

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、更生产力的工具。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。哲学反思是人类独特活动,这一点AI无法替代。因此,人文社科学家尤其是哲学家,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发给ChatGPT,它的回答基本能达到中等水平。现在我试了更高级的ChatGPT-4o以及新诞生的DeepSeek,它们都能“轻松”达到优秀,在许多问题上都超过了所有学生,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的地方,比如语气太重、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,但确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,未必是整个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,那么我们会很容易得出这样的结论:对整个社会的生产力提升作出更多贡献的人更有价值,

AI时代,哲学何为

胡翌霖



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,人类生产的目的只是为了生产。

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,最爱吃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,以便更好地享受生活,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、如何享受但又不会沉溺、如何扩大自由选择的空间但又不会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,这也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”,希腊人在体育馆锻炼的闲

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: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,这段话包含了智慧(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,更自由地思考生活和意义。

人因弱而独特,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人辩论技艺或者说“话术”的“智者”有区别。哲学家追求的并非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,更容易引发共鸣。

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人,可以同时在中国和美国的无数台计算机上拥有完全一致的备份,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 and 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,但他们的话语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哪个更大都不清楚。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他思想的深刻和精彩。

其次,AI很难形成哲学家那样的人格独立性和思想统一性,我们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,特别是人文社科学家,积极参与到训练和纠正AI的工作中。这需要我们打通思维隔阂,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学者,而文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,因而亟须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域和情感纽带,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制约。此外,技术的过度介入容易滋生技术滥用、教学异化等现象,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,循化贯道”“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,循化贯道;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(往往是西方白人中心男性的价值观),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……但这个问题的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)