

人工智能“模拟游戏”两种范式的哲学反思

郭鹏坤¹, 韩升², 王晶晶¹

(1. 河南工学院 马克思主义学院, 河南 新乡 453003; 2. 山东大学 马克思主义学院, 山东 济南 250100)

摘要: 就认知而言, 人工智能是对人类智能的模拟, 其实现方式主要有符号主义和联结主义两种主导范式, 二者分别映射了近代唯理论和经验论的认识论哲学。这两种主导范式都获得了一定程度的成功, 但是它们目前都存在某种预设, 具有一定的缺陷。从图灵的模拟游戏角度来看, 两种范式所实现的智能都是功能性的, 都属于物理主义, 与人类智能之间存在鸿沟。

关键词: 人工智能; 符号主义; 联结主义; 模拟游戏

中图分类号: B0

文献标志码: A

文章编号: 2096-7772-(2025)03-0064-06

0 引言

人类所经历的三次科技革命主要涉及人类感官功能的物理延伸, 而现代人工智能革命的兴起涉及人类智能的延伸。人们已经能创造出能够模拟甚至超越人类智能的系统。1955 年, 麦卡锡 (John Macthy)、明斯基 (Marvin Minsky) 等人共同起草了第二年夏季的《达特茅斯夏季研究项目关于人工智能的提案》, 首次使用了“Artificial Intelligence” (简称 AI)。该提案可以说是人工智能的宣言书: “该研究将基于这样的假设: 原则上, 学习的每个方面或智能的其他任何特征都可以被如此精确地描述, 以至于可以制造一台机器来模拟它。我们将尝试找到如何让机器使用语言、形成抽象和概念、解决现在为人类保留的各种问题, 并改进自己。”^[16] 人工智能自此成为一个新兴科学与技术工程领域的标志性词汇, 涵盖了认知 AI (Cognitive AI) 和工程 AI (Engineering AI) 两大分支及其互动性关系。认知 AI 是麦卡锡、明斯基等人在达特茅斯会议上关于人工智能阐述的出发点, 本文主要从认知 AI 角度对当前人工智能的两种主导范式进行哲学反思。

1 Artificial Intelligence 的构词内涵

从构词来说, artificial 和 intelligence 两个词都有

悠久的历史。artificial 来自拉丁词 artificialis, 由 ars 和 facere 组合变形而成。其中, ars 又源于古希腊语 τέχνη (technē), 主要指技艺, 其延伸含义还有工艺、艺术、知识等。facere 的拉丁词意为做或制造。合成的形容词 artificial 包含两种意思。其一, 人造的而非自然生成的。它所属之物不同于一般人造物, 是需要精湛技术的“精挑细琢”, 具有创造性知识、创新性技术要求的人造物。其二, Artificial 还被赋予了模拟性内涵, 这种人造物能模拟自然物的某些特性。例如, 在 20 世纪中叶该词就用于人造卫星 (Artificial Satellite)、人工心脏 (Artificial Heart), 传达了对自然实体某些功能的模拟。Intelligence 来自拉丁语 “intelligentia”, 是 “inter” (之间) 和 “legere” (选择、挑选等) 的合成词, 意为 “从中选择”, 对应希腊词 νοῦς (nous, 哲学翻译为努斯, 泛指心智、理性、理解等), 泛指心智或智能。智能到底是什么? 在人类思想史上一直充斥着莫衷一是的争辩, 学术界至今没有一个统一的定义。据统计, 目前, 人们对智能的定义至少有 70 多种, 在人工智能领域就有 18 种之多。如果用一个兼容广泛的定义, 智能可简化定义为 “完成复杂目标的能力”^[267]。但是, 智能形式是多维的, 包括人类智能、动物智能和人造智能或其他形式的智能^[32], 因此, 智能的实现并不依赖于某种物理基质, 它可以在碳基物

收稿日期: 2024-12-10

基金项目: 河南省高校人文社会科学研究一般项目 (2025-ZZJH-407)

