

生成式人工智能“生成性”的哲学考辩^[*]

——从马克思“感性活动”观点看

王 策

(西安邮电大学 马克思主义学院, 陕西 西安 710121)

[摘要]在当前人工智能与人类智能的诸多哲学解释中,马克思的认识论占有不可忽视的重要地位。生成式人工智能“生成性”的底层逻辑在于大型语言模型的开发和利用,体现为深度学习和人工神经网络所依赖的数理机制。马克思从感性活动论出发,有力地阐明了人类智能生成的实践性、社会历史性和多维性,使人类智能生成的默会维度和意向性维度得到彰显,从而睿智地预见到两类智能生成方式的基本区分。在人类智能生成过程中,感性活动论所揭示的逻辑与辩证性规范、形式化与意向性以及事实与价值的结合,对于未来人机融合智能的开发提供了极有价值的哲学启示。

[关键词]生成式人工智能;大模型;生成性;感性活动

DOI:10.3969/j.issn.1002-1698.2024.03.016

马克思作为19世纪致力于变革人类社会的思想家,他的认识论哲学,能否回应人工智能尤其是生成式人工智能所引发的诸多哲学问题?当前,大多数哲学研究者对这个问题持“敬而远之”式的漠视态度。他们普遍认为,如果非要从马克思角度探讨人工智能,也仅仅切中人工智能可能引发的社会效应问题,譬如社会分工、劳动异化、资本剥削、人的解放等等;一旦讨论人工智能与人类意识、大脑机制、逻辑机理、语言本质等更为基础的纯粹哲学,马克思必然失语并黯然离场。然而,如果细究马克思的认识论尤其是他的“感性活动”理论,上述看法则过于武断。

笔者比较认同张祥龙教授对哲学的一个见解:“哲学是对边缘问题的各种合理探讨”。^[1]如实说,人工智能哲学不属于对人工智能科学本身的研究,如果说图灵于1950年发表的哲学论文《计算机器与智能》是关于“机器智能”的核心思考的话,那么当前人工智能哲学则是对人工智能科学及其后果的“边缘性思考”。如果把各类人工智能哲学看作一个圆圈,这种“边缘性”亦有其层级性。若从现代逻辑、逻辑语义学、数学哲学、心灵哲学、意识哲学等方面看待人工智能,就属于距离圆点较近的“内圈性”思考;若从道德哲学、政治哲学、社会哲学、资本哲学、情感哲学等方面看待人工智能,则属于距离

作者简介:王策,哲学博士,西安邮电大学马克思主义学院教授,主要研究方向为分析哲学、马克思主义哲学。

[*]本文系国家社会科学基金项目“20世纪分析哲学与康德哲学的关系研究”(19XZX011)的阶段性成果。

原点较远的“外圈性”思考。须注意,这里“外圈性”并不等于“圈外性”,也不等于没有理论意义。如果这种分层能够成立,那么从马克思关于人类意识的生成角度看人工智能,则思考范围可划归于“内圈”与“外圈”的中间地带。不能因为马克思是19世纪的、专注于改造世界的思想家,就一概否定他与当代人工智能的理论关联。就一个时代最新的、最有力量的自然科学与哲学的关系而言,马克思在《1844年经济学哲学手稿》中作了如下表达:“自然科学展开了大规模的活动并且占有了不断增多的材料。而哲学对自然科学始终是疏远的,正像自然科学对哲学也始终是疏远的一样。过去把它们暂时结合起来,不过是离奇的幻想。存在着结合的意志,但缺少结合的能力。”^[2]本文试图把当前“最新的、最有力量”的自然科学之一——生成式人工智能——的生成性问题,置于马克思认识论哲学范围进行探讨,一方面希冀增进对人工智能和人类智能生成方式的双向理解,另一方面试图表明,马克思认识论尤其是他的“感性活动”理论,在当前人工智能哲学的各类探讨中,占有不可或缺的理论地位。

一、生成式人工智能的生成性机理

ChatGPT 全称是“生成式预训练转换聊天机器人”(Chat Generative Pre-trained Transformer),属于“生成式 AI 模型”的最新技术成果。生成式人工智能具有强大的文本生成能力,有着显著的反馈和预测能力,这种从量变到质变的涌现能力,体现为对声音、语言、文字、图像这些人类交流现象近乎不留缝隙的全盘模拟,在自然语言表达和图文生成方面,一度逼近人类智能的自主性和创造性。生成式人工智能基于深度学习和高度复杂的非线性人工神经网络,通过人类反馈强化学习的方法,实现了对人类个体智能和群体智能以及复杂思维机理的完美表征,在人机融合方面表现得更为自然成熟。

大型语言模型(Large Language Model,缩写为“LLM”,简称“大模型”)是生成式人工智能的技术关键。大模型具有无比巨大的数据参量和复杂的算法结构(预训练+人类反馈强化学习+预测扩展的深度学习),通常在强大算力支撑下利用海量数据集进行训练,表现出强大的通用性,几乎与人类的图文创造能力和使用自然语言的水平并驾齐驱。大模型的生成性和涌现能力,主要基于如下四个基本原理:^[3]

