武漢大學學報(哲学社会科学版)
WUHAN UNIVERSITY JOURNAL (Philosophy & Social Science)
第77卷第6期 2024年11月
Vol. 77 No. 6 Nov. 2024 025~033
DOI:10.14086/j.cnki.wujss.2024.06.003

生成式人工智能赋能哲学社会科学研究

马费成 陈帅朴

摘 要 生成式人工智能作为基于海量数据训练的模型,涵盖了丰富的社会价值和事实知识,其所展现的机器智能显著提升了各类社会生产与科学研究的效能。哲学社会科学强调与新一轮科技革命和产业变革交叉融合发展,生成式人工智能除了作为前沿技术工具丰富了哲学社会科学的研究手段,其更为突出的特征是成为哲学社会科学的研究对象,极大扩展了研究视角和观察视域。鉴于二者融合具有释放巨大势能的潜力,首先应当剖析生成式人工智能赋能哲学社会科学的前提基础,在此基础上围绕赋能过程中的创新模式与核心挑战进行深入探讨,最后系统提出赋能路径的建设策略,从而厘清生成式人工智能对于哲学社会科学的赋能机理,创新哲学社会科学的理论体系和研究方法,推进中国特色哲学社会科学的建设与发展。

关键词 生成式人工智能;哲学社会科学理论体系;哲学社会科学研究方法
中图分类号 C0;TP18 **文献标识码** A **文章编号** 1672-7320(2024)06-0025-09
基金项目 国家社会科学基金项目(19VXK09)

从数字化到数智化,技术演进已成为推动社会发展的关键要素。万物互联的数字文明新时代,人工智能作为人类探索机器智能的前沿技术,受到更为普遍和深入的关注,为各个领域带来了革命性的变化。为贯彻落实国家《新一代人工智能发展规划》,中华人民共和国科学技术部会同国家自然科学基金委员会于2023年3月启动人工智能驱动的科学研究(AI for Science)专项部署工作。如今,生成式人工智能(Generative Artificial Intelligence, GAI)的突破进一步推动科学研究边界拓展,权威科学期刊《自然》也将GPT-5等新一代AI的挑战列为未来最值得关注的科学事件之一。

生成式人工智能,亦称为大语言模型(Large Language Models, LLM),起源于1956年达特茅斯会议诞生的人工智能概念,近十年受生成对抗网络、强化学习、预训练模型等核心技术的驱动而快速发展。在其建构与发展过程中,理论研究经历了从规则逻辑到机器学习再到深度学习的演进,运作模式也从识别和预测现有数据转到创造和生成新型数据,应用实践更从空中楼阁到落地生根,为各行各业带来了显著的价值跃升。相较于过往各阶段的人工智能,生成式人工智能的典型特征是运作逻辑的转变,即从分析式系统转向为生成式系统,极大提升了技术的感知易用性和感知有用性,使人类能够更便捷高效地应用先进技术。此外,生成式人工智能得益于数智环境下海量的互联网多模态数据以及先进的计算机硬件支持,通过学习和理解数据的分布,已经具备强大的生成能力、迁移能力和交互能力,展示出类人的感知与认知智能。当前,以ChatGPT为代表的生成式人工智能正加速成为人工智能领域的前沿发展方向,推动人工智能从庙宇高堂走进千户万家,对经济、社会以及科学研究的发展产生重大影响。

生成式人工智能承载着人类对于机器智能的丰富想象,与哲学社会科学密切相关。生成式人工智作为人类思考存在与认知等哲学问题的现代化表达,能够模仿人类思维的复杂过程,重塑人类获取、处理和创造信息的方式。生成式人工智能作为数智环境中的变革性技术,正在为哲学社会科学提供新的

研究视角和工具。在新文科建设背景下,哲学社会科学正与时俱进,强调与新一轮科技革命和产业变革交叉融合发展[1]。探索生成式人工智能在哲学社会科学中的赋能作用,是学科持续发展的关键问题。交叉融合作为新文科建设的重要指导思想,赋予了现代社会科学研究对话科技与人文的重任。如何在技术变革的背景下抓住机遇? 这不仅是学科之问,更是时代之问。为科学地回答这一问题,需把握数智环境引发的研究思路转变,也需剖析与技术原理相契合的科学研究变化,才能系统地把握生成式人工智能对于哲学社会科学的赋能机理。