第一作者简介: 郭鹏坤 (1982—), 男, 河南封丘人, 讲师, 博士, 主要从事西方哲学和伦理学研究。

通信作者简介: 韩升 (1978—), 男, 山东安丘人, 教授, 博士, 主要从事马克思主义理论、政治哲学研究。

理基质上实现,也可以在其他物理基质如硅基上实现。

作为合成词的 Artificial Intelligence 具有如下特征,首先,非自然性。人工智能是非自然演化或非自然生成的生物智能,是人为设计、人为制造的智能。其次,技术性、知识性。人工智能是依赖于复杂的人为编程、人为算法和人为的数据等驱动生成的智能。最后,模拟性。目前人工智能主要以人类智能为参照系,是对“人类智能”的模拟,如学习、认知、感知、推理、语言识别、图像识别等能力的模拟,能够自主地处理信息和执行任务。欧盟委员会人工智能高级专家组(AI HLEG)将人工智能定义为:“通过分析环境来显示智能行为的系统,并采取具有一定程度的自主行动来实现特定目标。”Haroon Sheikh 等人的类似定义是:“通过分析环境并采取行动(具有一定程度的自主性)来实现特定目标,来显示智能行为的系统。”^[46]这两种定义至少可追溯到早期人工智能专家纽厄尔(A. Newell)和西蒙(H. Simon)对智能系统实现既定目标能力的描述。这种强调以目标为导向的智能行为的功能性定义可能削弱了 AI 本身包含的强面向,不过,过高的期望会导致不切合实际的幻想。目前,接受类似的 AI 定义是相对明智的,这可以为 AI 未来的发展留下空间。作为一个新兴领域,Artificial Intelligence 这个词汇指称了一种跨学科综合和工程技术创新。在后续的研究中,其广度扩展到其他众多学科,其深度除了涉及硬件系统的优化升级,更涉及多种智能形式的深度拓展。

从广度而言,这一新兴领域不只涉及计算机科学,还涉及认知科学、神经科学、脑科学、心理学、哲学等领域,具有跨学科的融合性。认知科学对人类认知过程的理解,为 AI 编程、算法设计提供了理论基础和启发;神经科学为 AI 提供了对大脑神经网络架构的重要洞见,有助于设计模仿人类神经系统算法系统;脑科学进一步揭示了大脑如何处理信息,为 AI 提供了灵感和生物学基础;心理学提供了对人类行为和心理过程的深入理解,对于设计能够与人类互动并理解人类情感和意图的 AI 系统至关重要;哲学为人工智能提供了必要的概念和反思视角等。从深度而言,AI 更体现了其作为一个新兴研究领域的复杂性和多层次性。由于智能的表现是多维的,不仅仅关注计算机程序的设计和优化,而且要创造能够展示出类似人类智能的系统,这包括学习、理解、推理、感知、交流、复杂任务处理等多种能力。

2 从“模拟游戏”看人工智能的近代认识论映射

早在 1950 年,图灵(Alan Turing)就发表了人工

智能的奠基之作《计算机器和智能》。文中提出了“模拟游戏”(Imitation Game)概念,也被称为“图灵测试”(Turing Test)。智能模拟离不开对人类一些显著智能特征的理解,如感知、理解、推理等认识能力特征。以自然语言识别模拟为例,语言是人类智能的表现形式,自然语言表征着世界的结构、含义和基于这个结构和含义进行的判断、推理与决策。人工智能可以以计算机语言形式实现这种模拟,语法、语义和推理这三个概念共同构成了大语言模型理解和处理世界信息的基础。语法(syntax)提供了一种结构化的方式来表示和处理信息。没有语法,人工智能系统就无法识别和生成有意义的语言。语义(semantic)是人工智能理解世界的深层含义的关键,它不仅识别语言的结构,还能理解语言所传达的实际信息和意图。没有语义,人工智能系统就无法真正理解它所处理的数据或语言的含义。推理(inference)代表了人工智能系统基于语法和语义的理解来生成新信息和知识的能力。推理使人工智能系统能够从已知信息中推导出新的结论,进行决策并执行人类设置的任务。没有推理,人工智能系统就无法实现真正的智能和自主性。通过图灵测试意味着人工智能系统在对人类自然语言的识别和生成上能达到人类一般智能水平,但并不能反映和测量认知能力的水平。同时,目前的人工智能很难通过图灵测试。

