

· 建构中国就业理论体系 ·

引领人工智能创造更多更高质量就业

蔡 昉*

内容提要 如同任何重大的技术创新，人工智能的发展也具有双刃剑性质。从就业影响来看，既可以在提高生产率的基础上创造新岗位，实现就业创造，也可能通过人工智能或机器人替代劳动者，导致就业破坏；并且可能出现就业创造在数量上少于、在时间上滞后于就业破坏的现象。正反两方面的效应孰轻孰重，归根结底取决于人工智能发展和应用的导向。本文围绕人工智能发展向就业优先战略和理念看齐这一命题，探讨重塑人工智能技术变迁动机的理论依据和实践路径。首先，讨论法律和政策如何引导人工智能创新者，使就业优先理念贯穿于技术发展的全过程。其次，通过论证减少劳动力投入不是提高生产率的唯一途径，以及提高生产率并非技术创新的唯一目的等命题，提出引导人工智能发展不以替代就业为动机的观点。最后，将人工智能的就业影响置于创造性破坏的分析框架内，提出基于理论和经验的政策建议。

关键词 人工智能 就业优先战略 对齐 技术变迁动机

一 引言

习近平总书记在二十届中央政治局第二十次集体学习时强调，人工智能带来前所未有的发展机遇，也带来前所未遇风险挑战。要把握人工智能发展趋势和规律，加紧制定完善相关法律法规、政策制度、应用规范、伦理准则，构建技术监测、风险预警、应急响应体系，确保人工智能安全、可靠、可控。习近平总书记在二十届中央政治局第十四次集体学习时强调，要坚持以人民为中心的发展思想，全面贯彻劳动者自主就业、市场调节就业、政府促进就业和鼓励创业的方针，持续促进就业质的有效提升和

* 蔡昉，中国社会科学院，电子邮箱：caifang@cass.org.cn。本文是中国社会科学院“学部委员工作室”、国家自然科学基金专项项目（批准号：72141310）的阶段性成果。

量的合理增长,不断增强广大劳动者的获得感幸福感安全感,为以中国式现代化全面推进强国建设、民族复兴伟业提供有力支撑。这两次重要讲话精神,是我们认识人工智能发展对就业影响的根本指导。人工智能的突破性发展,既可以成为创造更多更高质量就业的机遇,也可能造成技术性失业等不利后果。应对前所未有的挑战、抓住前所未有的机遇,一个重要的着力点是通过制定和实施相应的法律和政策,规制、规范和引领人工智能发展,确保其符合就业优先战略和理念。

人类发展史上的许多革命性技术突破,都被证明具有双刃剑的性质,人工智能技术的发展和运用也不例外。科技专家观察到大量现象,表明新技术扩散和运用的过程与技术发明的初衷之间,常有南辕北辙般的分歧(穆斯塔法·苏莱曼、迈克尔·巴斯卡尔,2024)。经济学家达龙·阿西莫格鲁等人也指出,人工智能技术天然具有提高人类能力和替代人类工作两种倾向(Acemoglu & Restrepo, 2020)。这两种倾向之间既有一致性,也存在不同之处。关于人工智能技术在提高人类能力与替代人类工作的方向上,理论和现实中都存在不一致之处,既是指劳动生产率提高与劳动替代并不必然是同一枚硬币的两面,也是指能够提高生产率的技术并不必然创造新的岗位。

所以,阿西莫格鲁等人特别强调,人工智能只是一个技术平台,固然可能通过自动化破坏岗位,但也可以通过改变生产流程创造生产率更高的新岗位。在后一种情形中,人工智能及其应用被称为“有助于重归就业的人工智能”,或“正确的AI”,或“正确的人工智能应用场景”。双刃剑的特点则表现在,人工智能技术既可以走上“正确的”道路,也可能走上“错误的”道路。布林约尔松及合作者也持有相同的观点,甚至为人工智能产生影响的几个重要方面,都分别指出了一个岔路口,由此可以走向正道或者歧途,从而产生大相径庭的结果。例如,取决于生产率表现,可能出现经济增长与停滞两种情形;取决于收入分配状况,可能出现较为平等与高度不平等两种情形;取决于产业集中状况,可能出现技术均衡分享与垄断两种情形(Brynjolfsson & Unger, 2023)。

人工智能是人类创造力的产物,如今应用最广泛的大语言模型也是由科学家用人类知识数据所训练,用户则按照自己的意图予以应用。因此,人工智能为人类利益和福祉服务是天经地义的。人工智能的双刃剑性质,固然包含着解决问题的可能性,但是,挑战总是最先显露出来。正如任何新技术都不会在所有的方面具有自动造福人类的涓流效应一样,这些新技术的负面效应能否显现出来或者得到遏制,也属事在人为。人工智能对就业积极效应的释放或者对消极效应的抑制,都要求技术发明者、投资者、企业家和用户取得尽可能一致的认识,从而在技术创新过程中具有共同的取向和行动优先序。

本文在以下部分中首先将讨论如何确立社会各当事人之间取得共识的基准,以及

人工智能发展向人类对齐的标准这一基础问题，指出就业优先战略和理念，应该成为人工智能发展和在经济社会领域应用的根本引领。在此基础上，本文探讨如何在技术发展过程的始终植入创新向善的基因。进一步，本文还将提出一个重要的理论命题并结合经验说明，即岗位替代并不是人工智能发展的必要动机。最后，本文就应该如何利用人工智能的技术能力，解决技术本身带来的矛盾等问题提出政策建议。