一、生成式人工智能赋能哲学社会科学研究的前提基础

哲学社会科学作为源自实证科学的学科,通过借鉴自然科学的思维逻辑和研究方法,研究人类社会的发展规律、结构制度等方面[2]。然而,与自然科学以严格的控制实验和系统思想为主导不同,哲学社会科学体现出更加复杂与多样的研究特性。总体而言,哲学社会科学不如自然科学拥有相对稳定的整体性规律,而在研究中呈现如下典型特征:一是研究方法论重机制而非法则,即哲学社会科学的理论和模型往往只能在特定的、有限的情境下成立;二是研究对象具备高度的复杂性,哲学社会科学研究的对象是人类社会及其行为,这些对象存在结构与功能之间的松散型连接;三是研究过程难度大且意义有限,自然科学通过严格的控制实验来设计研究过程,而社会科学实验往往难以在完全可控的环境中进行,即控制影响社会现象的复杂因素几乎是不可能的,一定程度上导致研究结论只见树木不见森林。这些特性决定了哲学社会科学研究存在显著的立场和价值观念问题。现代化进程赋予人更加多元的价值观念,这无疑为社会科学研究增添了诸多复杂性。面对纷繁复杂的社会现象,传统的归纳演绎以及基于小样本数理统计的方法,往往难以有效地拟合和重构。

现代社会正经历从信息时代向数智时代的演进,这为更好地解决哲学社会科学难题提供了新契机。信息时代以数据结构化为目标,关注数据处理与信息管理问题,将信息化建设视为主要任务,强调将数据和信息保存在信息系统中,实现各项业务信息的自动化管理。数智时代则致力于解决信息时代遗留的信息孤岛问题,其核心目标转向数据智能化,更加注重知识挖掘与智能决策,强调运用新型技术,集合数字资产积累和智能化分析手段,推动科技与研究的全面创新发展[3]。在这一过程中,技术演进始终是发展的关键驱动力。自工业革命以来,以自然科学为基础,自动化技术被用于解放人力资源。而今的数智革命,以大数据、物联网、云计算、人工智能驱动的数智技术,正表现出可量化、实时化、可迭代、可视化和智能化的核心特征,不仅加深了对自然现象的洞察,更重塑着人类社会的组织与运作模式[4]。在新兴的数智环境下,大数据驱动的研究范式和正在兴起的人工智能驱动的科学智能研究范式(AI for Science,AI4S)日趋成为主流[5]。前者通过对海量数据的深度挖掘,揭示潜在的关联规律及因果关系,捕捉事物的本质联系;后者则着重解决以往范式难以解决的科学组合的维度灾难问题,赋能决策者全面洞悉科学前沿动态,推动知识与技术的不断迭代更新,这两种范式可以挖掘以往社会科学研究中潜在且难以言喻的机制与模式。此外,数智时代相比信息时代更加强调类人的智能化,而不仅仅是机器的自动化,更需要前沿技术与哲学社会科学携手共进。

生成式人工智能,作为大数据和人工智能驱动最为突出的技术,正在密切加深与哲学社会科学的融合,使得哲学社会科学更加重视计算思维、协同思维、跨学科思维和关联思维,为解决复杂问题和构建新的研究体系提供了有力保障。从其迭代发展过程来看,生成式人工智能经历了从基于规则的逻辑推理到基于神经网络的感知和认知的演变。生成式语言模型作为目前最为高级的人工智能,秉持规模定律的显著特征,展现出卓越的理解、生成等能力。现代社会科学逐渐深化对定量方法挖掘社会深层规律的运用,而生成式人工智能的运作逻辑是数理统计,通过对人类现存知识和经验进行大规模训练,以高相关性的关联概率模拟人类社会的各种复杂关联。尽管这种方式在因果推断方面本质上是一种统计关系的推断[6],并不总能指向真正的因果逻辑,但在哲学社会科学领域的应用前景仍然令人期待。

首先，生成式人工智能为哲学社会科学研究提供了强大的新工具新手段，强化了数据驱动的实证研究范式。其一，生成式人工智能极大地丰富了数据的收集和处理方式，将受人类特性显著影响的主要数据来源扩展为具有多模态、深度语言建模以及更高稳定性和相关性的模型生成与理解的数据，提供了相对严格的类似自然科学控制实验的数据。其二，生成式人工智能极大地丰富了社会科学知识的来源广度。传统意义上社会科学研究依赖于学者的先验知识与价值视角，而生成式人工智能可以汇聚海量的世界知识库，包含不同场域的信息与知识，极大地扩展了社会科学的研究视角和观察视域。其三，生成式人工智能极大地促进了社会科学的研究分析过程。与传统的抽样调查方法相比，生成式人工智能通过对大规模数据的训练过程，汲取大量的人类经验和观点，提升研究结论的可推广性。生成式人工智能的应用更接近完全理性假设，使得社会科学研究得到的机制更具可靠性和普适性，有利于实施更为严谨的控制实验并扩大其研究结果的影响意义。