20 世纪 80 年代人工智能领域存在一个哲学“意识形态”之争,即“Neats”(精致派)和“Scruffies”(粗狂派)争论应该采用何种技术路线辩论。前者认为 AI 系统知识表征和处理应该像数学一样严谨和优雅,AI 技术路线的设定应该基于数学一样的形式化和逻辑推理方法;后者则认为,智能是复杂的,AI 技术路线不应该限制在数学或逻辑框架内,而应该从数据和实际应用中学习,以此才能模拟人类智能的复杂性。这个哲学意识形态之争发展为 AI 两种主导范式——符号主义(Symbolism)和联结主义(Connectionism)。从哲学认识论角度来说,二者分别映射了近代哲学认识论的两条主要路线——唯理论(Rationalism)和经验论(Empiricism)。

近代哲学自笛卡尔开启了从本体论(Ontology)重心向认识论(Epistemology)重心的转移,其后,逐步形成了唯理论和经验论两大认识论对立阵营。唯理论的基本观点是,知识主要来源于逻辑推理,知识不能依赖于感官经验,感官经验经常会欺骗人,只有从自明的公理出发进行推导、演绎等才能获得真理。数学和逻辑学就是其中的典范。以笛卡尔为代表的唯理

论建构就是依据这个思路。在《第一哲学沉思集》中,笛卡尔从普遍怀疑出发,推导出一个绝对存在的思维着的“我”,进而推导出“我思故我在”这一哲学元命题。从这一元命题出发,又推导出身体和物理世界的真实性^{[5]14-33}①。而后,莱布尼茨提出了“普遍语言”思想,即一种能够精确表达人类思想的符号系统,这种系统基于严密的逻辑规则,使得复杂的概念可以通过简单的符号组合来表示。这为知识表示提供了一种可能的符号化方法,为符号主义的发展提供了理论支撑^{[6]81-99}。这就是计算机程序所依赖的数理逻辑思想,计算机的程序编写和执行都依赖于逻辑公式的符号表示和推导。总之,莱布尼茨的“普遍语言”概念为知识表示和逻辑推理提供了理论基础,推动了布尔的逻辑代数和弗雷格、罗素的逻辑演算思想的产生,对符号主义的发展产生了深远的影响。

符号主义 AI 又称为经典 AI,也称为逻辑主义学派 AI,在一定意义上是唯理论认识论在人工智能上的具体实现。符号主义理论基础建立在物理符号系统假设上,符号代表了现实世界中的对象和概念。人类的语言或思维过程,都可以被还原为计算机符号和进行符号处理,且在技术上完全可以由计算机来模拟。简而言之,符号主义依赖于使用物理符号系统建模来模拟人类的思维过程,计算机智能行为的输出就是通过物理符号系统的推导实现的。符号主义在理论上具有一定的吸引力,在人工智能的发展历程中曾经占据主导地位,推动了诸如专家系统和定理证明器等应用的发展。

20 世纪末,随着大数据时代的到来以及算法优化和算力的不断升级,联结主义 AI 迎来复兴,深度学习标志着联结主义的又一次高潮。联结主义的核心观点是智能行为可以通过模拟大脑神经元之间的连接和交互来实现。简单来说,联结主义需要根据任务目标构建一个神经网络模型,通常包括输入层、隐藏层和输出层,通过反向传播算法和梯度下降等来训练神经网络,人工神经元单元根据训练数据、特定算法和激活函数在学习训练过程中相互作用,自主调整它们之间的连接强度和激活阈值,从而有效实现对训练数据特征的提取、识别和分类,整个神经网络模型经过不断的学习-反馈-调整-优化,最后实现依据输入目标输出人们期望的结果。这从智能的实现技术原理上映射了经验主义认识论。