原理一:大模型底层原理——基于关联度的预测。基于语元关联度的预测是大模型技术的底层原理。大模型的基本目标是,实现无障碍人机交流,即能听懂、说出像人那样的自然语言,能流利回答人的问题。无差异性地听懂并说出人的自然语言被认为是生成式人工智能机器的关键特征,它的设计原理是:首先以人类规模的信息累积当其语料库,然后自主选择关键性的语素痕迹,继而对具有语义关联的语言因素进行组合,最后自动生成新的语言。无论是人类使用的自然语言,还是其他类型的语言,通常会以各种文本形式留下语言痕迹,时间一久便形成语料库。语言痕迹的显与隐、多与少、丰与俭等一系列特征,构成了所谓的“语元关联度”。语元是语言意义的基本构成要素,语料库中语元之间的关联强度不同,遂形成不同的语元关联度。譬如,“我”“门”“们”是三个基本语元,在人类日常表达中,“我”与“们”的关联度就远大于“我”与“门”的关联度。大模型一旦获取海量数据的语元关联度,就可以预测下一个出现的语元,继而生成符合人类习惯、完整顺畅的语句。例如,当一个语句中出现“我”这个语元,大模型根据储存的语元关联度预测下一个语元是“们”的机率要远高于“门”,这样就生成了“我们”这一关联。

原理二:大模型基础特性——实例性模型。通常来讲,科学研究的建模思路大体分两种:概括性模型和实例性模型。前者由概括性法则描述普遍规律,后者则是对单个现象给予描述,即由大量实例堆积而成的模型。属于单个现象的实例构成了大模型的基础特征,同时决定着大模型的基本性能和

运行模式。GPT-4正是源于这种实例性,才实现它的可靠性、真实性、融贯性和可解释性。此外,它直接从语言文本提取实例性模型,再运用整个人类范围的原始语料库当作自身的训练样本,这样就极大提升了数据统计的覆盖率。可见,大模型生成语言的数据规模超过任何团体的成员,它还在不断地提取新的语言痕迹,每时每刻都运用语元关联度来生成提问所需要的新语元。面对任何人提出的任何问题,哪怕仅仅一个人知道答案,只要这个答案留下了信息痕迹,大模型就可以提取并回答提问。为了对提问生成更精准的答案,大模型运用两种优化和解决方法,一个是“精调法”(Fine-tuning Method),一个是“提示法”(Prompting Method)。前者的基本思想是采用已经在大量文本上进行训练的预训练语言模型,然后在小规模的任务特定文本上继续训练它;后者需要对提问设置更多的提示,这样大模型才可以激活与之相关的语元,以便符合预期的答案自然流出。

原理三:大模型拟人化的哲学根源——心物二元与投射效应。人类的自然语言除了包含信息之外,还同时蕴含着情感、情绪、愿望、意图等丰富内容,这与人工语言根本不同。大模型输出的话语与自然语言是否具有同样的内涵,至今还没有定论。大模型只能说类似于人,这种拟人化基于心物二元论和投射效应。一旦大模型输出的语言与人类自然语言的相似性达到难以分辨的程度,人们会自然而然地根据话语内涵把自我投射到机器的大模型上,认为这些内涵是大模型“大脑”中真实、自主的产物。心物二元论认为,机器是无心之物,只有人有“心”,语言是“心”存在的根本标志。因此,只有人才会说话,任何物包括机器都不可能说人类语言。然而,一旦遇到会说话的大模型,人们就作出新的判断:机器具备了人的智能,甚至具备了判断、意向、情感等类似人类才拥有的“心”。从心物二元论出发,人类对大模型难免产生一种畏惧,即生成式人工智能比历史上的任何机器都更令人担忧,它正在加速威胁人类文明。

原理四:人机融合原理——以功能模拟为主的集成智能。图灵早在1948年就设想了一种集成智能,可视为人机融合的思想萌芽。图灵的思路是,虽然机器的工作原理跟人的智力原理和行为完全不同,但还是尽可能地让机器模仿人的智力和行为。阿尔法狗赢得围棋比赛,就是采用落子胜率估算的决策原理,这与人类围棋知识截然不同。人工智能发展史证明,机器的运行原理虽然与人类智能不同,但前者会胜任一些更复杂的任务,其性能往往远超人类。大模型学习并大大扩展了阿尔法狗的经验,机器在行为执行和语言生成方面越来越接近人类,就在于被植入了语言关联度预测的程序。譬如,图灵测试依据的哲学原理是“行为的不可分辨性”,但人类的棋类游戏行为与阿尔法狗的下棋行为的确不同,二者是可以分辨的。由此机器被判定为“会下棋”。虽然大模型的语言输出和行为执行与人有着显著区别,但这并不影响人们对大模型语言能力的承认。近年来,生成式人工智能的成功使“行为模拟”逐渐替代以前的“功能模拟”,即让机器模拟人的智力功能,而无论该行为是否具有可分辨性。与功能模拟不同,科学家推出了最新的原理模拟,后者把人类智能最深层的原理植入机器,然后让机器模拟这些原理。集成智能正是以功能模拟与原理模拟为基础,在实践操作中体现为深度学习和人工神经网络,生成式人工智能、大模型乃至阿尔法狗都属于典型的集成智能。