二 向就业优先战略和理念看齐

一个现实的悖论是，人类智慧是人类劳动的必要根基，人工智能的发展逻辑也是替代人类劳动。所以，人类智慧与人工智能之间的竞争关系，也表现为人类如何保护自身就业权利的努力。其实，尽管这种竞争关系始终存在且愈演愈烈，直到雷·库兹韦尔（2011）意义上的“奇点”来临，即人工智能超越人类智慧并开始统治人类之前，关于“鸡生蛋还是蛋生鸡”的因果关系仍然是明确的：正是人创造出人工智能，训练了各种模型，使用了智能机器人，而不是相反。所以，人类在这场竞赛中首先是主宰，拥有主动权，其次才是参赛的一方。这使得人类要主动身兼规则制定者、裁判员和运动员等多重身份。

迄今为止，要求人工智能与人类意愿对齐，仍然是发展和驾驭科学技术的应有之义。虽然人工智能发展的必然逻辑，是走向人工智能代理之路，即无需人工干预机器自行做出决策，这也不意味着人工智能代理不能被引导和规制。诺伯特·维纳最早提出“看齐”（alignment）的概念：当我们使用智能机器服务于自身目的时，如果在过程中不确定能够有效干预其运行，则需要从一开始就确保机器的作为是我们真正所希望的（Wiener, 1960）。可见，看齐或对齐的目的在于引导人工智能符合人类整体或群体的预期目标、偏好及伦理原则。然而，虽然对于要不要对齐已经成为不言而喻的问题，与谁对齐的问题却尚未得到解决。这是因为，正如人类的世界观千差万别，对于现实的看法、所持的态度和采取的行动总是大相径庭一样，人们对于人工智能的发展也必然持有莫衷一是的观点，进而倾向于采取不同的策略。

从理念上看，人类尚有大量的观念需要从认识论、方法论、道德和伦理方面取得最起码的共识。迈克尔·桑德尔（2012）在哈佛大学开设了一门最受欢迎的通识课——“公正”，在其中他列举了哲学家津津乐道的种种道德悖论现象，介绍了从亚里士多德到伊曼纽尔·康德，从杰里米·边沁到约翰·罗尔斯，以及从约翰·穆勒到米尔顿·弗里德曼等思想家的对应答案。然而，最终的结论竟然是，在这些关乎人类道德取舍的著名悖论上，无论是跨越不同时代的哲学家，还是普通人，远未能够取得一

致的看法，更遑论遵守共同的行动守则了。

从政策角度看，对人工智能发展及其立法的实践，各国的表现也千差万别。以美国更倾向于采取自由放任策略、欧盟更注重以立法和规制约束人工智能发展这样两极态度为代表，各国分别以各自的方式，就人工智能及相关技术制定法律和规范性文件。例如，欧盟于2024年率先推出全球首部人工智能全面监管法案《人工智能法案》，旨在统一人工智能领域的各种规则，促进内部市场发展，提高人工智能所体现的自由流动性；在保护健康、安全 and 基本权利的同时，防止产生对公共利益的有害影响，同时保持对技术创新的支持。美国则是制定了相关的框架式文件，为企业提供自愿性的指南，在地方一级制定了限制性措施，同时正在拟议一些专项法案。此外，其他人工智能研发和使用大国也通过立法、发布政府文件、制定规范框架等方式，在或大或小的程度上规范人工智能的发展。

中国作为处于人工智能发展前沿的科技大国，通过立法、规制和规划等方式，从技术研发、产业化和伦理规范等方面，为人工智能的规范发展提供了对齐标准。值得特别关注的此类文件有两项。第一，2021年国家新一代人工智能治理专业委员会发布的政策性指导文件《新一代人工智能伦理规范》，从管理、研发、供应和使用等方面对人工智能发展做出了伦理规范；特别强调从机会均等、保护权益、分享收益和帮助弱势群体等方面入手，坚持以人为本、增进人类福祉和促进公平公正。第二，国家互联网信息办公室等七个部门于2023年联合发布《生成式人工智能服务管理暂行办法》，明确要求坚持社会主义核心价值观，也规定采取有效措施防止各种歧视现象。该文件特别强调把提倡什么和防止什么的做法，都体现在算法设计、训练数据选择、模型生成和优化、提供服务等过程中。上述规范文件强调的这些重要理念，可以说是在高屋建瓴的层次，对中国人工智能发展对齐问题作出的引领。

人工智能的对齐问题，归根结底需要从两个层面上解决。在稳定和扩大就业问题上的对齐问题，同样需要对这两个层面的基本问题给出答案。在第一个层面上，需要回答以谁为伦理标准看齐，或者向哪些发展理念对齐的问题。在就业问题上，应要求人工智能发展进行看齐的理念，包括所有我们在讨论就业优先战略时提出的原则。例如，坚持以人民为中心的发展思想，必然要求通过作为民生之本的高质量充分就业，提高居民收入和民生福祉，使人人都有充分的动机和平等的机会，通过劳动攀登社会流动阶梯，实现人的全面发展。从技术角度来说，对齐则意味着要求人工智能及其应用所执行的功能，具体而言就是在算法设计、训练数据选择、模型生成和优化、服务提供等过程中，始终服务于就业优先战略，努力做到就业创造大于就业破坏。在第二个层面上，需要回答通过何种引导方式和激励手段实现一致性对齐的问题。本文将在