经验论的基本观点是,人类并没有天赋的知识,人类所有的认识都源于通过感官获得的外部世界的信息,知识是心智对这些外部信息的抽象、记忆、归纳、联想、联结等的结果。经验论的代表洛克在《人类理解论》中提出了“心灵白板说”,即人出生时心灵就像一块白板,没有任何知识或观念,所有的知识、观念都是后天通过经验经感觉(Sensation)和内省(Reflection)产生的。洛克的“心灵白板说”暗示了学习是一个逐渐积累的过程,知识形成是逐步的,从简单到复杂,这与联结主义的训练过程相呼应。智能系统从没有任何先验知识的状态开始,通过数据“投喂”进行学习训练,最终形成对复杂数据结构的某种“理解”。如在监督学习中,算法通过大量标注数据集的学习模式与洛克所说的通过感觉和内省获得简单观念的过程类似。在深度学习方面,智能系统通过模拟人类大脑的神经网络建模,算法根据大量的数据学习训练,能够通过提取学习数据的特征(类似简单观念)来获取知识,这类似于单观念组合为复杂观念。休谟在《人性论》中进一步发展了经验论哲学。休谟认为,人类感官接受外部物理世界的刺激产生的初步感觉印象,是所有知识的最基本构成单位。感觉印象在心智中通过记忆保存和进一步抽象生成观念来表征外部世界的对应物,观念通过想象力再次加工,按照统计学概率规律对观念进行组合和分解,形成观念之间的因果性联结,观念的组合就是所谓的“习惯”使然^{[7]149-290}。休谟的认识论观点与联结主义采用统计学方法提取特征知识高度吻合。在联结主义智能系统中,神经网络通过大量的训练数据来调整其内部连接的权重,能够从原始输入数据中提取复杂输出,类似于人类通过经验来学习和理解世界。联结主义 AI 有很多优点,如能够通过数据学习训练做出合理的预测和决策;具有自适应学习能力,通过算法,模型能够自动调整内部参数来提高性能;模型具有很高的灵活性、多样性,可以通过不同的神经网络模型来实现各种不同的任务和需求等等。

符号主义和联结主义是唯理论和经验论的哲学“映射”,这反映了现代人工智能专家只要以人类智能为参照系,通过构建人工智能系统实现对人类智能的模拟,就无法绕开近代哲学家对人类认识机制的哲学解释。哲学家最早以哲学理论的形式来理解人类自身的智能机制,现代人工智能专家也就不可避免地在

① 笛卡尔同时代的机械唯物主义者霍布斯在《利维坦》中也指出,思维过程可以被理解作为一种内部进行的加减计算过程,思维就是计算,不只是数学,而且逻辑学、政治学、法学等的本质都是计算。见:霍布斯.利维坦[M].张妍,赵闻道,译.长沙:湖南文艺出版社,2011,第22页。

自身技术路线上映射它们,从而在精神上产生某种交汇而具有某种亲缘性。这并不是说符号主义和联结主义的人工智能专家通过阅读唯理论和经验论哲学,受其启发形成和发展出各自技术路线的。近代这些哲学家在其生活的时代也不会想到要为未来的人工智能发展奠定技术路线基础。

3 两种范式的局限性

从技术实现的范式来说,符号主义和联结主义在人工智能领域映射了两种截然不同的哲学认识论,反映了二者在方法论与智能理解的分歧,二者有各自优势所在,也各自面临一些难题和具有局限性。

