

马克思主义视域下的人工智能与未来治理之道^{*}

高奇琦

内容提要:马克思主义关于机器体系的论述对我们理解未来的通用人工智能具有重要意义。发展人工智能的目的原本是要帮助人类提高治理绩效和解决社会治理问题,然而人工智能的不当应用则可能会产生一系列异化后果。要破除异化,就需要让人回到共同体中,成为共同体中的自由人。在马克思主义看来,生产资料公有制是共同体的必要条件,而未来共同体则需要建立在生产力高度发达的基础上。在智能革命中,世界历史背景下的交往体现为相互依赖中的冲突与合作。可以预期的是,在生产力高度发展的背景下,世界性的制度变革会得以发生,而最终的理想世界是每个人都得到充分自由发展的新型共同体。从马克思主义出发的人工智能哲学是一种人本哲学,其强调人文本位、谨慎乐观和民本主义。

关键词:人工智能 机器 机器体系 马克思主义 异化

作为第四次工业革命中的引领性技术,人工智能正在快速发展,并且深刻地改变着人类生活和世界政治。中国要在新一轮科技革命中赢得主动权,不仅需要在科学技术上勇闯“无人区”,而且更需要在哲学社会科学理论研究上占据高地。作为19世纪以来最伟大的思想家,马克思和恩格斯尽管没有直接讨论过人工智能,但是在《资本论》《1857—1858年经济学手稿》《1844年经济学哲学手稿》《共产党宣言》和《德意志意识形态》等著作和篇章中,他们集中讨论了机器、机器体系、异化、人类解放以及未来理想社会等重要主题。其中在关于机器和机器体系的讨论中,马克思几乎贡献了所有的原创性观点。在异化、人类解放以及未来理想社会等主题中,恩格斯有很多精彩论述。这些讨论对我们今天研究人工智能及其对未来治理的影响具有重要的理论启示意义。

一、劳动的结晶:机器是否产生价值?

伴随着大量无人工厂的出现,人们会经常发问:人工智能能否产生价值或剩余价值?对这一问题的回答,要回到马克思劳动价值论的经典观点之中。具体来看,关于机器能否产生价值或剩

^{*} 本文为国家社会科学基金重点项目“加快数字化发展与建设数字中国的政治保障研究”(21AZD021)的研究成果。

余价值,马克思形成了如下观点:

第一,只有劳动者的劳动才能产生价值,而机器所从事的并不是劳动,因此其并不产生价值。在马克思看来,商品价值的本质是“结晶了的社会劳动”^①。马克思在这里强调的是社会劳动,而不是个体劳动。一个人所从事的是个体劳动,而只有放在整个社会分工的角度来理解,个体劳动才能变成社会劳动。

第二,机器是生产剩余价值的手段,并且其作为工具加剧了资本对劳动者的剥削。马克思将劳动力看成是一种商品,并指出整个资本主义的运行内核是以劳动力为中心榨取剩余价值。并且,机器存在一定程度的无形损耗,其往往以生产相对剩余价值的方式对劳动者形成压迫。^②

第三,机器对劳动者压迫的实质是机器被用作资本主义的手段。在马克思看来,机器本身并没有对错,但是由于机器被用于资本主义的生产过程,那么其就对劳动者形成了剥削。^③

可以看出,马克思所讨论的机器与目前的弱人工智能在特征上非常相近。人工智能分为弱人工智能和强人工智能。弱人工智能也被称为专用人工智能,其在特定领域可以协助人类完成某些任务。强人工智能也被称为通用人工智能,其能够跨领域地处理更为抽象和复杂的知识性问题。^④目前在产业界应用的人工智能依然是弱人工智能,而强人工智能是下一步科学界和产业界着力发展的目标。目前来看,对于弱人工智能,仍然可以适用马克思关于机器的这些结论。在关于人工智能的学术讨论中,研究者会习惯地把人工智能称为机器。关于人工智能的伦理研究往往也被称之为机器伦理学(Machine Ethics)。^⑤因此,马克思关于机器的分析和论断,可以为我们思考人工智能的发展提供宝贵的启示:

第一,人工智能发展的初衷是希望更大程度地解放人类,但目前对其的一些错误运用以及不当实践却可能反过来对人类形成巨大压力。如前所述,弱人工智能基本等同于机器。按照马克思主义的机器不产生价值和剩余价值的观点来看,弱人工智能也不产生价值和剩余价值。弱人工智能作为不变资本进入生产过程。在资本主义的生产条件下,弱人工智能是生产剩余价值的手段,并在某种程度上加剧了资本对劳动者的剥削。人工智能的发展也是如此,其使得工人不得不努力工作以跟上机器的节奏,并且处在随时失业的边缘恐怖之中。正是这种边缘恐怖让劳动者不断地处在高强度的工作之中,这也是马克思反复强调的工人不自由的源头。马克思的这一观点可以提醒我们要时刻思考发展人工智能的初衷。