其次，生成式人工智能带来了哲学社会科学研究基础理论的新问题新方向。生成式人工智能与其他前沿技术的显著不同之处在于其本身成了重要的社会科学研究对象。生成式人工智能在各种能力测评中表现出与人类相似的价值取向和心理特性，这推动了“人工智能心理学”“机器行为学”“机器伦理学”等新兴前沿领域的发展[7]，这些研究旨在通过将人工智能作为社会科学研究的对象来识别其能力、行为、决策机制以及伦理道德。此外，生成式人工智能的广泛应用为各类研究提供了新的方法和视角，极大地丰富了哲学社会科学中诸多学科的边界拓展，并积极推动了理论与实践的发展。

总之，生成式人工智能有助于哲学社会科学开辟新范式，绽放新活力。一方面，它为哲学社会科学开辟了新的研究方向，拓展了社会科学的应用场景，助推了从小数据辅助到大数据发现的范式进化。另一方面，由生成式人工智能引发的价值对齐风险、可靠可信服务等前沿问题亟待进一步探讨。赋能过程必须兼顾自然科学的规律与哲学社会科学的伦理准则，追求技术与人文的和谐共存，这愈加凸显了哲学社会科学理论和方法的重要性，不仅助力于科技的伦理发展，也推动了学科界限的延伸与整合。

二、生成式人工智能赋能哲学社会科学研究的创新模式

从信息化迈向数智化，人工智能正持续发力，不断促进信息技术与哲学社会科学交汇融合。在此基础上，科学认识生成式人工智能与哲学社会科学结合的典型特征和创新模式成为核心思考问题。

受生成式人工智能影响，哲学社会科学呈现出学术研究的两大典型特征。第一个典型特征是问题域的不断拓展。这主要指对原有问题的新突破以及未知领域的新探索，帮助研究者发现新的科学问题，通过对这些新问题的研究，揭示新的人文社会现象和规律。这一过程，或是发掘性问题，即哲学社会科学领域长期存在但未能解决的问题，如今借助新的数据和工具得以研究；或是原发性问题，随技术发展而呈现的新问题，如生成式人工智能的价值对齐、元宇宙的人机交互等问题。第二个典型特征是研究者参与度的不断提升。由于研究过程的智能化，研究者既可以远离研究过程本身，提高研究的客观性和效率，也能够以被试的角色融入实验过程，亲自体验和判断研究的特征、细节和真伪。在此过程中，研究者有着研究广度和深度的提升。生成式人工智能能够迅速生成和模拟高质量数据，为研究者提供丰富的数据源，研究人员能够跨越传统的学科界限，融合不同领域的知识，进行更为深刻的研究。研究者还有着自我体验的提升，生成式人工智能可以创建虚拟环境和模拟情景，使研究者能够以参与者的身份体验研究情境。这种沉浸式的体验可以增加研究的互动性，使研究者能够更好地理解研究对象的行为和决策过程。

两大典型特征表明生成式人工智能正在为哲学社会科学的研究模式注入新的活力，激发创新模式围绕三个方面展开：基础数据利用、人智协同、研究取向融通。

首先，数据利用的创新涉及数据来源的多样化以及数据处理过程的进化。具体而言，数据来源多样化体现在量级和类型的丰富，传统社会科学研究通常依赖于调查问卷、实验室研究或案例研究收集的小

规模数据集,这种方式已无法满足生成式人工智能背景下的科研需求。如今生成式人工智能的训练语料有着全量样本特点,能够通过海量数据集来进行大规模的模式识别和趋势预测。数据类型的扩展体现在不同模态、不同结构数据的融合,社会科学研究的数据来源已经从传统的文本、问卷数据扩展到图像、视频、音频和实时交互数据[2]。这些数据为研究人类行为和社会环境的交互提供了全新视角,使得研究可以更全面地捕捉和理解社会现象。此外,数据处理过程朝向全局关联和持续细化发展。在生成式人工智能产生以前,数据处理往往局限于部分领域、部分场景的关联分析,但社会作为复杂信息系统,数据关联往往呈现出非线性耦合和复杂关联的特性[8]。生成式人工智能凭借其全局视角,将研究领域向外系统延伸,形成跨领域、跨场景的全局关联的数据网络。这使得生成式人工智能能够持续细化社会科学研究中的数据处理流程。在数据收集阶段,生成式人工智能能够更加高效地筛选和整合多源数据,确保数据的全面性和多样性;在数据清理阶段,生成式人工智能展现出强大的异常值检测、变量转换和缺失值处理能力,能够有效去除噪音数据,保留含有丰富信息且具备研究价值的数据;在数据注释阶段,生成式人工智能不仅能够实现与人类能力相似的高质量标签标注,甚至在一致性上表现得更为出色,确保数据的准确性和一致性,为后续分析奠定坚实基础。在高效处理已有数据的基础上,生成式人工智能在数据生成与扩展能力上展现出卓越的创新性。通过模拟复杂的多维数据关系,生成式人工智能能够生成具有现实参考意义的合成数据,进一步丰富数据集的多样性和覆盖范围,填补现有数据集中的空白,为稀缺数据场景提供高质量的替代数据,全面提升数据的多样性、平衡性和适用性,不仅弥补了传统数据处理方法的局限性,而且为研究复杂问题提供了全新的数据资源和研究可能性。