从方法论上来说,符号主义采用“自上而下”的方法论策略,强调从一般到特殊的推理过程。这种方法的核心在于使用符号系统来表示信息和知识,通过预先定义的规则、逻辑来推导特定问题的解决方案。这种方法在处理结构化问题和需要逻辑推理的任务时表现出色,如数学定理的证明、逻辑问题的解答以及开发基于规则的专家系统等。它的每一步推理都可以追溯和解释,表现出透明度和可解释性的优势。然而,符号主义的局限性在于其符号形式化的表达和高度依赖人工编程,在处理复杂问题、非结构问题时就显得力不从心了,如缺乏对模糊性、不确定性和非结构化数据的处理能力,也难以将逻辑推理和运算扩展到像自然语言识别、感知等智能形式领域。联结主义采用了“自下而上”的方法论策略,侧重于从特殊到一般的学习过程。联结主义智能模型通过模拟人脑神经元的复杂交互作用,在处理非结构化数据和模式识别任务时表现卓越,如图像和语音识别、自然语言处理等。这使得联结主义在处理模糊、多干扰性或不完整数据方面具有独特的优势。但是,联结主义复杂的神经网络模型和算法在输出结果时不仅会表现为结果的不可解释性,即所谓的“黑箱”问题,还可能造成和放大算法偏见。训练数据的数量、数据的质量、多样性和代表性也会直接影响模型的性能和能力等,这些问题可能会导致其具有一定的不可信赖性。

从对智能的理解来看,符号主义将智能视为符号处理的能力,智能可以通过符号的逻辑、计算的操作来解释其产生的原理。这个路线可追溯至纽厄尔和西蒙的“物理符号系统假设”(Physical Symbol System Hypothesis)。该假设认为“一个物理符号系统是具有实现一般智能行为的必要和充分方式”^{[8][16]},反之也成立。豪格兰特(J. Haugeland)又进一步指出,只要满足具有足够组织规模和适当组织形式的物理符号系

统就是智能系统,以此为基础可实现对人类逻辑思维的模拟。问题是,符号主义存在一个元问题,即“物理符号系统假设”。基于这个假设,将人类认识的复杂过程通过构建基于符号和规则的模型来模拟人类智能,实现化繁为简。符号代表世界中的对象和概念,规则定义了这些符号如何组合和操作以产生意义和推理,将复杂的推理任务又进一步还原为一系列符号操作和逻辑规则来实现。如果这个假设有问题,那么,这个建立在物理符号系统的智能局限性就凸显出来了,如编程不能穷尽规则,编程规则不能泛化以及编程的静态性难以应付动态多变的物理世界,从这个角度来说,这很难有效实现对人类智能复杂性的模拟。从本质上看,这种技术路线是一种(机械)物理主义。

联结主义通过神经网络中的连接权重、算法来模拟人类智能。联结主义不需要预设物理符号,正如休谟所认为的,符号不过是前符号的感觉材料心理加工后的输出物(对应联结主义AI的训练数据)。现实生活中,人们对不同的感觉材料会形成不同的观念和联结知识,就像不同的算法神经网络模型通过相对应任务数据“训练”可以自主产生不同的输出结果,整个系统就实现了将整体任务的实现最终分解为单个神经元的信息传递状态,它们的最终“合流”就是目标任务趋势。因此,联结主义“不关注在符号层面对人类信息的处理过程,而是注重如何在亚符号的层面上,以数学方式模拟人类神经网络的运作方式,并通过对于此类神经网络的‘训练’以使得其能够给出用户所期望的合格输出”^{[9][17]}。联结主义方法的设计灵感来自于模拟人类大脑神经元及其突触连接的方式,特别是深度学习模型,通过大量简单计算单元间的加权连接,能够捕捉和学习输入数据中的复杂模式,在诸如视觉和听觉感知等任务上取得的成功,被一些学者认为是对生物神经网络功能的一种粗略但有效的近似。但是,这些成功并不意味着人类智能可以以此来理解。如同符号主义,联结主义其实有一个算法预设,即“智能就是算法”,这是AI领域的核心观念。这种观念可以追溯到近代唯物主义哲学家霍布斯那里。其实,人类智能的本质远比算法更为复杂和多维,简单地将智能归结为算法忽视了智能的复杂性。人类智能的许多方面,如情感、直觉、想象等,在人类智能中扮演着重要角色,也难以用算法来表达。更不用说,联结主义只模拟了人类神经网络表面的、简单的运行机制,其复杂程度远远不如人类大脑,更无从谈及模拟整个人类大脑的智能。正如福多(Jerry A.Fodor)所指出的,联结主义过于简化了认知过程。尽管联结主