第二,政治制度对于人工智能的发展至关重要。马克思认为,机器压迫劳动者问题的实质是机器被当做资本主义的手段。马克思阐述的这一问题在智能时代会更加明显。人工智能的发展可能会导致就业失重的问题。人们对就业的准备和思考是按照之前的经验进行的,这样在新的条件下会出现类似于真空中失重的情况,而这种失重可能成为社会不稳定的重要来源。因此如何建立一种对不同社会群体(特别是弱势群体)包容的政治制度,在离心化倾向非常明显的智能时代就变得非常必要。

① 《马克思恩格斯文集》,第3卷,第47页,人民出版社,2009年版。

② 《马克思恩格斯文集》,第5卷,第467—468页,人民出版社,2009年版。

③ 《马克思恩格斯文集》,第5卷,第469页。

④ Kareem Ayoub and Kenneth Payne, “Strategy in the Age of Artificial Intelligence”, *Journal of Strategic Studies*, Vol. 39, No. 5 - 6, 2016.

⑤ Michael Anderson and Susan Leigh Anderson, *Machine Ethics*, NY: Cambridge University Press, 2011.

二、自动的机器体系对价值的复杂影响

上文分析了马克思关于机器的相关讨论,而马克思关于机器体系的阐述则具有更为丰富的内涵,对于我们理解未来着重发展的通用人工智能有非常重要的意义。在一些马克思主义研究者看来,《1857—1858年经济学手稿》是目前马克思主义研究中尚未充分开发的思想资源。^① 马克思在这份手稿中充分讨论了“自动的机器体系”这一概念。马克思关于机器体系的讨论主要有如下观点:

第一,机器体系不是纯粹的工具性存在,而是具有某种类主体性。马克思引用了安德鲁·尤尔(Andrew Ure)《工厂哲学》的一段论述,这段论述中出现了“自我意识”和“受一个自行发动的动力的支配”等表述,^②而这些表述与目前通用人工智能发展的方向是一致的。通用人工智能发展目前要克服的关键问题是,希望可以让智能体拥有自我意识,且进行自我认知和决策。简言之,马克思所描述的“自动的机器体系”,几乎可以等同于当下讨论的通用人工智能。

如果说在之前的论述中,马克思将机器看成是一种工具化的存在,那么在这里马克思更多是从类存在的意义上讨论机器体系。在马克思看来,工人在传统机器和机器体系中的地位和作用是不同的。对此,马克思写道:在传统机器中,工人是主人,而机器是工具。然而,机器体系“本身就是能工巧匠,它通过在自身中发生作用的力学规律而具有自己的灵魂”,^③工人则沦落为工具。马克思把机器体系看成是具有自己灵魂的存在,这就类似于在某种意义上具备意识的通用人工智能。与前文关于机器的论述相比,马克思关于机器体系的观点更具前瞻性,对目前的通用人工智能讨论具有更加直接的启示意义。

第二,作为劳动资料的机器向更加主体性的机器体系的转化,意味着未来生产力的变革性发展。科学技术在机器体系背后发挥着巨大作用,并成为未来生产力发展的关键。马克思已经看到,机器体系的发展对必要劳动的否定是历史趋势。^④换言之,马克思已经认识到科学的巨大影响以及在其基础上机器体系对生产力的重要推动作用。^⑤这启示我们,发展智能相关技术应该成为我国在智能革命中赢得战略主动的关键问题。

第三,最大限度地否定必要劳动时间使得未来产品的价值不断下降,这为未来社会的按需分配提供基础。在生产力极大提高的背景下,商品几乎完全由机器来生产,商品中所蕴含的无差别社会必要劳动时间就会越来越少,其价值也就会越来越小。马克思明确指出:现实财富的创造“取决于科学的一般水平和技术进步”。^⑥这一观点可以理解为是马克思对其劳动价值论的一个新发展。劳动价值论认为,价值是由直接劳动时间的量来决定的,但是由于大工业的发展,现实财富的创造已经越来越少地取决于劳动实践,而越来越多地取决于科学技术。在这样的背景下,所有的商品就会趋向于免费。这种商品价值趋于免费的趋势是马克思在设计未来共产主义社会按需分配的一个重要物质基础。

第四,机器体系的资本主义应用会对劳动者形成更大的压迫,并产生更为严重的社会危机。机器体系产生之前资本家靠剥夺工人的剩余价值来获得利润,而在机器体系下工人连被剥夺剩

① 安启念认为,马克思在《1857—1858年经济学手稿》中关于“自动的机器体系”及其影响的论述,乃至整个《1857—1858年经济学手稿》至今没有得到应有的重视。参见安启念:《马克思关于“自动的机器体系”的思想及其当代意义——兼论马克思主义哲学时代化的文本依据问题》,《马克思主义与现实》,2013年第3期。