鉴于上述四种基本原理,大模型在图文输出方面实现了强大而逼真的拟人性,表现出无与伦比的生成性和涌现能力。然而,大模型引发的社会后果正在引起企业界和政府的高度重视。Open AI公司创始人埃隆·马斯克联合一帮科技领袖发表了一封公开信,号召科技人员停止开发比GPT-4更强大的人工智能机器,信中提到对人工智能导致错误信息传播、劳动力市场自动化风险以及文明失控可能性的担忧。2023年7月,中国官方公布了《生成式人工智能服务管理暂行办法》,旨在促进生成式人工智能健康发展,防范生成式人工智能服务风险。不同于以往对新科学技术的单纯向往,人们对生成

式人工智能既有欢呼,又有恐惧,既要发展,又要防范。人们对其态度为何会“冰火两重天”?接下来,笔者从马克思认识论角度探讨人类智能的生成性,以便与大模型的生成性机理形成比照性理解。

二、“感性活动”与人类智能的生成

马克思并没有直接用“人类智能”“生成性”“涌现”这些词汇,而是用意识、自我意识、观念、智慧等语词描述人的思想存在和现实状况,用发展、劳动、改造、生命活动、历史性、辩证性等语词描述人、人的思想观念以及人类社会的形成和变化。马克思特别把“哲学”“理性”“纯粹理性”“纯粹意识”“绝对理念”“自我”等唯心主义哲学的关键概念作为批判标靶,目的是跟包括康德、黑格尔、费尔巴哈在内的德国古典哲学划清界限。可以这样说,马克思的新唯物主义认识论突破传统笛卡尔式的纯粹意识哲学,通过赋予“感性活动”以深远的哲学意义,成功描述和解释了人的发展和人类智能的生成。

(一)感性活动是人类智能的前提和基础

近代认识论哲学基本是笛卡尔“我思故我在”的各种发挥,即“思维”与“存在”的二元论,人的主体性着重体现为思维,思维的极致就是逻辑和数学,这是所有人工智能的基础,即人工智能实质是运算过程,仅仅是一个没有血肉的“思维体”。马克思通过“感性活动”这一概念,强调人类意识的发展、人类智能生成的丰富性和多维性,继而强调了意识和语言的感性本质。在《德意志意识形态》中,马克思的哲学思想臻于成熟,他把外部世界、意识和人的全部生活界定为“活生生的感性活动”。“感性活动”与后来更为成熟的“实践”概念并不完全等同,虽然后者包含了前者的主要内涵。比起“实践”来说,“感性活动”这一概念携带着马克思在新旧哲学分界之际的思想痕迹,更能表达出他从黑格尔概念发生论走向活动发生论、从费尔巴哈直观唯物论走向辩证唯物论的思想转向过程。用“感性活动”来表达马克思认识论的实质,不但更为准确、传神和生动活泼,且完美显示了马克思的思想张力。因此,感性活动理论是笔者分析马克思关于意识、人类智能生成的认识论总纲。

首先,意识来源于感觉器官与世界的“对象性”联系。与笛卡尔的“我思”、康德的“先天范畴”以及黑格尔“绝对理念”等肯定纯粹意识存在的哲学不同,马克思强调意识自身没有任何神秘的内部机制,仅有外部的来源。纯粹的意识哲学虽然承认人类通过感觉器官接受外部经验,但意识的内部机制与感觉经验始终是二元分裂的。与此不同,马克思认为人类感觉器官是以整体性、联合性和社会性的方式与感觉对象处于一个统一关系中,即“对象性”关系。他没有把器官对对象的“占有”理解为器官的生理机制的单独作用,而是归结于人的感性活动。在马克思看来,人在感性活动中激活了感觉器官对对象的占有,感性活动使各个器官的功能得以发挥,感觉的本质是社会性的活动而非生理性的静观,感性活动是人类意识、人类智能产生的基础。换言之,感觉的感性功能并不能产生人类智能,人类智能只来源于感性活动。马克思在《1844年经济学哲学手稿》中对这一思想作了充分表述:“人对世界的任何一种人的关系——视觉、听觉、嗅觉、味觉、触觉、思维、直观、情感、愿望、活动、爱,——总之,他的个体的一切器官,正像在形式上直接是社会的器官的那些器官一样,是通过自己的对象性关系,即通过自己同对象的关系而对对象的占有”,^[4]“感觉为了物而同物发生关系,但物本身是对自身和对人的一种对象性的、人的关系,反过来也是这样。”^[5]

其次,意识没有独立性,但体现为能动性。马克思一贯主张,人类的思维、意识和精神现象,并不是黑格尔所谓的独立存在,思维的形成、存在和发展,离不开人本身乃至人类环境的整体状况。他说道:“人还具有‘意识’。但是这种意识并非一开始就是‘纯粹的’意识”,^[6]“思维和存在虽有区别,但同时彼此又处于统一中。”^[7]思维没有自身的独立性,因为人虽然是特殊的个体,是独立的社会存在,

但同时人也是总体,他能够把全部观念包含在自己之中,因而是观念的总体。此外,人还是被感知的、社会的、自为的主体存在。马克思沿用黑格尔的“自为”概念强调思维不是对外物的简单反映,而是有着能动性和自主性,这就预示了人的主体性,人类正是在认识世界和改造世界的感性活动中,确证自己的主体性。