第三、第四和第五部分对相关问题进行讨论：一是探讨应该如何引导人工智能投资者和开发者，把就业优先理念体现在技术本身；二是探讨应该如何引导人工智能的应用者或用户，能够自觉地做到不以替代就业为动机。

三 重塑人工智能技术变迁动机

本部分的标题涉及两个与对齐最相关的命题：第一，自动化最擅长的减少劳动力投入的做法，并不是提高劳动生产率的唯一途径；第二，提高劳动生产率也不应该是人工智能应用的唯一目的。传统经济学理论认为，技术变迁是由节约稀缺要素的需求引致发生的。对于任何具有更大稀缺程度从而更高相对价格的要素，节约要素成本的动机和市场创造的激励，都会推动以减少该要素使用为目标的技术创新（Hayami & Ruttan, 1985）。不过，这个理论的有效性依赖于一个条件，即存在着资源约束和技术制约。一旦这类约束被实质性打破，诱致性技术变迁理论的效能便大大弱化。

人工智能及其赋能的技术进步具有提高劳动生产率、创造充足物质财富的无限能力，因而正在不断拓展资源约束的边界，从而打破传统理论范式，技术变迁不再需要要素相对稀缺性或相对价格信号来引导，而是可以做到指哪儿打哪儿。第一，技术可望达到这样的程度：要素不再具有绝对或相对稀缺性，技术创新也就不需要诱致性过程。第二，通过大数据分析和机器学习，人工智能可以预测未来需求和技术趋势，主动引导各层面创新。第三，人工智能可以超越物质需求，识别变化着的社会性需求，而无须通过市场信号。

由此引申的政策含义是双重的。一方面，即便在劳动力短缺的禀赋特征消失的情况下，最新的技术应用也可以不顾市场信号，遵循惯性以人工智能和行为能力替代劳动力。另一方面，无论劳动力是短缺还是过剩，新技术都可以做到忽略不计，并不必然引致出节约劳动的技术变化。这两个方面的变化都意味着，技术变迁过程与就业替代结果之间已经具有脱钩的可行性。所以，人工智能发展对于就业的影响方向如何，越来越是主观的选择，而不是客观必要性。

迄今为止，各产业、行业的自动化需求，仍然产生于劳动力短缺和技能人才不足。对中国来说，自动化或替代就业的技术变迁，受到老龄化和劳动力短缺，以及工资成本上升的要素禀赋驱动。根据国际机器人联合会发布的数据（Müller, 2024），在2018–2023年期间，全球运行中的工业机器人总量年均增长12%，目前达到428.2万台；同期中国的年均增长率为17%，达到176万台，在全球占比为41.1%。在2023年全球54.1万台新安装的工业机器人中，中国为27.6万台，占比高达51%。这种机器人替代劳动者

的趋势与实际工资增长趋势密切相关。根据国际劳工组织的数据（ILO，2024），进入21世纪以来，工人实际工资水平在各国均有显著提高，各国之间的工资差距也有所缩小。在很大程度上，这是由于在中等收入国家，人口年龄变化导致劳动力短缺，因而工资增长速度明显快于其他收入组的国家。作为中等偏上收入国家，中国对全球工资水平提高和差距缩小的贡献最为显著。

根据国际劳工组织这一口径，在2006–2021年期间，中国工人的平均工资实际增长了128%，不仅比全球平均增长率（57%）快1.24倍，而且分别大幅快于低收入国家、中等偏下收入国家、中等偏上收入国家和高收入国家的平均增长水平。这一劳动要素相对稀缺程度加剧和工资上涨趋势，在中国始于2004年的民工荒现象，并在理论上被称为刘易斯转折点（蔡昉，2022）。随后，用不同统计口径表示的劳动力资源，相继经历了从增长到减少的转折点。如图1所示，15~64岁劳动年龄人口在2013年、城乡劳动力在2015年分别达到峰值，随后进入负增长。同时，城镇就业人口也已处于平缓的顶部，继续增长的势头基本消失。劳动力短缺改变要素稀缺程度和相对价格，诱致产生节约劳动力的技术应用，一直以来都被认为是产业优化升级的表现，也被认为是中国经济转向新阶段的特征。然而，由于人工智能及其赋能的数字经济发展，具有巨大的生产率提高潜力，以及显著的成本降低空间，因此值得担忧的是，在劳动力短缺程度得到缓解，甚至结构性就业矛盾已经显现的情况下，节约劳动力的技术变迁仍会继续。这意味着人工智能对就业的冲击可能趋于长期化。