义模型在某些感知任务上表现出色，但它们往往忽略了智能行为背后的复杂认知结构和功能^{[10]33-51}。因此，联结主义从本质上看也是物理主义。

4 “模拟游戏”的智能难题

自人工智能诞生以来，领域内的一些研究者不仅对之充满乐观，而且寄予厚望。如西蒙在1965年预言，人工智能将在20年内完成人类能做的任何工作。古德（I. J. Good）预言，人工智能将引起“智能爆炸”，它将是人类的最后一项发明，它将是地球上最聪明的生物，可以完成所有的发明创造。类似的预言太过于乐观，是充满幻想的AI神话。

我们回到1950年图灵在论文中提出的“模仿游戏”，即用“图灵测试”来解释为何这种预言具有神话性。图灵测试传达的主要思想是：一名人类审问者通过打字的方式与两个隐藏参与者交流，其中一个是人，另一个是计算机。审问者的任务是通过提问和回答的方式，来判断其中哪个参与者是计算机。如果审问者无法一致地区分出二者，那么机器就可以被认为具有思考能力，即展现出了人类水平的智能^{[11]433-434}。简而言之，这是通过模拟智能行为表现来推断或追溯计算机是否具有智能的方法，这是一个简单而直观的标准。模拟游戏的关注点是计算机输出的智能行为表现，而不是它的内部状态或者说思考过程。也就是说，人工智能所表现出的智能是基于计算机语言的信息处理，它的理解或者说思考并不对应于真实世界及其意义，它是基于对数字世界感知和理解所表现出的符号或数字智能，而非像人类一样的生物智能。从技术发展未来角度来看，符号主义与联结主义智能系统都有通过图灵测试的可能。但从更深层来看，这个模仿游戏暴露了人工智能与以人类为代表的生物智能之间存在难以跨越的鸿沟。这个鸿沟由塞尔（J. Searle）的“中文屋思想实验”暗示出来。

塞尔想象了一个不懂中文的英语说话者被封闭在一个房间里。这个房间有一个小窗口，通过窗口，外面的人可以递进一些用中文写的纸条。房间里有一本用英文写的中文操作指南，告诉这个英语说话者如何根据纸条上的中文符号来查找和操作一系列中文符

号，并将回应通过窗口递出去。对外面的中文使用者来说，似乎房间里的人能够理解中文，但实际上，房间里的人只是在遵循与中文对应的英文操作指南来回应，他并不理解中文^{[12]417-418}。可以说，塞尔用这个实验揭示了人工智能与人类生物智能之间存在难以跨越的鸿沟，它们功能相似，但智能实现机制差别很大。当然，塞尔这个“中文屋思想实验”的主要目的不是揭示这种鸿沟，而是通过这个思想实验来界定清楚弱人工智能与强人工智能的不同。弱人工智能是在特定领域内表现出类似于或超越人类的能力，但它们并不具有理解现实世界的能力。强人工智能是具有认知能力泛化的智能体，不仅能在多个领域内执行任务，还能模拟出与人类相似的意识。可见，符号主义和联结主义都属于弱人工智能。