其次,当哲学社会科学研究注入生成式人工智能之后,人智交互作为创新的协同模式,愈发重视生成式人工智能在社会科学研究中的重要地位。传统研究模式中,研究主体完全依赖人工实施,主要关注解释型任务,通过实验和调查数据,利用成熟的理论模型和统计方法来阐释研究对象的部分特性。然而,随着各类交互设备、计算机技术以及以机器学习和深度学习为核心的人工智能技术的发展,社会科学的研究视角也转向了更加复杂的预测、启发等任务。在这一过程中,研究主体由人类和各类计算机与信息系统所组成,形成了人机交互模式,此时机器主要依赖于预设规则和有限的数据处理能力,更多承担着信息检索、数据整理和初步分析等辅助性任务,且机器的智能化程度有限,缺乏自主学习和深度理解能力,难以应对复杂的社会现象和动态变化的研究需求。这一阶段,人类仍旧在方案制定和决策分析等研究流程中占据主导地位。近来,生成式人工智能的快速突破使得研究者看到了机器具备创造性解决问题的能力,前沿研究主体正进化为人智交互的协同模式。在这一模式中,研究者作为需求提出者和反馈提供者,为研究提供了丰富的语境、价值观和行为特征输入。生成式人工智能不再停留于传统人机交互模式中辅助者的角色,而是承担了包括方案制定者在内的多种角色,凭借其全域知识和预测能力,帮助研究者突破传统思维局限,为研究者制定更加准确和高效的解决方案。人智协同正在整合包含交互环境、交互任务在内的众多要素,以构建“正和”博弈的动态反馈机制为前沿方向,促进智慧增量,支持更高质量、效率、智能的社会科学研究。

最后,在生成式人工智能推动背景下,研究更应注重数智环境中不同研究取向和方式的结合。一是重视模型驱动与数据驱动相结合。模型驱动侧重于从理论构建模型和演绎推理,而数据驱动则依赖大量数据进行归纳推理和模型构建。两者结合可以挖掘出传统模式无法预见的新现象和趋势。生成式人工智能通过增强数据处理、自动化实验设计、跨领域知识整合、复杂系统分析等方法,不仅优化模型驱动和数据驱动的研究质量,还加速二者的融合,为社会科学研究面临的复杂问题提供有力工具。二是重视相关分析与因果分析相结合。社会科学研究的核心是揭示事物之间的因果关系来探究基本规律。然而,大数据环境下的数据复杂性和非线性关系使得因果分析变得困难,相关分析可以快速识别数据间的潜在关联,为初步洞察提供依据。为实现相关和因果的融合,可以发挥生成式人工智能的作用,例如,设计和模拟随机对照实验设计、增强因果发现算法等,通过结合大数据分析的广度和因果推理的深度,揭

示更精确的社会模式和因果关系。三是重视空间分布与时间序列分析相结合。空间分布不仅关注物理的地理空间，也涵盖权力空间、关系空间、心理空间、信息空间等多维的社会空间及其相互作用，而时间序列分析揭示了社会现象随时间的变化。二者结合能深化对社会动态的理解。生成式人工智能具有多模态数据的表示和处理能力，提供了深入洞察社会科学的时空研究的可能性。例如，能够实现时空数据整合和清洗，能够实现时空数据的复杂依赖关系建模以及预测和模拟等，推动对社会科学现象更深刻且全面的理解。四是注重实证与思辨相结合。实证研究强调用经验材料证明或证伪理论假说，关注对客观事实的说明；而思辨研究则注重概念分析、逻辑推理和理论构建，也关注价值理想。两者存在鲜明的互补性质，思辨帮助形成理论假设，实证则验证这些假设，共同推动科学知识的积累和发展。生成式人工智能通过处理和分析大规模数据，促进实证研究得到更为充分的经验证据，挖掘思辨研究中隐藏的潜在模式和抽象概念，实现二者在经验广度和理论深度的统一。此外，生成式人工智能也促进了两者研究中事实与价值的协同，使得研究者能够更全面地理解和解释复杂的社会现象，开发出更符合伦理和社会需求的人工智能技术。这不仅为生成式人工智能的健康发展提供了重要指导，也推动了哲学社会科学领域的创新和进步。

三、生成式人工智能赋能哲学社会科学研究的核心挑战

生成式人工智能赋能哲学社会科学的创新模式，不仅提高了传统社会科学研究中各个环节的效率，也扩展了社会科学研究的边界。然而，实现生成式人工智能在社会科学领域的深度赋能绝非一项平地起高楼的简单任务。一系列研究证据表明，生成式人工智能受训练过程和训练语料影响，已经显现出歧视偏见、伦理争议以及模型弱解释性等诸多问题[9]。为了把握和应对生成式人工智能技术赋能哲学社会科学研究所带来的巨大机遇与挑战，需要对核心关切的问题与挑战进行系统分析和深入思考。