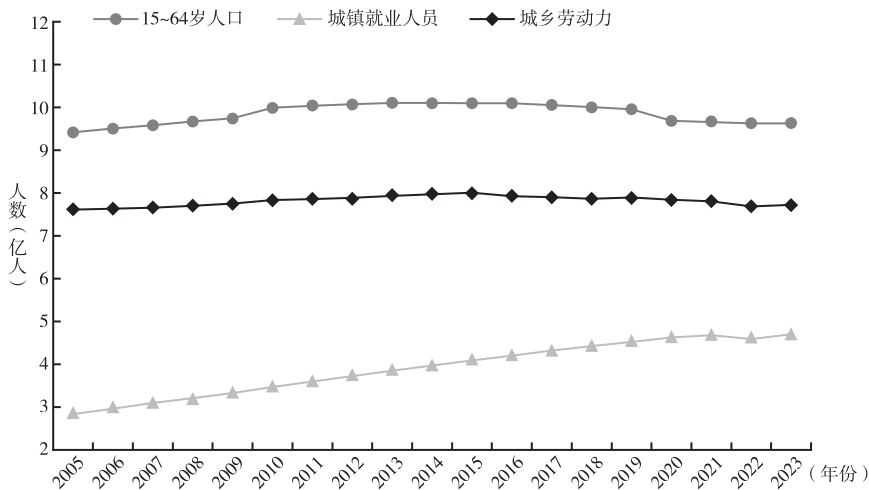


图1 中国进入劳动力短缺时代

资料来源：根据国家统计局“国家数据”（<https://data.stats.gov.cn/easyquery.htm?cn=C01>）计算得到。

对于是否存在将技术变迁过程与就业替代结果相分离的可能性，从现有的理论和经验出发，可以提出两个命题来回答。第一，提高劳动生产率虽然是人工智能助力自动化的重要目标，但减少劳动力投入却不是提高生产率的唯一途径。现实中，人工智能提高生产率的途径可谓多矣。例如，劳动生产率计算公式表明，在包括减少劳动投入在内的通过降低分母达到提高生产率目标的方式之外，还可以通过增加产出、扩大销售和增加营业收入等提高分子的方式，达到相同的目标。这种方式包括优化资源要素的配置、提高产品和服务的质量、畅通产业链供应链、改进工艺流程和创新产业模式、改善用户体验和消费者感受等具体途径。

第二，提高劳动生产率也不是技术创新的唯一目的，有很多与提高生产率同等重要的目标，同时与创造就业的效果兼容。这与阿西莫格鲁等人指出的“正确的AI”方向有异曲同工之妙。具体而言，一是人文、社会发展目标。通过赋能儿童发展与教育、医疗与公共卫生、养老服务事业与产业等，直接服务于促进人的全面发展。二是可持续发展目标。包括赋能生态环境保护、应对气候变化和气候危机，促进资源合理利用和绿色转型等。三是公平正义目标。借助先进技术手段实现收入再分配、更均等配置公共资源，发展普惠型金融和工商业模式，提高社会福利水平，促进社会流动。四是公共安全和公共服务目标。包括提高产业链供应链的安全性，防范和应对公共危机，提高政府服务效率，建设智慧城市基础设施等。

四 引领人工智能创新向善

在谈及许多关键科学技术时，正确地指出其具有“双刃剑”性质固然是有益的，但既然预见到这种双刃剑特征，着眼于从一开始就未雨绸缪，将防范措施植入技术基因之中，无疑更为重要。究竟如何把向善基因植入人工智能，仅仅是通过算法写入那么简单吗？很多人开始为奇点来临，进而人工智能不再受人类控制，反而取代人类成为世界主宰的可能情形忧心忡忡。这的确是一个问题。然而，在思考和准备应对这个问题之前，更应该关注的是在人类尚能支配、控制人工智能的时候，如何以自己的认识、理念、追求和动机引导人工智能技术的创造和应用。作为与人工智能接触最密切的科技人员、投资人、企业家和用户，都会成为上述主观意识的代理，从而推动形成不同的技术创新和应用方向。所以，人工智能创新向善这个理想目标，应该通过推动这些当事人形成正确的共同理念，进而做出合理行为和行动得以实现。

政府应该通过立法和规制，在体现针对人工智能的一般约束，如保护国家安全、个人隐私、知识产权、信息准确性、公序良俗等规范性目标的同时，把就业优先的理念转

变成人工智能当事人之间的共同意图、约束性要求和激励相容行动。在理念上,首先需要设立一个保护劳动者和就业岗位的公共利益底线。由于就业作为民生之本,是人民生活品质、民生福祉和人的全面发展的基础保障,因此,在涉及资本与劳动、生产率与岗位、成本与收益及至效率与公平等权衡取舍时,理所当然应该把就业、民生和人的全面发展作为抉择标准,规范人工智能发展和应用的方向、路径和实务优先序。

在处理涉及这些方面的现实关系,甚至解决相互之间可能存在的冲突时,政府不应仅仅将自身视为一个发挥制衡作用的第三方机构,也不应简单地充当一个中立的裁决者,因此不能采取通常意义上的不偏不倚立场。这是因为在类似的关系中,通常不存在一个先天就存在的利益均衡点。事实上,劳动者和就业岗位天然便是弱势的一方,倾斜性地施加保护是一种符合社会利益的行为。同时,政府采取这种有意识的偏向,可以得到一些政治哲学思想的道义认同和理论支持。例如,美国哲学家约翰·罗尔斯的“社会正义”思想,就被视为现代收入分配和社会福利政策的哲学基础。