②③④⑤⑥ 《马克思恩格斯全集》,第31卷,第88页,第91页,第92页,第99页,第100页,人民出版社,1998年版。

余价值的机会也丧失了即机器体系下出现了工人失业的现象。因此,在资本主义条件下,机器体系并不是为了解放劳动者而出现,同时资本家使用机器体系的目的是为了更大程度地压迫劳动力。马克思同样看到了机器体系可能造成的严重失业问题,并明确指出“达到一定地步,机器就会代替工人”。^① 在第四次工业革命中,这种机器体系对人产生的替代效应可能会更加显著。^②

第五,如果运用得当,机器体系的发展可以为人类自由时间的增加以及最终解放提供物质基础。马克思认为,这种机器体系导致的最终结果是,人的自由时间增加,而这一点则为个人的全面发展提供基础,即为“每个成员创造大量可以自由支配的时间”。^③ 在马克思的理论体系中,自由时间是极为重要的。由于缺乏自由时间,劳动者不得被绑在其劳动岗位上,无法进行知识和技能的提升,也无法从事自己所希望从事的工作。然而,在自由时间充分利用的基础上,劳动者可以逐步实现最终解放。

从上述讨论中可以看到,机器体系与通用人工智能的特征非常相似。马克思关于机器体系的讨论不仅在当时具有超前性,即使在当下仍然有非常重要的未来意义。马克思讨论的机器体系的主体性与目前关于通用人工智能讨论都在强调智能体本身的主体性相关。马克思将机器体系作为一种主体性的存在加以讨论,并且强调劳动资料向机器体系的转化等观点能够帮助我们思考未来通用人工智能的发展方向。

三、智能革命中的异化与新型共同体的构建

马克思在讨论机器和机器体系的资本主义应用及其对劳动者形成新型压迫的批评性文字中,使用最多的概念是“异化”。马克思的异化思想对于我们理解当下日益出现的人工智能异化问题有重要帮助。发展人工智能的目的原本是要帮助我们提高治理绩效和解决社会治理问题,然而人工智能的不当应用则可能会产生一系列的异化。这些异化主要表现在:

第一,人工智能的快速发展可能导致较为棘手的失业问题。人工智能的发展会使得失业问题加剧,而失业又会加剧社会的分裂,并在社会就业的两极中形成巨大的张力。长此以往,这样的异化问题就会演变为社会危机。汉斯·汉森(Hans Hansen)等人的研究则表明这一风险往往被大众忽视。^④

第二,人工智能的发展使得平台型企业的权力得到快速增加。马克·安德烈耶维奇(Mark Anderjevic)认为,大数据的发展可能加剧数字时代的权力失衡。^⑤ 阿里尔·扎拉奇(Ariel Ezrachi)和莫里斯·斯图克(Maurice Stucke)也指出,当一些超级平台企业将触角延伸到物联网、智能设备时,其数据优势就会演化成为一种竞争优势。^⑥

第三,人工智能应用并没有使得人们的自由时间增多。一方面,人工智能的应用在某些方面可以节省人们的时间,例如打车软件可以通过高效率的信息匹配,节省人们的等待时间。另一方

① 《马克思恩格斯全集》,第31卷,第99页。

② 埃里克·布莱恩约弗森、安德鲁·麦卡菲:《第二次机器革命》,第248页,中信出版社,2016年版。

③ 《马克思恩格斯全集》,第31卷,第103页。

④ Hans Krause Hansen and Tony Porter, “What Do Big Data Do in Global Governance?”, *Global Governance*, Vol. 23, No. 1, 2017.

⑤ Mark Anderjevic, “The Big Data Divide”, *International Journal of Communication*, Vol. 8, No. 1, 2014.

⑥ 阿里尔·扎拉奇、莫里斯·斯图克:《算法的陷阱:超级平台、算法垄断与场景欺骗》,第315页,中信出版社,2018年版。

面,人们又会花费大量的时间在一些新的应用之中(例如快手、抖音等),并形成新的沉溺行为。这些沉溺行为为新型平台企业带来巨大流量,也为其进行资本融资提供业绩数据支持。但对于个人而言,结果却是大量自由时间被吞噬。现代社会形成了一种填充机制。现代人在消费主义的指引下将流行产品的更新作为其填充休闲时间的标志,然而却走向了另一种形式的虚无。^①

第四,人工智能会被率先使用在一些黑色产业中。作为一种工具,人工智能技术是中立的,然而其却极为容易被不法分子所利用。例如,智能语音系统被用来拨打骚扰电话进行推销。再如,犯罪团伙会运用深度伪造技术(Deepfake)等生成虚假视频或音频进行诈骗、勒索、舆论操纵等。

马克思和恩格斯对异化问题有非常深刻的讨论,并且对其提出了完整的解决方案。而从他们对异化问题的论述中,我们可以获得对于人工智能异化问题的解决之道。具体而言,马克思和恩格斯对于异化问题的深刻论述及其启示包括以下方面:

第一,异化最本质的表现是关系的颠倒和反转:“对象化竟如此表现为对象的丧失”^②。这种颠倒和反转对于我们理解今天的人工智能有重要意义。人工智能技术本来是人们解决问题和改善生活的工具,但其在发展中越来越变成目的本身。算法是帮助人们解决问题或者辅助决策的优化方案,然而算法越来越成为决策的主导,并且人们对算法的依赖成为一种常态。另外,人们在使用人工智能时,本来希望获得更多的自由时间,但是人们的自由时间并没有增多,反而会花费大量的时间在一些新的应用中。

第二,残酷竞争和非自愿分工是异化的最直接来源,而资本主义私有制则是残酷竞争和非自愿分工的根源。由于这种分工不是个体自愿进行的,那么长此以往劳动者就会感到一种从外向内的强制和压迫,即“一种异己的、同他对立的力量”^③。这种非自愿分工在某种意义上是残酷竞争的结果。在资本主义的竞争条件下,为了避免竞争失败,劳动者被迫接受这种非自愿分工,以致于在长期的竞争文化中,这种非自愿行为似乎会逐渐自愿化。在马克思看来,这种竞争的过度发展无疑会导致人的异化,即人的存在目的和意义逐渐丧失。在资本主义制度的背景下,人工智能无疑会导向更加残酷的竞争。恩格斯明确指出,在资本主义私有制条件下的竞争导致了人类社会的孤立和对立,即“私有制把每一个人隔离在他自己的粗陋的孤立状态中”^④。如果管控不当,人工智能的技术发展会进一步凸显资本主义所有制背景下的个人主义。如果对此不加限制,那么人工智能在应用中就会出现强者愈强的马太效应。

对于异化的破除,马克思和恩格斯的解决方案是新型共同体。其中包含了如下观点:

第一,让人回到共同体中,成为共同体中的自由人。在共同体中的自由不是随心所欲或摆脱自然,而是在掌握自然规律的基础上积极主动地实现目标。可见要破除人工智能的异化,一方面需要人们进一步掌握发展人工智能的规律,另一方面则需要从人本身出发来制定人工智能的发展方案。目前人工智能的发展正处在第三波浪潮之中,而这一波智能技术的主要进步集中在神经网络和深度学习的算法上。^⑤然而,目前这一算法最大的问题在于算法黑箱,即内部的不可解释性。^⑥因此,我们需要进一步发展可解释的智能技术,例如将知识图谱和知识工程作为未来的发展方向,以及推动人工智能从数据驱动的智能转向知识驱动的智能。否则,人类不断增长的情

① 高奇琦:《填充与虚无:生命政治的内涵及其扩展》,《政治学研究》,2016年第1期。

②③④ 《马克思恩格斯文集》,第1卷,第157页,第537页,第72页,人民出版社,2009年版。

⑤ 松尾丰:《人工智能狂潮:机器人会超越人类吗?》,第41—42页,机械工业出版社,2015年版。

⑥ 阿米尔·侯赛因:《终极智能:感知机器与人工智能的未来》,第49页,中信出版集团,2018年版。

性会使得黑箱算法的主导越来越常见,并可能最终演化为算法独裁。算法独裁最终可能会导致人的自主性完全丧失,因此问题解决的关键是回到人。这里的“人”并不是个体的人,而是处于共同体中并获得了充分自由的人。个体的力量非常有限,而只有在共同体中,每个个体的聪明才智和主动性才可以得到有效发挥。因此,破除智能异化的关键是构建这种新型的共同体。

第二,要减弱或消除分工对人带来的压迫和约束性。马克思批判的重心并不是机器或者科学技术,而是运用机器体系的资本主义制度和逻辑。由于资本家要获得更大的剩余价值,因此他们会竭力对工人进行剥削。正是资本的操纵使得人的自由成为稀缺品。在马克思看来,私有制导致了资本家贪欲的不断上升,而且这种贪欲还具有扩散性的特征,导致整个社会无法以互助友爱的方式运行。因此,马克思希望实现的是一种简单的、平凡的、但是又温暖的共同体。^① 马克思所描述的共同体社会可以使得每一个体都得到自由的、全面的、充分的发展。在这样一个社会里,每个个体不再受到非自愿分工的绝对约束,而可以自由地从事自己选择的职业。在智能时代,传统的教育模式、学习内容与互动方式都会面临巨大挑战。之前教育的重心是把劳动者培养成某一狭窄领域的专业人士。这样一种专业区分会影响到劳动者一生的职业选择,而初期的职业教育或路径选择则决定了该个体的职业轨迹。劳动者一旦选择某个职业后,他将很难从一个岗位换到另一岗位。同时,即便自己内心对目前从事的职业和岗位并不认同,但是迫于生活的压力,劳动者也不得不按照要求来履行岗位职责,这就是马克思所指出的由分工导致的不自由。在智能时代,教育方向需要发生根本性的变化,即转向培养全面发展并且可能会在多个领域形成贯通效应的学习者。