一是需要关注生成式人工智能生成内容的可靠性问题。生成式人工智能是联合概率模型，仍不可避免地会出现生成内容的真实性和可信度问题，包括最为典型的幻觉问题，其定义为生成文本内容与可查证的现实世界事实相矛盾，或者与用户的指令和给定的上下文不吻合的情况。这通常是由于训练数据的偏差、模型训练的过拟合以及推理过程的不透明等因素所造成[10]。幻觉问题的出现不仅会影响用户的交互体验，例如模型答复生成过程存在胡编乱造甚至前后矛盾的情况，也有可能导致不良的社会影响，例如虚假信息的广泛传播导致互联网空间数据环境的进一步恶化。为改善这种情况，未来的社会科学研究需要引入事实核查机制、检索增强机制并完善模型架构。需要提升用户的媒介素养和批判性思维，加强模型输出的监督管理，以降低虚假信息的传播风险，确保生成式人工智能技术的健康发展，发挥其在促进社会进步中的积极作用。

二是需要关注生成式人工智能存在的偏见与伦理风险。生成式人工智能在社会各界的应用日益广泛，然而伴随而来的歧视偏见、伦理道德等方面的挑战亦不容忽视。首先在歧视偏见层面，互联网数据的质量参差不齐，尽管生成式人工智能的训练经过了科学的数据清洗与过滤，但数据偏见问题难以避免，会使得生成式人工智能继承并放大现有的意识形态、性别、种族或年龄偏见，会产生包括侮辱、仇恨、威胁在内的各类有毒语言。其次在伦理道德层面，涉及敏感场景的不当使用问题日益凸显，ChatGPT诞生初期，有人用逆向提示工程的方式诱导其撰写犯罪说明书。涉及法律监管的领域更是如此，以知识产权领域为例，模型训练过程的训练数据和推理过程的生成内容都可能存在侵权风险。此外，心理健康和医疗诊断领域作为伦理道德和法律敏感的高风险领域，更需审慎使用生成式人工智能技术。为正本清源、科技向善，需要以价值对齐作为模型建设观念，依赖哲学社会科学价值理论的支撑，建立更全面的价值观指导原则和法规政策，从数据建设、模型对齐、评测标准等层面引领人工智能向着更加负责任更透明且与人类利益相符的方向迈进。

三是需要关注生成式人工智能引发的科研质量问题。生成内容的可靠性与偏见风险着重揭示了生

成式人工智能通过社会应用的较高复杂性间接影响哲学社会科学的研究。进一步地,生成式人工智能也能通过更加直接的方式影响哲学社会科学的研究质量。首先是生成式人工智能助力的科研工作者的“知识革命”,显著提升了科研效率。然而,生成式人工智能往往会臆造知识、臆造结论、臆造引用来源,甚至臆造论文。这极易导致研究质量和准确性的降低,形成“假作真时真亦假”的不良学术环境。其次,模型开源与闭源问题对科学研究繁荣产生了显著影响,这本质上也是基础设施建设必须考虑的关键因素。闭源的生成式人工智能往往秉承商业化应用的发展理念,得益于企业高额投入,其性能通常更佳。然而,源代码的非公开性使得研究难以理解模型的内部机制和决策过程,不仅降低了用户的信任程度,也难以进行个性化的模型改进与优化。更需要注意的是,闭源模型往往迭代速度较快,模型版本不同可能加剧模型结果的不一致性,一定程度上降低研究的可复现性。相对而言,开源模型虽一定程度上效果落后于闭源商业模型,但其透明度能够增加社会科学研究的稳定性和可复现性。社会科学是众多学科领域交织的学科,通过对开源基础模型的持续优化,能够训练更多垂直领域模型,解决领域特定问题。因此,未来加强基础开源模型的建设不仅有助于生成式人工智能的普适化应用以及平权发展,也能够推动知识的持续创新与共享。