罗尔斯关于“无知之幕”这样一种思想实验,设想人们在决定社会基本结构和分配原则时,犹如被遮蔽在幕布后面,不知道且不能自主选择可能具有的身份、地位、天赋、财富、信仰等个人特征。于是,为了在最不走运的情形下,自己的社会生活仍不至于陷入极端困境,人们一般来说愿意接受的社会福祉和社会保护的理念便是,为社会保护水平设置的边界水平,应该以社会中处于最不利地位的那些成员的需要来确定(Stanton, 2007)。诺贝尔经济学奖获得者约瑟夫·斯蒂格里茨(2023)对此做了一个补充,认为在这张无知之幕背后,人们同样不知道自己即将面临的风险。也就是说,政府在这个事务上的职责,正是确保这一原则的实施,从制度上天然地偏向于弱势的一方,即那些可能被替代的就业岗位和可能受伤害的劳动者。

确保这一原则得到充分体现的重要抓手,是构建一个就业友好型或岗位保护导向的法律、规制和产业政策框架。相关的内容应该包括:反歧视与促进公平就业,如推动算法透明化,避免在人工智能招聘工具中出现算法性别歧视和年龄偏见;建立“人类最终决策权”机制,确保涉及人机协作方式、员工雇用、大规模解聘,以及其他重大劳动关系的决策,均通过人为的程序和机制做出,或经由人工的严格复核;保护和拓展劳动者权益,把“平台零工”纳入社会保障体系和劳动力市场制度的覆盖;保障“人机协作”场景下的工作安全与健康标准。此外,把对劳动者的再培训纳入政府和企业的义务、分行业差异化维护效率与就业平衡、监测细分岗位变化、建立动态评估就业影响机制等做法,也应纳入规制范围,确保得到实施。

人机协作是人工智能保留和创造就业岗位的重要途径,劳动者与人工智能代理之

间的分工关系和性质，则是健康、有益、可持续人机协作的关键考量因素。有学者指出，人机分工协作应该重点关注以下关键因素。首先，复杂度越高的任务，自动化的可能性就越小。其次，执行频率越高的任务，越容易被自动化替代。再次，任务衔接的难度越大，特别是在任务可能出现碎片化的情况下，人机协作的难度也就加大。最后，任务产生的结果越是生死攸关，人的作用就越是超过人工智能（Ales & Combemale, 2025）。实施各种监管、规制和产业政策的目的，正是以岗位数量最大化和高质量就业为出发点，改变上述考量中各因素的相对权重，进而转变当事人的成本收益预期，使人类岗位与人工智能的相依相存关系达到合意的平衡。

五 降低技术进步的岗位替代动机

在以前的讨论中，本文初步回答了这部分标题提出的问题，答案是人工智能并不需要以替代就业为发展的必要动机。然而，要想真正做到这一点，显然也不可能将其视为一种自然而然的结果。也就是说，使人工智能的发展激励与替代就业的动因脱钩，应该作为产业政策、社会政策和社会福利体系的共同目标，为此必须构建和依托相应的制度框架。推动必要的制度建设，既要顺应人工智能时代的挑战，与时俱进改变治理观念，也有一般规律可供遵循，以及各国发展的共同经验可供借鉴。作为市场主体的人工智能技术使用者，通常以利润作为生产和经营目标，最大限度降低成本和增加收入。所以，减少对劳动力的雇用进而节约人工成本，是生产经营的天然动机和生存之路。这种市场竞争逻辑不可能改变，迄今为止也没有做出改变的条件。然而，这种微观动机也完全可以在顺应发展阶段变化的条件下，按照一般发展规律的要求，通过制度安排予以重塑，在不损害信号作用和激励机制的同时，使其与社会利益形成一致性。

生活在19世纪的英国经济学家威廉姆·杰文斯发现一个悖论现象，即提高煤炭使用效率的新技术，不仅没有减少对煤炭的需求，反而扩大了煤炭的使用数量，从而加速了资源枯竭的速度（约翰·M·波利梅尼等，2014）。也有人借用这个“杰文斯效应”，预测人工智能的不断进步，导致这一进步本身对算力的需求扩大，从而对能源的过度依赖和惊人耗费。如果从乐观、正面的意义上，把杰文斯效应借鉴来认识人工智能的就业影响，可以预期劳动者的生产率提高可以具有两个效应：一是岗位供求失衡会导致市场工资的降低，二是整体生产率提高可以支撑更多鲍莫尔性质的有需求产业。可见，就业增长的潜在可能性是存在的，归根结底在于各种政策如何做到因势利导。

一般来说，产业政策、宏观经济政策和社会政策，可以通过自身工具手段引导国

民收入分配和再分配，从而改变社会生产和消费模式，以致形成这样一种格局，社会总产出、社会消费需求和投资需求都有越来越大的部分由公共品构成，具体表现形态分别为基础设施、公共服务和相关产品。例如，博物馆、图书馆或者其他文化机构的数量，会随着经济发展水平的提高而显著增加，其本身属于公共设施，为民众提供公共服务或准公共服务，通常又以特定的商品（如图书、文创产品）作为实物载体。一旦在社会总供给和总需求中，公共品或准公共品取得了比私人品更为显著的份额优势，就会相应改变劳动力再生产的成本构成，进而改变劳动要素的市场价格。