第三,实现生产资料的公有制是未来共同体的必要条件。马克思强调,在劳动过程之中,个人劳动力不再是“单枪匹马”的个体行为,而是社会劳动整体行为的一部分。因为社会产品是社会共同拥有的,即“这个联合体的总产品是一个社会产品”^②,所以社会成员可以将部分社会产品重新投入生产,这样可以进一步提高劳动生产率。在智能革命基础上生产力得到高度发展,这已经成为必然趋势。但如果智能革命产生的巨大生产力仅仅掌握在少数精英手中,那么人类社会就会变成分裂的社会。只有对社会产品进行整体性调节才能有效解决这一问题。通过形成整体性调节机制,共同体可以对整个社会的公共产品和剩余进行弹性分配,如通过征收人工智能税,对失业的劳动者进行补偿。失业的劳动者可以利用补偿金通过学习重新进入工作岗位,也可以在不就业或自由就业的状态下保障其基本生活。在整体性调节的保障下,智能时代的社会公平更容易实现。当然,在实现生产力高度发达的目标之前,保持一定程度的私有产权有利于社会激励机制的建立。

第四,未来共同体需要建立在发达生产力的基础上。马克思和恩格斯指出:“如果没有这种发展,那就只有有贫穷、极端贫困的普遍化”。^③ 在之后的人类社会发展史中,巴黎公社、苏联的计划经济以及“大跃进”等尝试将马克思和恩格斯的未来社会构想加以实践,也都遭遇了一些困难,然而这并非马克思和恩格斯的观点错误或方向错误。马克思和恩格斯明确指出,这样的共同体必须建立在生产力高度发达的基础之上,即建立在机器体系对价值重新定义的时代背景之下。而现在看来,此前的种种社会实践都远远没有达到马克思所认为的“生产力高度发达”的前提。

综上所述,马克思和恩格斯的思想对我们理解智能时代下的异化问题有重要帮助。马克思批判的关键是异化,而异化的本质是一种关系的错位或颠倒。机器本来是劳动者的生产工具,但

①③ 《马克思恩格斯文集》,第1卷,第537页,第538页。

② 《马克思恩格斯文集》,第5卷,第96页。

机器却在异化中变为目的,并成为压迫劳动者的工具。同时,马克思客观地分析了机器体系对未来生产力提高的巨大潜能。异化的本质是在资本主义条件下人的机器化。而在马克思和恩格斯看来,要解决异化问题,关键要回到人本身,即回到共同体中去。因此,未来的关键是要以共同体的整体观念来理解智能时代下的社会。人工智能会形成巨大的离心力,可能会将社会撕裂为精英阶层和大众阶层,那么共同体意识在其中就发挥重要的凝结作用,即把整个社会紧密地联系在一起。因此,中国提出的全面建成小康社会的理念和精准扶贫的政策就变得至关重要。相较而言,西方已经出现了社会分裂的明显态势。在智能革命的背景下,人工智能已经在安全风险、公民隐私和公平难题等方面对国家治理提出新的挑战。^①同时,在智能革命加剧的结构性失业影响之下,人工智能的离心作用会进一步加大,并可能导致更大程度的社会危机。只有通过建构新的共同体意识和采取切实的共同体建设措施,才能够真正破除异化难题,推动人工智能的健康发展以助力人类社会的解放实践。

四、智能革命时代的世界历史与人类命运共同体

人工智能的革命性影响不仅体现在一个国家的国内治理实践之中,而且也反映在世界范围内的全球治理实践之中。马克思使用的“世界历史”概念与当下的“全球治理”内涵非常接近。马克思主义对世界历史的讨论对我们从整体上思考智能革命与全球治理的关系具有重要意义。马克思主义关于世界历史的观点,主要为如下几点:

第一,工业革命对世界历史的形成具有推动作用。马克思和恩格斯肯定了工业革命对世界历史的开创意义,“因为它消灭了各国以往自然形成的闭关自守的状态”^②。并讨论了在发达国家工业革命的背景下,不发达国家被动卷入的情景。^③也正是因为不发达国家的被动卷入,世界历史得以形成。智能革命对世界历史和全球治理的形成也具有重要的推动作用。在马克思主义看来,世界历史就是将世界作为一个整体进行思考和行动。这就启发我们,全球治理同样是把人类社会的问题放在全球的整体思维中加以考虑,并构画其解决方案。智能革命所带来的生产力变革以及全球范围内的互动增加,都使得全球治理的实现更为可能。