四是需要关注生成式人工智能在科学想象方面的能力有限性。这种想象力的有限性主要体现在概念抽象不足和创造力缺失。首先,生成式人工智能的概念抽象能力不足,具体表现为总在基础的计数问题、算数推理等注重概念抽象的数学问题中表现不佳[11]。推测生成式人工智能内部不存在通常意义的计算和因果逻辑,主要依赖上下文语境和语料相关度作出判断。与之相对,人类能够对现实世界进行抽象,构建系统的概念框架,进行深入的逻辑推理。与自然科学相比,哲学社会科学所关注的社会运动规律具有更深的内蕴性、趋向性和复杂性,仅依赖生成式人工智能的相关度预测无法满足科学想象的需求。其次,创造力的缺失是制约科学想象的另一个关键挑战。真正的创造力不仅是信息或知识的重组,而是超越现有知识边界,探索未知领域,提出突破性观点和假设。尽管生成式人工智能在概率统计模型的帮助下能够产生一些新奇内容,但难以自行超脱现实条件去探索和假设客观世界的未知规律。人类在某些情况下会产生偏离标准推理方式或正常的逻辑规则的现象,但也增强了人类创造力、想象力和情感等能力,使其能够更好地适应和应对复杂多变的科学现象,而机器往往难以突破算法逻辑的理性约束,对于生产与建构具有意义性质的知识作用有限。由此可见,哲学社会科学研究过程仍要保持研究者作为主体的核心地位,防止过度依赖机器智能而丧失研究者的主体性与独立思考能力。

四、生成式人工智能赋能哲学社会科学研究的实践路径

生成式人工智能在赋能哲学社会科学的过程中,机遇与挑战并存。因此,探索二者有机融合的研究路径对于赋能过程至关重要。在研究路径的建设过程中,首先需要确立指导原则。创新模式和核心挑战的研究趋势均显示,增强生成式人工智能的可解释性以及人类信任对赋能起着至关重要的作用,可解释人工智能(Explainable Artificial Intelligence, XAI)是赋能路径的重要目标,应将其实现作为指导原则。具体的着力路径需要结合哲学社会科学研究特色展开。在哲学社会科学研究中,观察—归纳—证实作为一套普遍的科学发现实证机制的研究路径,生成式人工智能正利用其独特的能力与优势,推动这一研究路径的创新发展。在此基础上,激发生成式人工智能的感知能力是创新基础,拓展认知能力是必要条件,完善分析体系是关键核心。

(一)以可解释人工智能实现为指导原则

指导原则的建设是实现深度赋能路径的内在逻辑。可解释人工智能旨在使人类用户可以理解并信任人工智能创建的结果和输出[12],这既是创新模式中人智协同的重要目标,也是应对生成式人工智能核心挑战下的重要原则。生成式人工智能是端到端的复杂智能决策系统,其整体复杂性加剧了不可解释性,使得人类难以了解决策原因和过程。这既不利于科学研究的优化与改进,也无法保障社会应用的安

全。在此背景下，亟须为其建立行之有效的实现框架。设计科学旨在通过创造与评估等一系列规范流程来设计创新成果以扩展人类和组织能力的边界，可解释人工智能作为复杂信息系统，设计科学能够为可解释人工智能的开发提供系统性指导，具体通过研究理论、研究方法、研究内容以及研究评估展开。在研究理论方面，应当扩展至多学科交融的视角，不仅需要计算机学科提供算法与模型的持续优化，更需以哲学社会科学作为价值指导。例如，哲学可以提供思辨讨论与价值规范，心理学可以迁移行为主义来对机器进行测试与评估，法学可以探索技术合法性，信息资源管理学科可以整合多方知识资源进行系统规划。在研究方法方面，需注重解释方法的改进和优化，既要追求预测过程的可解释性（模型自解释），也要追求预测结果的可解释性（建模后解释）。生成式人工智能的底层神经网络架构面临“不可能三角”，即同时实现并行训练、低成本推理和良好的扩展性能。这导致可解释人工智能也面临着解释性与准确性之间的平衡难题。因此，研究方法的改进需要引入数学、计算机乃至物理等学科进行底层逻辑建模，以实现高质量的解释方法优化。在研究内容层面，人本人工智能的概念正在兴起，人类作为人本人工智能和人智交互过程的关键要素，可解释人工智能的开发需着眼于用户需求，重视人智交互过程中涉及用户层面对于可解释性因素的理解，并构建一个能够接受人类反馈和拥有迭代能力的可解释人工智能。在研究评估方面，可解释人工智能的实用性、质量和有效性必须经过设计良好的评估方法的严格论证。在这一过程中，评估目标与评估指标的构建尤为重要。针对评估目标建设，目前多数研究仍基于模型视角进行效果测评，但人类作为关键主体，环境作为关键要素，测评需要关注不同环境下人类的体验问题，因此，需要将这二者纳入评估目标之中。针对评估指标建设，目前研究大多聚焦于单一指标，而可解释性是多元评价体系的综合。因此，未来需要构建覆盖多层次、多角度的评估体系，以实现全面有效的可解释性评估。