我们可以通过跨国数据的比较，观察到这样一个经验事实，并尝试解读其背后蕴藏的一般规律性，即随着人均收入水平的提高，人民生活品质和居民福祉的改善，越来越同政府的公共品供给水平密切相关。或者说在更高的发展阶段上，公共服务成为劳动力再生产的一个必要条件和重要组成部分。反过来说，劳动力市场上的供求关系及其变化，也在越来越不显著的程度上决定劳动力的相对价格或工资。相应地，引导技术变迁的要素相对价格信号也便发生变化。根据世界银行数据库，我们收集了118个国家和地区的数据，观察随着人均GDP的提高，居民消费率和政府支出率（政府支出占GDP比重）之间关系的变化趋势。如图2所示，在较早的发展阶段上，这两个指标之间的差距非常大，整体表现出此消彼长的关系，即在整体贫困的状况下，居民必须花掉收入的绝大部分才能满足温饱。同时，治理能力和财政资源都限制着政府所能提供的公共服务。

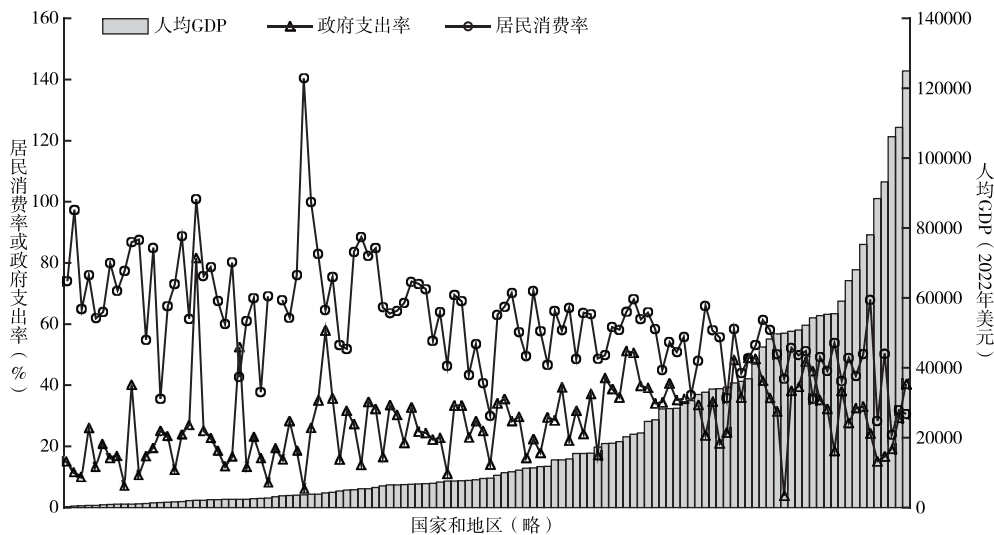


图2 不同发展阶段的政府支出和居民消费

资料来源：根据世界银行数据库（<https://data.worldbank.org/>）计算得到。

这种资源制约状况随着发展水平的提高逐渐改变。我们暂且撇开从中等偏上收入阶段跨入高收入阶段时，需要有一个跨越性的消费提升这种情形，而是着眼于更长期的趋势性变化。从更长的历史跨度看，两个指标的走向之间呈现出收敛的趋势——居民消费率向下，政府支出率向上。这说明，居民消费需求和民生福祉的满足，越来越不那么依赖于把国民收入的极大比重支出在私人产品与服务中，而是越来越需要依靠公共品的供给。

上述变化既是一个具有一般规律性的趋势，也同时带来一个附带的结果，即在劳动力市场上形成的工资水平，以及企业在生产经营过程中面临的劳动力成本，都会在十分显著的程度上趋于降低。或者说，生产经营中使用的劳动力，不再构成市场主体最迫切需要替代的昂贵要素。而在经济发展和技术进步过程中，涉及机器人技术发展和应用的经济动机，则从替代劳动力转向补充和增强劳动者能力，人工智能发展的就业破坏倾向可以得到扭转。反过来，这还会形成一种信号和激励，引导人工智能的发展更加主动地弱化替代劳动力和破坏岗位的功能。

六 应对创造性破坏的破题之策

当人们越来越认识到人工智能将无所不能，并危及各种就业岗位，进而为此忧心忡忡时，实际上陷入了一个悖论式的思维定式。如同用“以子之矛，陷子之盾，何如？”的诘问指出楚人的荒谬一样，对人工智能就业冲击的悖论，也可以提出类似的质问：既然人工智能无所不能，其应用场景无远弗届，何不借助这种威力来解决岗位替代和就业破坏问题？具体来说，应对人工智能可能造成的就业冲击，确实不应该听天由命，更不应该试图阻碍技术本身的发展。唯一不可回避且正确的选择，是主动引导人工智能发展和应用方向，避免或应对人工智能冲击就业的自发效应。

无论是大语言模型“自动补全”的技术特点（即按照指令或用户要求以最大可能性实现任务目标），还是大规模强化学习被用于训练人工智能代理，都倾向于赋予人工智能模型一种内在动力，进而成为一种不断提高的能力，即以一往无前的决心，乃至无所不用其极的手段达到要求的目标（Booth, 2025）。从积极的方面来看，这固然意味着，人工智能系统自身解决问题的能力不断提高，达到目标的愿望也日益增强。问题在于，应该如何赋予其正确的目标，并且使其善用自身完成任务的愿望。下面，本文从四个方面探讨应该采取什么样的手段，才能让人工智能的功能达到人类和社会的意图。