第二,世界历史背景下的交往体现为相互依赖下的冲突与合作。马克思和恩格斯讨论了国际产业竞争对不发达地区的影响:“由广泛的国际交往所引起的同工业比较发达的国家的竞争,就足以使工业比较不发达的国家内产生类似的矛盾”。^④这一点在智能革命中表现得更加明显。新一轮的智能革命最初会在工业发达国家出现,然而作为结果,那些发展中国家更容易受到智能革命的影响和冲击。长期来看,受到自动化冲击最大的不是发达国家,而是劳动力密集型的发展中国家。^⑤在实际运行过程中,世界历史首先表现为相互依赖下的不断冲突,而这种冲突在智能革命的背景下会更加常见。例如,西方发达国家(特别是美国)希望进一步控制智能技术的主导权和话语权,不希望这一技术向发展中国家转移。美国针对华为和中兴的一系列事件,都表明了美国的这一倾向。马克思和恩格斯深刻地看到,在工业革命过程中工业发达国家和不发达国家之间存在巨大紧张关系。同时,要克服这一问题,恰恰需要从马克思的交往概念出发。交往体现在以开放和平等的态度来看待人工智能的技术进步。一方面,发达国家要更加开放,并接纳发展

① 高奇琦:《智能革命与国家治理现代化初探》,《中国社会科学》,2020年第7期。

②③④ 《马克思恩格斯文集》,第1卷,第566页,第567页,第568页。

⑤ 埃里克·布莱恩约弗森、安德鲁·麦卡菲:《第二次机器革命》,第252页。

中国在新兴领域的技术进步。另一方面,发展中国家要抓住新工业革命的机遇,积极进取并尽力取得跨越式发展。

第三,在生产力高度发展的背景下,世界性的制度变革可能会出现。马克思和恩格斯强调,未来理想社会的创建一方面是以发达生产力为重要前提,另一方面要“以生产力的普遍发展和与此相联系的世界交往为前提”。^① 马克思和恩格斯强调了各民族之间交往的重要性。通过这种交往,各民族之间的世界分工彻底被消灭,并且在此基础上人类历史演变为世界历史。未来世界制度性变革的基础是生产力的高度发展,而人工智能是新一轮科技革命的关键技术。未来生产力高度发达的关键,在很大程度上取决于该国在智能相关技术上的发展。从这一意义上讲,中国在人工智能的发展上就具有非常重要的历史使命。中国代表发展中国家来参与世界范围内人工智能新话语权的竞争。

在工业革命发展史上,发展中国家第一次进入最为前沿的领域并与发达国家来展开竞争。在前三次工业革命中,主导者都是西方发达国家,而这一次中国代表发展中国家来参与竞争。这里反复强调中国代表发展中国家,这就意味着中国在技术的相关问题上需要更加开放性地对待发展中国家。中国将来需要将前沿技术更广泛、更深入地同发展中国家开展应用、分享和治理,这也是中国提出的“一带一路”倡议和人类命运共同体思想的基本出发点。^② 中国强调自己是“第三世界”的一员,^③同时强调自己是发展中国家的代表。中国反复强调,要让发展中国家分享中国改革开放带来的红利。同样,中国在新技术革命中取得的领先地位和先进成果也会同发展中国家共享,这就是马克思主义意义上的世界历史的最终变革。

第四,最终理想世界是每个人都得到充分自由发展的新型共同体。在人类解放过程中,每一个单个的人都需要被关注,即“每一个单个人的解放的程度是与历史完全转变为世界历史的程度一致的”。^④ 因此,新型共同体还需要回到对人的关心上。从本质上看,人类命运共同体要使每个个体都得到充分的发展。人类命运共同体是全球范围的“自由人的联合体”,而人工智能恰恰可以提供物质基础,这是因为人工智能首先是一次普遍性的生产力革命,同时人工智能可以更有效地赋能弱势群体。例如,在人工智能技术的辅助下,低视力人群可以降低其脆弱性。^⑤

整体而言,交往理论的物质基础是要在全世界范围内形成较为发达的生产力,这就需要各国不能狭隘地看待技术进步,即不能把技术进步看成是某一国家或某一群体的特殊利益,而要将其看成是人类的整体利益。如果不能对人工智能的进程进行充分掌控,未来国际社会很可能会形成较大的技术鸿沟。而要避免这一鸿沟,就需要用开放共享的态度来看待技术,要让发展中国家的科学家、工程师和民众更加容易接触到这些技术。人工智能自身的一些特点是有利于学习的,但是如果发达国家以控制智能技术为目的,通过专利或其他手段来限制智能技术的利用和传播,那么就会导致一个割裂世界的出现,而发展中国家的技术发展就更加遥不可及。但是,如果用开放的态度看待这些技术进步,用交往的心态来理解世界历史,那么人工智能助力人类命运共同体

① 《马克思恩格斯文集》,第1卷,第539页。

② 习近平:《迈向命运共同体 开创亚洲新未来——在博鳌亚洲论坛2015年年会上的主旨演讲》,《人民日报》,2015年3月29日。

③ 《毛泽东文集》,第8卷,第441—442页,人民出版社,1999年版。

④ 《马克思恩格斯文集》,第1卷,第541页。

⑤ Zhi Heng Tee, Li-minn Ang and Kah Phooi Seng, “Smart Guide System to Assist Visually Impaired People in an Indoor Environment”, *IETE Technical Review*, Vol. 27, No. 6, 2010.