（二）以激发感知能力为创新基础

指导原则为着力路径提供了目标导向。在着力路径建设中，需要激发生成式人工智能的感知能力，以实现对社会科学中的研究现象和研究数据等客观事实的增强观察。感知能力指的是对客观存在的事物、现象进行理解的能力，涵盖对物理空间和信息空间的多维感知。在物理空间感知方面，生成式人工智能及演进的人工智能智能体（Artificial Intelligence Agent）能够显著增强感知能力，主要体现在两个方面：一是时空维度的扩展，其能够突破时空束缚，提供全景式体验，帮助人类更全面地理解不同的社会文化现象。二是感知视角的扩展，其超越了人类的生理感知限制，提供更加立体、全面和持久的感知，避免单一视角带来的感知偏差，深入洞察隐性和微观层面的现象本质。在信息空间感知方面，生成式人工智能同样展现了强大的能力，一者，其能够整合和理解多源异构数据，打破领域和数据模态的鸿沟，在单一模态数据上具备更加细粒度的理解能力。例如，在处理文献类型的文本数据时，其能够深入分析章节、表格、公式等不同粒度层面。二者，其具有强大的数据生成能力，随着生成式人工智能对高质量训练数据需求的与日俱增，互联网中的优质资源将逐渐枯竭，而生成式人工智能通过生成合成数据来弥补这一缺口的能力，将成为推动科学研究的关键力量。除此之外，生成式人工智能虽然不曾经历社会化进程，但其感知能力已经显现出模拟人类群体行为的潜力。这为丰富基于代理的建模（Agent-Based Modeling，ABM）范式提供了基础，生成式人工智能可以凭借这种能力研究复杂社会系统中群体之间的互动和行为，不仅提升了社会科学研究中真实模拟群体过程的可行性，还增强了对社会系统的复杂性、非线性和演化过程的理解，能帮助分析和预测社会系统中的模式和趋势。

（三）以拓展认知能力为必要条件

生成式人工智能的认知能力需要拓展，以促进信息和知识的归纳综合。认知能力旨在赋予机器模式建构、逻辑推理的功能，模仿人类思维对数据加工处理，转化为信息和知识。从早期人工构建的特征工程辅助机器理解知识到当前基于深度神经网络的参数自动学习知识，人工智能的认知能力边界被进一步扩展。为使生成式人工智能推动社会科学研究的归纳综合能力实现纵深发展，需要从整合已有知

识体系以及创新知识体系两方面尝试。针对整合和利用现有知识体系,首先是异构知识库构建,生成式人工智能能够从大量异构信息源中自动抽取知识,通过创新异构知识表证与组织方式,有助于捕捉异构知识的多样性和复杂性,构建起多模态、多维度的统一知识库,为知识的高效整合和检索奠定基础。其次是关联性知识挖掘,基于生成式人工智能的强大的能力,可发现隐性的知识关联模式,揭示不同领域、不同理论之间的内在逻辑联系,促进知识的融合贯通。针对创新知识体系,首先是跨领域知识迁移,生成式人工智能擅长捕获不同领域知识间的关联,能够实现知识的跨领域高度共享和迁移,有望推动不同学科之间的知识互鉴,将其他领域理论和方法引入社会科学视角,催生跨界、原创的知识体系,开辟新的研究思路和范式。其次是创新问题发现与理论建构,一方面,生成式人工智能正推动认知科学、人机交互、人工智能等与社会科学紧密相关的领域发生理论变革,促使社会研究重新审视新技术环境带来的全新研究问题。另一方面,生成式人工智能可以在现有知识基础上,在假设空间实现高效的组合探索,减少试错成本,提出更有前瞻性的创新研究问题,面向特定研究问题提供实验方案,加速理论框架的创新构建和优化。

(四)以完善决策分析体系为关键核心

在事实感知和知识认知的基础上,赋能路径中关键核心在于完善融合生成式人工智能的决策分析体系,以更好地支撑科研发现的验证与推演。决策分析体系,是指吸纳科学证据,在此基础上构建一套能够弥补知识缺口、处理不确定性的理论架构。生成式人工智能的兴起正基于这一思想,利用统计概率得到的强相关性来获取相对精确的近似解,而非追求最优解。在人工智能驱动的科学研究浪潮中,为真正发挥生成式人工智能辅助哲学社会科学研究的分析能力,需构建一套决策分析支撑体系。首先是新型决策逻辑的确立,机器智能和人类智能的融合发展催生了人机混合智能,衍生出一种新型的决策分析逻辑,将人的抽象思维与机器的规模计算相结合,形成人机深层协同交互的合作模式。在这一模式下,新型决策分析逻辑以人类价值为导向,机器模型会不断学习和理解人类思维,动态调整自身的逻辑推理和分析过程。其次是将新型分析逻辑具体应用于研究发现的验证与推演过程。根据研究难易及证据充足程度,采取实证式分析和启发式分析两种应用方法。针对科学证据较为充足、不确定性程度较低的科学研究,可采取人类为主导、机器辅助解释和验证的实证式分析方法。这种方法可以发挥人类的创造力和概念抽象能力,结合机器的形式化推导和数理验证能力,以“白盒模型”的分析方式实现对简单科学问题的机制理解和高效验证。针对科学证据较为薄弱、存在高不确定性的复杂科学研究,可以采取启发式分析方法。这一过程需更加重视人机混合智能的创新驱动作用。其有区别于传统的线性流程,是以复杂系统理论为指导,综合考虑人类、机器及各种环境的作用影响。适度承认“黑盒模型”的合理性,充分发挥生成式人工智能在大规模计算、跨领域知识关联、复杂逻辑处理等方面的优势,与人类的洞察力和创造力相结合,获得更为准确、客观、科学的分析结果,助力科学发现的验证和推演。