最后,语言与意识一体分殊,语言具有感性和社会历史性。不论在笛卡尔式的意识哲学中,还是在当今人工智能哲学中,语言与思维被看作一种表征与被表征的关系。20世纪初分析哲学之所以产生,乃在于哲学家运用现代符号逻辑的方法,把传统的哲学问题转换为语言问题,认为只要对语言进行逻辑分析,以往哲学难题都可以迎刃而解。从莱布尼茨设想的“通用语言”到今天的计算机编程,人们对机器智能的实现,就体现为不断创造和优化机器可以识别的语言。因此,语言是思维、意识、精神的外在表征,至今仍是哲学家的牢固共识。不同于传统语言观,马克思有着自己独特的语言哲学。他认为语言不是对意识的单纯表征,语言是发生性的,其来源和不断丰富离不开人的感性活动。他指出:“语言是一种实践的、既为别人存在因而也为我自身而存在的、现实的意识。语言也和意识一样,只是由于需要,由于和他人交往的迫切需要才产生的。”^[8]马克思继而认为,语言与意识一体分殊,“思想、观念、意识的生产最初是直接和人们的物质活动,与人们的物质交往,与现实生活的语言交织在一起的。”^[9]他形象地把语言比喻成“思想的物质外壳”,语言和意识都具有实践性和社会历史性。

(二)感性活动与人类智能生成的默会维度

在人类智能的生成方面,马克思的感性活动论弥合了思维与世界的分裂,以更宏伟丰富的维度诠释了人类智能的生成过程,即“人对世界的任何一种人的关系”涵盖极其丰富的内容,不但包括人的五种外部感觉,还包括内在的思维、直观、情感、愿望、活动、爱等多样的要素。马克思一方面避免了近代哲学对思维本身的盲目尊崇,另一方面对人们把世界万有不假思索地视为一种单纯的数理架构的世界观作了反思。遵循这一思路,他的感性活动论完全可以视为哲学家波兰尼“默会知识论”的先兆和预表,二者心照不宣地把人类认知、人类智能的生成从单一的“智能”维度提升到丰饶的“智慧”维度。

波兰尼把人类知识分为两类:一类是“明述知识”(explicit knowledge),另一类是“默会知识”(tacit knowledge)。前一种知识亦称为“显性知识”,后者亦称为“隐性知识”。前者指能够用命题明确表达的知识,或者可以用数学和逻辑进行描述的知识。教科书的内容基本属于显性知识,这种知识在人际交流中较少发生意义损益,也基本不会因人而异地被理解,除非理解者是对既有人类认知改天换地的天才。自然科学知识、大多数现行的社会规则,大都属于此类知识。显性知识的最大特征是它的可复制、可通约和可传达性,用数理逻辑表述这类知识,机器语言就成为可能。当前人工智能的底层逻辑结构、符号规则、数学建模等再怎么复杂,这一套东西都可以在另一台机器上照样运行,除非算力大小影响机器运行速度。近代西方哲学自笛卡尔把“我思”界定为第一实体,加之近现代科学的数学化表述,显性知识的形而上学地位不断得到强化,最终获得“知识典范”“真理”“自明”等标贴与殊荣。莱布尼茨的“通用符号”、休谟的“观念的知识”、康德的“分析命题”以及黑格尔的“绝对理念”等,皆彰显了显性知识的形而上学地位。显性知识的地位在20世纪初出现的分析哲学中达到极致,弗雷格、罗素、前期维特根斯坦和维也纳学派通过“反心理主义”的方式,把直观、情感、意志、道德等人类认知的默会维度从自然科学、数学和逻辑等知识系统中统统清除出去,极力追求世界的逻辑构造。^[10]这种状况在20世纪中叶后逐渐得到改观,在波兰尼看来,虽然人类的认知机制及其过程非常复杂,但心灵本身所包含的默会能力在整个认知中居于主导地位,比之于显性知识来说,认识的默会维度具有实践上和理论上的优先性。他在 *Knowing and Being* 一书中指出:“语言拓展了人类的智力并大大超越了纯粹默会

知识的范围,但语言的使用方式及其内在逻辑仍然是默会的。默会知识本自具足,明述知识却无从自立,它必须依赖于人们对之默会地运用和理解,实际上,一种完全明述的知识是不存在的。”^[11]他强调,默会性的认知本质上是一种洞察力和领悟力,它体现为人们关于经验世界的感性与理性相互交融的理解。因此说,默会性的认知乃是具有时间意识的存在者对世界整体的某种“领会”和“悟入”。与波兰尼的思想极为接近,马克思诉诸感性活动以期实现对世界的能动地改造,因而隐含着关于人类知识、人类智能生成的默会维度。感性活动的默会维度着重体现为如下两方面:

一是人类智能生成的身体维度。人类认知的默会维度实际上蕴含着认知的身体性。笛卡尔式的二元论把心灵与身体割裂开来,认为思维高于物质性的身体;与此相反,马克思的感性活动论和波兰尼的默会认识论,则强调身体在认知中的重要性,从而把身心整合起来。波兰尼主张心灵本质上是身体化的,心灵虽然依附于身体,但又不等同于身体。他富有洞见地说道:“心灵是它的各种外在活动的意义,如同部分与整体的关系,这些活动体现为心灵的线索与心灵综合体的关系。因此,理解他者心灵的方法乃是融入这个人的活动过程中,只有融入于一个棋手的走法,我们才能理解他的心灵。同样地,只有融入因疼痛而扭曲的一副面孔,我们才能理解另一个人的痛苦。”^[12]心灵是身体活动的意义,表明身体活动的基础性和优先性,这个观点并不意味心灵与身体的分离,否则会再次陷入笛卡尔式的二元论。以此观之,马克思的感性活动论所暗含的身心关系,与波兰尼的观点深度契合。关于人的身体性,马克思在早期文本中常常使用“肉体生存”“动物性机能”“无机界生活”“肉体的主体”等字眼描述人在异化劳动中的奴役状态,后来使用“劳动”“生命活动”“自由活动”说明人类正常的感性活动。就人的身体对意识产生的作用,马克思在《德意志意识形态》中明确说道:“人们之所以有历史,是因为他们必须生产自己的生命,而且必须用一定的方式来进行:这是受他们的肉体组织制约的,人们的意识也是这样受制约的。”^[13]马克思所谓身体的认知作用,不是人的静态的物质性肉体,而是人的身体性活动,是身体与意识的交互作用、整合为一的人的创造性活动。

身体维度体现了人类智能生成过程的整体性,这是感性活动的重要特征。我们知道,“劳动”是马克思思想的核心概念,是判别人究竟是一种“思维存在”还是“社会存在”的唯一标尺。劳动是人类生命生产的形式,劳动产生了智慧,马克思用劳动来强调人类认识、人类生命乃至人类社会产生和发展的整体性。如果说动物同自己的生命活动是直接同一的,那么人在劳动过程中既与自己的生命活动同一,同时还与这种活动分裂。所谓“分裂”就是人能够意识到自己的劳动,产生了苦、乐、爱、恨等意识,人的劳动是有意识的生命活动,人在劳动中使得“视觉、听觉、嗅觉、味觉、触觉、思维、直观、情感、愿望、活动、爱”等这些生理性的感觉器官和精神性的认知途径,以联合的、整体的方式成为人的对象并为人所占有。

马克思的感性活动论蕴含人类意识生成的默会维度,能够把人类智能与世界融合起来进行整体观照。默会知识无法被命题化,但它有效,能够影响人的思想、身体,从而作出行动。如果说显性知识是可见的冰山一角,默会知识则属于更广大、更丰富、更深不可测的人类智能的基底。感性活动的认识论意义在于,把笛卡尔诉诸神秘“松果腺”的身心链接关系实践化和现象学化,突破了把人类智能的生成纳入机制主义、行为主义和功能主义的庸俗解释之中的做法。

二是人类智能生成的意向维度。马克思通过感性活动强调人的本质的社会历史性,在他看来,人类智能不应该单纯体现为数理性认知,即笛卡尔式的理性和思维,还应当包含很多纯粹理性以外的东西,如直觉、意志、情感、希望、审美、爱等等。在人类智能的展现方式上,除了形式逻辑之外,还应包含辩证逻辑、直觉、矛盾、价值等等。马克思在感性活动诸要素中特别提到了“激情”,他说道:“对象性的

本质在我身上的统治,我的本质活动的感性爆发,是激情,从而激情在这里就成了我的本质的活动。”^[14]马克思对激情、意志等要素的强调,表明人的感性活动蕴含着人的活动本身、人的意识以及人类智能生成的意向性。

从心智哲学讲,意向性通常被定义为心智摄取特定对象的一种内在活动和能力,这个过程被称为意向活动。通过意向能力,被摄取的诸对象能够实现对自身与心智之间物质性时空阻隔的突破,从而成为一种心智关涉项。与动物心理或机器程序不同,即使不存在像孙悟空这样的对象,人类在心智中仍然可以把握它,使之成为心智的相关项。布伦塔诺认为,意向性是区分“有心智者”和“无心智者”的基本标准。^[15]简单而言,意向性就是人在感性活动过程中同时伴随着“意图”的生成,即人能够知道自己的“知道”,能够意识到自己的“意识”,通常体现为“我意愿……”“我希望……”这样的表征。因此,人类活动具有明显的指向性和目的性,它是心智的一种穿越时空指向未来的能力。

马克思感性活动理论的意向性涵义,在他早期的哲学手稿中有着最为充分的表述:“自由的有意识的活动恰恰就是人的类特性”,“正因为人是类存在物,他才是有意识的存在物,就是说,他自己的生活动对他来说是对象。仅仅由于这一点,他的活动才是自由的活动”,“劳动的对象是人的类生活的对象化:人不仅像在意识中那样在精神上使自己二重化,而且能动地、现实地使自己二重化,从而在他所创造的世界中直观自身。”^[16]“意识在任何时候都只能是被意识到了的存在,而人们的存在就是他们的现实生活过程。”^[17]在人类感性活动中,人不但直观到自己的劳动对象,还能够意识到自己的自由本性,意识到自己是一个整体性的、抽象的类存在。劳动生产了美,人按照美的规律来构造,人的意向活动不仅使实物成为自身的意向相关项,还使抽象对象——美、美的规律、自由、类存在——成为自身的意向相关项。马克思感性活动论突破了传统现象学诉诸纯粹意识的意向性,为现象学的多维度发展开启了新的思考方向。^[18]