的实践就更加可行。在生产力高度发展的前提下,人类社会在平等互助的基础上进行交往,那么马克思主义意义上的“自由人联合体”就可以进行更大范围的扩展。

五、马克思主义视角下的人工智能新哲学

在未来共同体中,实现的解放是作为种类的人的解放。这种解放是“以宣布人是人的最高本质这个理论为立足点的解放”。^①换言之,人的解放最终是人的本质的解放。要实现人类解放,关键要有新的理论指导。这一部分主要讨论从马克思的基本立场和观点出发的未来人工智能哲学。从马克思主义的角度来讲,这里的理论就是哲学,也就意味着我们要形成人工智能时代的新哲学,而且这里的哲学主要是指政治哲学。目前,与人工智能相关的理论研究中,已经出现了较为丰富的认知哲学,^②并成为科学哲学的重要组成部分。然而,认知哲学并不是本文所要讨论的重点。西方政治哲学中关于人工智能的讨论已经出现一些成果,但是从马克思主义的视角来看,西方的人工智能哲学实际上存在严重的问题,这主要表现为如下三个特征:

第一,智能体本位,即极为强调智能体的主体性,而相对忽视人的主体性。例如,作为西方机器伦理学界的代表性学者,温德尔·瓦拉赫(Wendell Wallach)和科林·艾伦(Colin Allen)希望设计主体性极强的人工道德智能体(artificial moral agent)即道德机器,使得机器可以明辨是非。^③他们希望,智能体可以独立进行道德判断,并通过“道德图灵测试”(moral Turing Test)。^④近年来,美国等西方国家发布了一系列人工智能的相关准则,例如“阿西洛马23原则”、谷歌AI原则、英国上议院特别委员会的AI代码五项原则、欧盟的AI伦理指导方针,这些原则几乎都较为强调智能体的主体性,而对人类主体性的强调明显不足。

第二,悲观论,即强调人工智能发展的最终结果将是人类的灭亡。西方学者在描述人工智能的政治后果时,几乎都是完全悲观的。^⑤另如,阿约巴和佩尼布认为,尽管专业人员的预估值从20年到几百年不等,但长期来看,人工智能可以发展出与人类相匹配甚至远远超过人类的智能。^⑥这些观点无疑都对人工智能的未来影响表现出了非常悲观的倾向。

第三,超人主义,即通过激进的方式来推动人类进化,并使其将来可以与人工智能有更好的竞争。人类在脑机接口、可穿戴设备以及人体增强等技术的支持下拥有更强大的能力,同时,在基因编辑以及生物工程等技术的辅助下,人体可以更快地进行进化,这样就可以与人工智能进行充分地竞争。^⑦这种超人主义背后的思想可以溯源到尼采哲学。^⑧西方人工智能的产业推动者们都在实践尼采的超人哲学。诸如埃隆·马斯克(Elon Musk)、杰夫·贝佐斯(Jeff Bezos)这样的

① 《马克思恩格斯文集》,第1卷,第18页。

② 重要成果参见 Paul Thagard, *Computational Philosophy of Science*, Cambridge: MIT Press, 1988; John Haugeland, *Artificial Intelligence: The Very Idea*, Cambridge: MIT press, 1985.

③ Wendell Wallach and Colin Allen, *Moral Machines, Teaching Robots Right from Wrong*, New York: Oxford University Press, 2009, pp. 3-11.

④ Colin Allen, Gary Varner and Jason Zinser, "Prolegomena to any Future Artificial Moral Agent", *Journal of Experimental Theory of Artificial Intelligence*, Vol. 12, No. 3, 2000.

⑤ 杰瑞·卡普兰:《人工智能时代:人机共生下财富、工作与思维的大未来》,第200页,浙江人民出版社,2016年版。

⑥ Kareem Ayoub and Kenneth Payne, "Strategy in the Age of Artificial Intelligence", *Journal of Strategic Studies*, Vol. 39, No. 5-6, 2016.

⑦ 尼克·波斯特洛姆:《超级智能:路线图、危险性与应对策略》,第43—52页,中信出版社,2015年版。

⑧ 尼采:《查拉图斯特拉如是说(译注本)》,第9—12页,生活·读书·新知三联书店,2007年版。

产业精英在推动人工智能发展时,很大程度上并没有考虑绝大多数人的想法。

智能体本位、悲观论和超人哲学都受到西方基督教文化的影响。在西方基督教文化中,上帝是第一位的,而人匍匐在上帝脚下。只是在近代启蒙运动以来,人的主体性才得以凸显,因此人的主体性是相对短暂的。因此,在未来发展中,人的主体性被机器的主体性所取代,在西方主流文化看来是极为正常之事。在基督教文化中,只有上帝是永恒的,而其他(包括人的主体性)都是短暂的。悲观论的源头也可以追溯到基督教,即世界末日是基督教的主题叙事。因此,在西方主流文化视野中,人工智能的发展最终就是走向世界末日。超人哲学的实践则类似于耶稣拯救世界的一个世俗版本。人工智能的产业精英把自己想象成耶稣,从而希望可以通过各种激进的方式来拯救世界。