在数智环境中,生成式人工智能为哲学社会科学研究带来了前所未有的发展机遇。基于对大数据和人工智能驱动下的研究特点与规律的深入把握,推进人类智能与机器智能的深度融合,已成为实现科研创新迭代的关键。然而,我们不能忽视生成式人工智能对社会价值观和认知体系可能带来的冲击。生成式人工智能带来的智能化革命是否能真正突破塞尔“中文屋”实验中的计算程序逻辑,达到图灵测试中类人的智能水平,仍需要人们以更加审慎的态度加以考量。在此背景下,践行“以人为本”,始终将“人”作为核心研究理念、对象和目的,将有助于突破技术与人文的壁垒而引导人工智能有序发展。在这一过程中,哲学社会科学各学科的研究特点决定了其形式、节奏和路径的多样性,但随着现实问题的日益复杂,许多研究和应用难以简单归类于单一学科。未来,跨学科合作将愈加关键,以跨学科和系统化思维在不同研究领域高效整合人工智能技术,有助于促进学科交流合作,实现领域协同共进,持续推动中国特色哲学社会科学理论体系和研究方法的创新。

参考文献

[1] 马费成，李志元．新文科背景下我国图书情报学科的发展前景．中国图书馆学报，2020，(6)．
[2] 赵鼎新．社会科学研究的困境：从与自然科学的区别谈起．社会学评论，2015，(4)．
[3] 马费成，张帅．我国图书情报领域新兴交叉学科发展探析．中国图书馆学报，2023，(2)．
[4] 马费成．推动哲学社会科学创新发展．中国社会科学报，2021-07-20．
[5] 孙坦，张智雄，周力虹等．人工智能驱动的第五科研范式（AI4S）变革与观察．农业图书情报学报，2023，(10)．
[6] 陈小平．大模型关联度预测的形式化和语义解释研究．智能系统学报，2023，(4)．
[7] 矣晓沅，谢幸．大模型道德价值观对齐问题剖析．计算机研究与发展，2023，(9)．
[8] 陈国青，任明，卫强等．数智赋能：信息系统研究的新跃迁．管理世界，2022，(1)．
[9] C. A. Bail. Can Generative AI Improve Social Science? *Proceedings of the National Academy of Sciences*, 2024, 121(21).
[10] Z. Lin, S. Guan, W. Zhang et al. Towards Trustworthy LLMs: A Review on Debiasing and Dehallucinating in Large Language Models. *Artificial Intelligence Review*, 2024, 57(9).
[11] B. Romera-Paredes, M. Barekatain, A. Novikov et al. Mathematical Discoveries from Program Search with Large Language Models. *Nature*, 2024, 625.
[12] 吴丹，孙国烨．迈向可解释的交互式人工智能：动因、途径及研究趋势．武汉大学学报（哲学社会科学版），2021，(5)．

GAI-Empowered Philosophy and Social Sciences Research

Ma Feicheng, Chen Shuaipu（Wuhan University）

Abstract Generative artificial intelligence (GAI), an AI technology based on models trained on vast datasets which encompass a wealth of social values and factual knowledge, has significantly enhanced the efficiency of various social production sectors and scientific research with its machine intelligence. With the interdisciplinary integration and development of philosophy and social sciences with the new round of technological revolution and industrial transformation, GAI has become an advanced technological tool that enriches research methodologies within these fields, as well as a critical research subject, significantly expanding both research perspectives and observational frameworks. Given the profound potential for synergy between these domains, it is crucial to first examine the foundational conditions under which GAI can effectively empower philosophy and social sciences. Building upon this foundation, a thorough investigation into the innovative models and key challenges encountered during the empowerment process is necessary. Ultimately, a systematic strategy must be proposed to construct empowerment pathways, thereby clarifying the mechanisms through which GAI enhances philosophy and social sciences, fosters innovation within theoretical frameworks and research methodologies, and facilitates the advancement of philosophy and social sciences with Chinese characteristics.

Key words generative artificial intelligence; the theoretical system of philosophy and social sciences; research methods of philosophy and social sciences

■ **作者简介** 马费成，武汉大学人文社会科学资深教授，武汉大学信息管理学院教授，湖北 武汉 430072；
陈帅朴，武汉大学信息管理学院博士研究生。

■ **责任编辑** 何坤翁