首先,利用人工智能加速技术进步的巨大能量直接创造就业岗位。人工智能优越于以往任何新技术乃至很多核心通用型技术之处,在于其具有加速所有变革过程的能力。无论就其积极意义而言,还是就其消极方面而言,皆是如此。对于创造就业还是破坏就业这个问题来说,长期存在的所谓不对称问题和时间差问题,或者说就业创造在数量上小于、在时间上滞后于就业破坏的问题,可望在这种强大的技术能力下得到解决。作为这方面的一个例子,人工智能技术平台能够产生显著加速产业孵化的效应,因此可以大大缩短初创市场主体诞生的时间,并在更短的时间内释放出就业创造潜力。

社会总供给与总需求之间的对立统一关系,既可以提出新的岗位需求,也可以创造新的岗位供给。新时代中国社会的主要矛盾,已经转化为人民日益增长的美好生活需要和不平衡不充分的发展之间的矛盾。这意味着,尽管需求侧因素越来越成为经济增长的主要制约,城乡居民希望提高生活品质和改善民生福祉的很多需求,尚未在供给侧得到完全满足。此外,人口老龄化、气候变化和绿色转型、发展方式转变、“一带一路”建设,以及向以国内大循环为主体的双循环格局转型等战略的推进,也创造出对于新的消费品、投资品、服务和公共品等的强劲需求,因而也对人工智能赋能产业发展,扩大优质供给提出了有效需求,必然相应创造出新的就业岗位。

其次,借助人工智能技术之力找到或设计出人机互补性,进而推进人机协作,尽可能接近一种人机共进的境界,具有创造就业岗位的无限可能性。虽然关于未来人类智能与人工智能谁将占据统治地位的问题,目前尚无定论,但是,可以设想在相当长的时间内,两种智能各自具有相对优势,仍将和平共处。这提供了一个足够大的空间,可以利用人工智能相关技术,通过重新安排工作任务、重塑生产流程、增强劳动者能力边界,一方面在企业内部推动人机协作模式,另一方面催生人机协作的新业态,稳定乃至增加人类劳动力的岗位。既然人机协作可以通过多种路径实现,也就意味着有多种创造就业的可能性。譬如通过任务的重新组合和配置,让机器人接管那些程式化、重复性的工作,同时帮助劳动者转向决策性、更具创意、需要随机应变性质的岗位;通过增强任务复杂性和延伸产业链,增加人机协作型岗位;在推动技能升级的同时创造互补性岗位;降低各行业的岗位技能门槛,雇用更多就业人员,特别是大龄劳动者,等等。

再次,通过制度安排使生产率提高的成果获得分享,为那些生产率尚未提高却有需求的岗位提供继续存在的理由和条件,或者通过改变就业的定义,创造出以往不被认为是就业的新类型岗位。这相当于把可能发生的“逆库兹涅茨过程”引向“鲍莫尔过程”。也就是说,如果任由人工智能系统自发地替代就业岗位,那些被替代的劳动者就会大规模转移到低生产率部门,在由此造成整体生产率降低的同时,这部分劳动者

的就业质量和工资水平也会下降，全社会的收入差距也会扩大。如果在生产率分享和社会需求扩大的合理范围内，可以做到使新岗位的报酬水平与生产率水平脱钩，则劳动力重新配置的结果，即意味着创造出更多有社会需求的鲍莫尔类型岗位^①。

最后，所有假设的就业创造可以大于、先于或者同步于就业破坏的可能情形，都不会自然而然地产生，或者说不存在所谓的涓流效应，而是需要通过政策调整和制度建设，创造出实现这些目标的必要条件。一个缺一不可的必要条件链条可以表述如下：第一，人工智能技术确实能够转化为提高生产率的手段；第二，生产率的提高能够相应转化为居民收入的提高；第三，人均收入的提高能够转化为新的消费需求。由此可见，技术本身不会给出关于就业影响的全部答案，经济社会体制的因应性变革及其成效，才是影响最终结果的决定因素。

这里，需要特别强调消除可能存在的“人工智能鸿沟”的重要性。任何新技术的发明和使用，都可能制造出针对弱势群体的技术鸿沟，数字鸿沟是迄今为止的一个最新表现。可以预期的是，一方面，人工智能可能产生史无前例的机会，创造出新的工作岗位和就业形态；另一方面，人工智能也会对部分人群形成前所未有的技术阻碍。在就业创造最终惠及全体劳动者之前，就业困难和劳动力市场冲击总是率先降临在特定的脆弱群体身上。因此，利用人工智能创造就业的能力和机会，同时保护好自身权益，防范劳动力市场冲击和风险，必然要求所有劳动者都是平等的人工智能用户：拥有同等的进入机会、便利程度、必备知识和成本承受能力等。