三、从“感性活动”再思人机融合

人工智能与人类智能存在巨大差异:前者是硅基的、计算的、非自然的集成智能,是纯粹的思维体,它不具有严格的时间性意义,数理运算是其生成性的本质;后者是碳基的、身体的、意向性的自然智能,是有血有肉的生命体,它伴随社会历史的进程逐渐形成,感性活动代表着它的生成特征。笔者无意对两类智能作出高低优劣的价值判断,而是从马克思感性活动论出发,通过甄别两类智能的根本差异,以期寻求二者的共通,试图为两类智能的和谐共生——人机融合智能——提供新的哲学思路。

(一)数理逻辑与辩证性规范的融合

当代美国逻辑学家R·托马森在其著作中区分了“人工智能中的逻辑”(logic in AI)与“哲学的逻辑”(philosophical logic)之关系,他说道:“所谓人工智能中的逻辑,指在人工智能建模中所运用的逻辑,而所谓哲学逻辑,则是指在哲学研究中所运用的逻辑。后者是前者的‘秩序输出者’。”^[19]人工智能的运作机制是数据集成和逻辑运算,人类智能来源于人的感性活动、劳动等各种形式的生命活动,后者遵从辩证性的生成过程,简称“辩证性规范”。^[20]辩证性规范究竟是什么?简单地说,它就是可以灵活表征、非单调推理、直觉决策、随机应变,从而准确有效地处理“人一机—环境”复杂系统的一种思维方法,它不是仅用有约束边界条件的形式化方法计算世界,而是可以产生人的洞察机制的。因此,辩证性规范是一种开发逻辑,把时序和极限有机统一,实现事实、归纳、猜想与反驳序列、价值的一致化。从马克思感性活动角度看,智能不等于意识,人除了有智能,更有丰富的其他意识,智能仅仅凸显人类意识整体中的逻辑推理和计算维度。因此,辩证性规范是人工智能逻辑的来源和基础。^[21]

目前,弗雷格发明的一阶谓词逻辑已逐步发展为当今人工智能科学家所普遍采用的符号逻辑或数理逻辑,然而,生成式人工智能所依赖的各类数学模型仍需要不断革新,理由在于复杂神经网络和深度学习的运算基础已经大大超越了传统命题式的符号逻辑体系。徐英瑾教授归纳了现代数理逻辑的“五宗罪”,他指出符号 AI 所运用的哲学逻辑工具的缺陷有五个:“其一是无法自我检查其所处理的经验性命题自身的真假,其二是真值与内涵的剥离,其三是‘极化思维’,其四是反心理直觉,其五是不考虑信息的贮藏空间的局限问题。”^[22]数理逻辑的这些缺陷,对我们从马克思感性活动论所衍生的辩证性规范角度审视人工智能,从而实现人机融合智能,具有很大的启发意义。

人工智能逻辑只注重事项之间的纯粹逻辑关系,导致休谟所谓“观念的知识”与“事实的知识”的分离。人工智能以削足适履的方式迎合自身高度僵化的逻辑要求,彰显着现代数理逻辑的不可置疑性。换言之,现代形式逻辑仍旧是一种在规则层面上起效的设计活动,与事实无关,因而基于数理思维的人工符号必然会面临“经验事实输出不足”的缺陷。与之相反,马克思的感性活动恰恰把意识与对象、观念与事实依据辩证性规范结合起来,即所谓“意识并非一开始就是纯粹的意识”,原因在于“精神一开始就受到物质的纠缠”。在人类智能的判断活动中,这两类知识不分彼此地共同起作用,无法相互还原。辩证性规范保证了在经验事实的易变形、混杂性和逻辑规律的永恒性与条理性之间,存在着一定程度的收缩能力,因为人类本是能够熟稔地在逻辑、次逻辑和非逻辑之间进行自由穿梭的。因此,辩证性规范守护着人类智能的跨界性、跳跃性和流动性,保留着语言意义模糊的“灰色地带”,但又不至于妨害人类知识的严密性。因此,如何在数学建模的顶层设计中添加某些辩证性规范,或许是实现人机融合智能的有效思路。但话又说回来,要把辩证性规范融入数理逻辑,要么需要比原来更大的数据量和更加复杂的算法结构,要么需要彻底打破已有的顶层设计架构,这的确是人工智能专家面临的一大挑战。

(二)人类意识的形式化与意向性的融合

英国人工智能专家玛格丽特·博登认为,当前制约人工智能进一步发展的瓶颈在于,如何把逻辑算法的形式化与人类独有的意向性进行巧妙结合。然而,人工智能源自形式逻辑框架,人类智能脉于辩证思维体系,人机融合智能的根本在于逻辑与非逻辑的思想结合。逻辑与非逻辑、形式思维与辩证思维,二者似乎井水不犯河水,是永恒的矛盾?马克思感性活动论所蕴含的意向性思想,对我们破解这一难题提供了思路。