相较之下,从马克思主义出发的人工智能哲学乃是一种人本哲学。马克思高度关注和强调人在社会中的主体性作用,即创造性实践的过程。这种人工智能新哲学将主要体现在如下几个方面。

第一,人文本位,即将人的主体性以及人本原则作为人工智能哲学的基础。对于此,笔者提出了“人工智能新三原则”:(1)智能体永远是辅助;(2)人类决策占比不低于黄金比例;(3)人类应时刻把握着人工智能发展的节奏,并随时准备好暂停或减速。^① 强调智能体辅助的原则是希望可以突出人类对自身命运的主导。同时,人类决策占比不低于黄金比例也是希望人类不能将所有的决策都交给算法,而是要有一个限度。当然,如何衡量决策比例是一个高度复杂的技术操作问题。同时,人工智能的发展也不应该总是线性的。就像核武器一样,如果这一技术的发展最终会导致人类的灭亡,那么及时暂停或减速将是必要的。

第二,谨慎乐观论,即对人工智能的发展未来表示乐观,但同时持谨慎态度。马克思对人类社会的未来是乐观的,因此提出了新型共同体这样一种未来社会的构想,而人工智能的发展为这样一种未来社会可以提供物质基础。这种对未来的判断是乐观的,但同时这种乐观也不是盲目的。在人工智能的发展过程中,人类需要辅之以严格的伦理原则和法律制度,需要用长期的制度性行为来保障人工智能的健康发展,确保其目的是为人类解放服务而不是增加人类社会的冲突和紧张。

第三,以民为本。这种立场是与西方的超人主义相对应的。人工智能的发展会产生巨大的离心机效应,而人类社会的利益分化与冲突会进一步显现,因此需要一种向心哲学。以人民为中心的民本主义哲学更多强调从弱势群体角度出发来思考问题,可以更好地协调各方面的利益,以保障人工智能的健康和良性发展。例如,在这一观念指引下,未来智能税的征收就显得尤为必要。征收智能税对失业群体进行补偿,这样就可以使失业群体的基本生活得到保障,避免失业者对未来的智能社会做出颠覆性举动。

六、结语

在马克思和恩格斯的经典文本中虽然未出现人工智能这一用语,但是马克思和恩格斯对机器、机器体系、异化、人类解放等主题的分析,可以对当下人工智能的思考提供重要启示。从马克思和恩格斯观点出发的人工智能思考与他们的哲学、政治经济学以及科学社会主义之间存在着内在的紧密关联。马克思主义政治经济学的基础是劳动价值论。马克思希望通过这一理论来抵

^① 高奇琦:《全球善智与全球合智:人工智能全球治理的未来》,《世界经济与政治》,2019年第7期。

抗资本主义对劳动者的压迫。这一理论反映的是一种以人为本的人文关怀和价值取向。马克思关于机器体系的讨论则看到了科学技术改变未来社会的巨大潜能,同时也看到了机器体系可能对劳动者所形成的新的压迫。这种对新压迫的关心是马克思以人为本思想的一种延续。这种压迫实际上就是马克思意义上的异化,而在智能革命时代人工智能同样可能会出现异化。异化的本质是人主体性的缺失。而要避免异化,关键就是要回到人本身,回到共同体中重新找到人。这种整体观在智能革命时代至关重要。我们应该一方面积极地推动人工智能的发展,使得人工智能对生产力推动的巨大作用充分发挥出来,同时还要谨记马克思主义对机器体系资本主义应用的批判,避免机器体系滥用导致的智能异化。

马克思主义视域下的人工智能思想,首先从马克思主义政治经济学入手,但最终要回到马克思主义哲学和科学社会主义。要破除异化,马克思认为需要新的哲学。构建新时代的人工智能哲学意义重大。未来社会变革的最终目标是人类解放,其可以通过世界历史的交往过程来实现。无论是国内社会还是国际社会,都需要各方行为体在生产力高度发达的前提下,通过开放和共享的思维来分享科技进步的成果,而不是少数精英通过控制科技进步为其特殊利益服务。社会主义制度恰恰可以提供一种整合机制。全面建成小康社会的理念和精准扶贫政策等把整个社会凝结在一起,这对人工智能来临之后形成的紧张关系有较好的消除作用。在这样一个背景下,通过社会主义的长期实践和人类命运共同体的建设,马克思主义经典作家理想中的人类解放和自由社会的形成就更加可能。通过这种开放性的交往过程,人工智能带来的生产力红利就可以从内到外来扩展,助力实现建设人类命运共同体的最终目标。人类命运共同体是人类解放的另一种重要表述,其目的是让每一个人都进入自由和全面发展的状态。

作者:高奇琦,华东政法大学政治学研究院(上海市,201620)

(责任编辑:刘杰)