则,实现从知识建构到全面发展的“人→ 技术、技术赋能、技术→人”的责任伦 理创新,防范和规避数智时代教师责任伦 理风险,使教育的技术支撑不可异化。

教师的育人功用不可替代,教师的责任 主体不可转嫁。作者提出,确立教师主体角 色——以人为本,循化贵道;发挥技术伦 理规范——以道驭技,推学助师;提高教师 引导作用——和谐共生,情知共生;重塑 教育契约精神:以法为界,契约协调。

未来,在教育数字化发展过程中,还会 出现更加繁复的伦理问题,而连续性的伦 理审视是建立与内化伦理规范的有效途 径。



《自然辩证法研究》封面。

论文题目:
人工智能伦理治理范式:从价值 对齐到价值共生
作者:夏永红
出处:《自然辩证法通讯》,2025 年 01 期

当前人工智能(AI)伦理治理的主导范 式是价值对齐,其目标是确保机 器价值与人类价值一致。本文 提出了一个被称为价值共生的 替代范式。这一新范式的初衷 是:由于人类价值的复杂性,不 应该也不可能寻求机器伦理向 人类伦理的单向对齐,而应探 索人类与机器的价值共生,让 它们相互受益和相互承认。

价值对齐要确保 AI 模型在 设计时就以人类的方式行事。 然而,人类并不存在一种单一 的价值观,甚至同一个时代的 同一族群中也存在相互冲突 的道德学说,因而难以提炼出 一套普遍适用的机器伦理框 架,如果强行将机器伦理与 人类伦理对齐,有可能会强 化目前的主导价值观(往往 是西方白人中产男性的价值 观),也有可能强化操作者的 权力。

价值对齐的技术实现依赖于 计算机对人类价值观念的全 面把握,但大量人类常识都 没有被记载在文本中。诸如 人生经验和人情世故等多 为默会规则;很多人类常识 的习得体现在与他人和世 界的具身交互之中。比如 护理中肢体动作的细节,体 现同情、耐心等价值观念等。

这些使得价值对齐既不可能, 也不可取。本文提出的价值 共生范式,其核心有两条原 则。一是相互受益原则,即 机器与人类可以因为彼此 的存在而相互受益,以对 方的繁荣作为自己繁荣的 条件。为此,可以考虑采取 生态位并行和技能互补两 种不同的 AI 设计准则。二 是相互承认原则,即机器 与人类都在价值上承认对 方,不仅承认对方的工具 价值,也承认对方的内禀 价值。而主体间性、跨主 体性和机器间性将是未来 人机和平的拱顶石。

价值对齐和价值共生两种范 式基于不同的技术语境。 本文呈现了一个可称为“创 新-风险螺旋”的现象:有 限 AI 的应用引发价值对 齐问题;为了解决该问题, 需要发展超级智能;超级 智能的应用又带来价值共 生问题……但这个问题解 决后依然可能引发新的风 险,比如人类的自我退化 和淘汰。在这个无尽的 螺旋中,风险与创新不断 交织。(尹一)