感性活动是人类能动的有意识的活动,人类意识突破时空阻隔的指向性以及把过去和未来统统纳入“当前存在”的意识活动,是意向性的重要特征。感性活动论就包含着人类的形式思维与非形式思维,逻辑思维与直觉、想象、矛盾思维的有效融合。马克思生动地说:“蜜蜂建筑蜂房的本领使人间的许多建筑师感到惭愧。但是,最蹩脚的建筑师从一开始就比最灵巧的蜜蜂高明的地方,是他在用蜂蜡建筑蜂房以前,已经在自己的头脑中把它建成了。”^[23]建筑师头脑中的“建筑作品”是非现实的意向性产物,但其设计原理必定要符合建筑学规律,符合结构力学与各项安全要求,如果是高明的建筑师,其作品还要符合美学要求。客观规则本无所谓美,主观审美也无固定规则可言,但建筑师能够在想象中把二者完美结合,创造出一种客观之美——建筑美学。马克思认为“人根据美的规律来构造”,黑格尔所谓“美是理念的感性显现”,都表明客观之美是感性活动与理性规则在创造活动中的融合。人工智能专家刘伟教授在同等意义上说:“在人机融合智能中,意向性是联结事实与价值、真善美的唯一桥梁,形式化可以某种程度地实现这种意向性。”^[24]可见,逻辑与非逻辑、形式思维与辩证思维、意识的形式化与意向性的结合,完全可以在马克思的感性活动论中找到依据。

进一步说,人机融合智能与社会学密切相关。从感性活动论看人机融合,人是一个意识与身体、心理与生理、能动与受动的结合体,具有社会历史性,人类智能是一个时间性的生成物;机器乃是一个数理和物理的聚合体,它是非社会、非历史的,机器智能是一个非时间性的、人工的集成物。但机器智能必定是人造的,是人的意向对象的现实产物,人机融合也不完全是逻辑、形式思维所能实现的,还有人的因素在里面。因此说,机器思维的形式化与人类智能的意向性的结合、逻辑与非逻辑的思想结合,其方式不止是一种简单叠加,还有更为复杂的纠缠。马克思认为人是一种社会性存在,人类智能也是如此,所以智能是群体意向性的产物,包括数理逻辑在内也是群体意向的产物。单个机器属于个体智能,但即使个体的人也属于群体智能。个体智能与群体智能的区别是,个体智能的逻辑如果是 a 大于 b 且 b 大于 c ,则 c 不能大于 a ;而群体智能的逻辑如果是 a 大于 b 且 b 大于 c ,那么 c 可以大于 a 。从今天物理学观点看,马克思诉诸的感性活动过程,开启了“暗信息”这一人类意识发现的“新大陆”,单纯的感受不是智能,没有主体性的、能动性的觉知和理解,就不会存在人类智能的灵动性。因为人的智能不应该完全是计算的,其中充满了非计算的“事”和“情”。因此说,人机融合不同于人类智能,亦不同于人工智能,它是一种面向未来的全新智能,是一种跨物种的、将传统智能属性合而为一的未来智能科学体系。如果说真就是存在,善就是意向,美就是“存在+意向”的融合;假设机是存在,人是意向,那么人机就是“存在+意向”的融合,是“形式化+意向性”的融合,是马克思感性活动中的“人”与数理运作的“机”的融合,是机器智能与人类智慧的融合。这样一来,从今天逐步发展的人机融合到未来的人机之“和”的过程,才更符合马克思把人看作一种“类存在”的思想。

(三)事实与价值的融合

马斯克等科技大佬呼吁暂停开发比GPT-4更强大的人工智能系统,中国出台《生成式人工智能服务管理暂行办法》,这一系列举措反映了人工智能对人类文明的威胁并非空穴来风,已经引起产业界和政府的担忧和防范。生成式人工智能的深度学习机制正在对人类既有的人文资源进行剥夺,加速侵犯人类自然智慧的空间。徐英瑾教授从生存论哲学出发,认为生成式人工智能所依赖的深度学习机制,正是海德格尔所谓的“常人”或“庸人”(Das Man)的技术化表述,因为该概念代表着人类智慧的平均意见,缺乏创见和生机是其主要特色,因而深度学习就在于它以人类常识判断的结论为其自身存在的前提。如果把进行判断的人类个体以及该判断所依赖的人文背景视为广义的“人文资源库”,那么生成式人工智能所依赖的深度学习则被视为攀附在人文资源上的“技术寄生虫”。^[25]它正在快速挥霍人类经由千百年缓慢累积起来的人文资源的红利,自身却难以产生新的人文向度的可能性。长此以往,深度学习对人类自然智慧的侵蚀、占领乃至最终扼杀,极有可能成为人类文明的最大威胁。

人类文明之所以不断创生,绵延至今,源于不同种群和个体在不同境遇中的偶然遭遇,众多偶然遭遇的相互碰撞、交流甚至攻伐,便催生了更强韧、更有生命力的新文明。与此相反,人工智能所依赖的深度学习技术,则是以“取平均值”的方式消灭这种偶然性,使人类多元价值观趋向扁平化,逼迫不同文明下的种群和个体在不知不觉中变成“常人”,最终导致人类自然智慧趋于衰竭。如何在人工智能系统中实现事实与价值的融合,实现机器功用与人类文明多样性的和谐共生,则成为人机融合智能的理想目标。在笔者看来,马克思的感性活动论仍旧给我们提供最有价值的理论资源。