在这个传统意义上的“技术鸿沟”之外，人工智能的发展还创造出一种特殊类型的“鸿沟”，表现在人工智能的创造者和潜在应用者之间的不对称现象。或许是由于人工智能技术的突破性太强、速度过快和步伐太大，导致潜在的技术应用者对技术本身的理解力相对滞后，在如何应用和使用新技术方面知识不足。或者说，关于人工智能技术创新性、想象力和主动精神，均远远不如技术和模型创造本身。与此同时，大型科技公司也好，中小型新创企业也好，由于面临着巨大的竞争压力，每时每刻都在追求可以使自己脱颖而出的技术突破，因而，如何善用技术本身以及如何通过与使用者协同行动，实现科技向善的目标，自然不在它们的优先议程之中。可见，消除这种人工智能在创造与应用之间的知识鸿沟，也应该成为规制和激励机制的取向。

^① 指那些劳动生产率低且没有显著提高的潜力，却具有较大的需求收入弹性的服务。随着收入水平的提高，人们通常愿意支付社会必要水平的价格，以获得这种服务。关于这个概念本身及其经济学含义，可参见 Heilbrun（2011）。关于拓展鲍莫尔类型行业的讨论，可参见威廉·鲍莫尔等（2023）。

参考文献:

- 蔡昉 (2022),《刘易斯转折点——中国经济发展阶段的标识性变化》,《经济研究》第1期,第16-22页。
- 雷·库兹韦尔 (2011),《奇点临近》,董振华、李庆诚译,北京:机械工业出版社。
- 迈克尔·桑德尔 (2012),《公正:该如何做是好?》,朱慧玲译,北京:中信出版社。
- 穆斯塔法·苏莱曼、迈克尔·巴斯卡尔 (2024),《浪潮将至:技术、权力与未来的冲击》,贾海波译,北京:中信出版集团。
- 威廉·鲍莫尔等 (2023),《增长的烦恼:鲍莫尔病及其应对》,贾拥民译,北京:中信出版集团。
- 约翰·M·波利梅尼、真弓浩三、马里奥·詹彼得罗、布莱克·奥尔科特 (2014),《杰文斯悖论:技术进步能解决资源难题吗》,许洁译,上海:上海科学技术出版社。
- 约瑟夫·斯蒂格里茨 (2023),《21世纪的福利国家》,《比较》第1辑,第146-174页。
- Acemoglu, Daron & Pascual Restrepo (2020). The Wrong Kind of AI? Artificial Intelligence and the Future of Labour Demand. *Cambridge Journal of Regions Economy and Society*, 13 (1), 25-35.
- Ales, Laurence & Christophe Combemale (2025). The Jobs AI Can Do – and Those It Shouldn't. Last access date: 2025-02-06, <https://www.ft.com/content/9b34b76c-938b-4de4-838f-9c6ef506da19>.
- Booth, Harry (2025). When AI Thinks It Will Lose, It Sometimes Cheats, Study Finds. Last access date: 2025-03-06, <https://time.com/7259395/ai-chess-cheating-palisade-research/>.
- Brynjolfsson, Erik & Gabriel Unger (2023). The Macroeconomics of Artificial Intelligence. *Finance & Development*, 60 (4), 20-25.
- Hayami, Yujiro & Vernon Ruttan (1985). *Agricultural Development: An International Perspective*. Baltimore: Johns Hopkins University Press.
- Heilbrun, James (2011). Baumol's Cost Disease. In Ruth Towse (ed.), *A Handbook of Cultural Economics* (2nd). Cheltenham: Edward Elgar, pp. 67-75.
- International Labour Organization (ILO) (2024). *Global Wage Report 2024-25: Is Wage Inequality Decreasing Globally?* Geneva: International Labour Office.

- Müller, Christopher (2024). *World Robotics 2024 – Industrial Robots*. Frankfurt am Main, Germany: VDMA Services GmbH.
- Stanton, Elizabeth (2007). The Human Development Index: A History. *PERI Working Paper Series*, No. 127.
- Wiener, Norbert (1960). Some Moral and Technical Consequences of Automation. *Science*, 131 (3410), 1355–1358.

Harnessing Artificial Intelligence to Create More and Better Jobs

Cai Fang

(Chinese Academy of Social Sciences)

Abstract: The development of artificial intelligence (AI), akin to other transformative technologies, exhibits a dual nature. Regarding its employment impact, AI possesses the capacity for both job creation, driven by productivity gains, and job destruction, through the substitution of human labor with AI or robotics. Moreover, job creation may be quantitatively insufficient to offset job destruction and may also temporally lag behind it. The ultimate balance between these countervailing effects hinges on the strategic orientation of AI's development and application. This paper argues for aligning AI development with an employment-first strategy and philosophy, exploring the theoretical underpinnings and practical approaches to reorient the motivations driving AI's technological trajectory. First, it examines how legal and policy frameworks can steer AI innovators towards integrating an employment-first philosophy throughout the technological development lifecycle. Second, by contending that labor input reduction is not the exclusive means to enhance productivity, nor is productivity enhancement the sole objective of technological innovation, the paper advocates for directing AI development away from a primary focus on job displacement. Finally, the paper situates AI's employment impact within Schumpeter's framework of creative destruction and proffers policy recommendations substantiated by both theoretical analysis and empirical evidence.

Keywords: artificial intelligence, employment-first strategy, alignment, motivation of technological change

JEL Classification: O33, O38, J23, J24

(责任编辑：西 贝)