马克思认为,在异化劳动状况下,人类文明、人类价值的增长与自己劳动成果的价值增长呈负相关。他说道:“工人的产品越完美,工人自己越畸形;工人创造的对象越文明,工人自己越野蛮;劳动越有力量,工人越无力;劳动越机巧,工人越愚笨,越成为自然界的奴隶”,^[26]“劳动生产了美,但是使工人变成畸形……劳动生产了智慧,但是给工人生产了愚钝和痴呆。”^[27]马克思虽然批判生产资料资本

主义私有制下的异化劳动,但不可否认的是,当时最先进的蒸汽机和自动纺纱机在劳动过程中的普遍使用,对工人境况的迅速恶化起到推波助澜的作用。因此,马克思认为只有实行生产资料的公共占有,人才能够从事自由劳动,才能使人的“自由个性”成为可能,最终实现每个人“自由而全面的发展”。当前的人工智能可谓最完美、最机巧、最智慧的劳动工具,然而“码农”“程序猿”“加班文化”“网络诈骗”“全民网红”等现象伴随着人工智能产业一同被生产出来。人类思想的扁平化、价值取向的单一化、工具人横行天下以及信息时代的诸多负面后果,倒逼人机融合智能的研发和设计在改进技术手段的同时,必须考虑人类的价值需求。

当前主流人工智能在犯一个类似“经济人假设”的错误,即把人看成单纯的理性人,殊不知人是一个感性活动的人,人除了理性思维之外,还是一种具有直观、情感、情绪、愿望、爱欲等等生动活泼的要素的感性存在,而作为数理性的人工智能对此还无能为力。如果人类一思考,上帝就退隐,那么在计算开始的地方,理解就终结了。马克思的感性活动论就涵盖了人对世界的整体性理解,以及对各种关系的理解,人对世界的价值投射恰好构成了理解的前提和基础。因而怎样让机器学会理解和反思,是人机融合智能面临的两大难题。当前人工智能只使用了人类理性中可程序化的一小部分,距离人类理性差距很大,更遑论接近人类更神奇的部分——感性及其活动。人机融合智能要瞄向对人的感性层面的仿生研究,做到像自然人一样去感受外部世界,去理解人与物、人与人的各种关系,这样的人工智能才是完善的。因此,只有解决未来人工智能如何融合感性元素的问题,使深度学习介入人的感性活动场域,人机和谐才有望真正实现。

注释:

[1]张祥龙:《中西印哲学导论》,北京:北京大学出版社,2022年,“序”,第1页。

[2][4][5][6][7][8][9][13][14][16][17][26][27]《马克思恩格斯文集》第1卷,北京:人民出版社,2009年,第193、189、190、533、189、533、524、533、195、162-163、525、158、158-159页。

[3]本文就大模型基本原理的总结,参考了陈小平教授《大模型:人工智能思想及其社会实验》一文,载于《文化纵横》2023年第3期,在此表示衷心感谢。

[10]王策:《分析哲学思想的溯源研究》,北京:人民出版社,2018年,第133页。

[11][12]Michael Polanyi, *Knowing and Being* The University of Chicago Press, Chicago, 1969, pp. 144, 215.

[15]德国哲学家布伦塔诺是意向性理论的创立者,他的弟子胡塞尔丰富了这一学说并提出意向性理论的基本方法“现象学悬搁”。自19世纪末直到今天,意向性仍然是理解人类心灵和机器智能的重要哲学理论。

[18]法国哲学家梅洛·庞蒂的《知觉现象学》诉诸认知中的身体性,成为现象学运动中的里程碑;此外,当前认知科学中交互隐喻与涉身哲学等理论,均与马克思感性活动论存在或深或浅的理论勾连。

[19]Richmond Thomason, “Logic and Artificial Intelligence”, *The Stanford Encyclopedia of Philosophy*, Winter Edition 2022.

[20]教科书通常把辩证法的具体内容归结为辩证逻辑,即对立统一规律、否定之否定规律以及质量互变规律。在笔者看来,辩证逻辑并不是近代知性意义的、描述事物运动的精细的思维律,它毋宁是一种关于人类意识、行为、情感、意向等多维活动的整体性、内在性的规范,故命名为“辩证性规范”。

[21]关于现代数理逻辑、辩证逻辑与先验逻辑三者之间的关系,可参看王策、金廷:《从康德到弗雷格——论先验逻辑对现代逻辑的奠基》,《科学技术哲学研究》2016年第4期。

[22]徐英瑾:《人工智能哲学十五讲》,北京:北京大学出版社,2021年,第29页。

[23]《马克思恩格斯文集》第5卷,北京:人民出版社,2009年,第208页。

[24]刘伟:《追问人工智能——从剑桥到北京》,北京:科学出版社,2019年,第169页。

[25]参见徐英瑾:《试论深度学习技术对人类社会持续发展所造成的风险》,《当代美国评论》2019年第1期。

[责任编辑:刘毅]