随着科学技术的发展，人类是否有可能实现强人工智能？强人工智能或许代表了人工智能的最终目标，它似乎揭示了Artificial Intelligence更强的面向。不过，根据上文的反思，符号主义和联结主义两种范式是根本无法实现强人工智能的。即使人类完全理解了人类智能的产生机制，并以此为基础制造出了强人工智能和实现人类意识的模拟，将会面临更大的难题。这个难题不在于意识产生机制的完全理解，而在于中间还横亘着一个查尔莫斯难题^②。我们从理论上有可能完全理解大脑的功能以及它如何产生意识，但无法解释这种意识为什么和如何产生，更遑论制造出这种意识了。没有这种意识，人工智能即使是强大的，也只可能是一个没有“灵魂”的“僵尸”。如果解决了上述全部问题，一旦这种强人工智能“神话”真的照进现实，我们不知道这将会带来什么样的不确定风险和挑战。我们不要忘记了麦卡锡在发展AI的宣言书中所暗示的目的：发展AI是为了提升工具的智能性，帮助人类解决问题——换言之，发展AI是为了辅助人类、增强人类、造福人类而非取代人类。

5 结语

从认知科学角度来说，人工智能是对人类智能的模拟，其两种主导范式——符号主义和联结主义都取得了一定的成功，但也都存着明显的局限性。符号主

② 澳大利亚哲学家查尔莫斯（David J. Chalmers）区分了关于意识的“容易问题”和“困难问题”。容易问题涉及的是意识的机制和功能，这个问题虽然复杂，但原则上可以通过科学方法来解释。困难问题则是关于主观经验如内在体验、感受等，关注的核心问题是意识的主观经验——即“物质过程为什么会和如何产生主观经验”，即“意识困难问题”（hard problem of consciousness）。困难问题挑战了传统的物理主义观点，即所有的现象都可以归结为物理过程，但是涉及意识的主观经验无法用物理过程来解释。见：CHALMERS D J. Facing up to the problem of consciousness[J]. Journal of Consciousness Studies, 1995, 2(3): 200-219.

困难问题仍然是意识研究中的一个未解之谜，并且对于人工智能也提出了挑战，即如何创建能够体验主观感觉的机器，或者这样的机器是否可能存在，笔者另有文章讨论此问题。

义 AI 暴露的物理机械性忽略了认知过程中的多维性和复杂性, 而联结主义 AI 虽然在模拟神经网络运作上取得了很大的进展, 但从本质上与符号主义区别不大, 两种范式的智能模拟都是物理主义, 其背后的底色是计算功能主义 (Computational Functionalism)。塞尔的“中文屋实验”和人工意识难题进一步揭示了这两种技术路线存在着人工智能与人类智能之间横亘着的难以跨越的鸿沟。不过, 人类发展 AI 技术旨在辅助人类而非取代人类, 即使它们未来能在技术上跨越鸿沟、迈向强人工智能时, 人们务必要保持谨慎的态度。同时, 要对其发展不确定性进行前瞻性哲学思考和准备。

参考文献:

- [1] BOSTROM N. Super intelligence: Paths, Dangers, Strategies[M]. Oxford: Oxford University Press, 2014.
- [2] 麦克斯·泰格马克. 生命 3.0[M]. 王婕舒, 译. 杭州: 浙江教育出版社, 2018.
- [3] HOFFMANN C H. Fighting human hubris: Intelligence in nonhuman animals and artifacts[J]. Ethics and Bioethics, 2023, 13(1): 1-14.
- [4] SHEIKH H, PRINS C, SCHRIJVERS E. Mission AI: The New System Technology[M]. Cham: Springer International Publishing, 2023.
- [5] 笛卡尔. 第一哲学沉思集[M]. 庞景仁, 译. 北京: 商务印书馆, 1996.
- [6] 王自贵. 莱布尼茨逻辑思想研究[D]. 重庆: 西南大学, 2016.
- [7] 休谟. 人性论[M]. 石碧球, 译. 北京: 九州出版社, 2007.
- [8] NEWELL A, SIMON H A. Computer science as empirical inquiry: symbols and search[J]. Communications of the ACM, 1976, 19(3): 113-126.
- [9] 徐英瑾. 人工智能哲学十五讲[M]. 北京: 北京大学出版社, 2021.
- [10] FODOR J A, PYLYSHYN Z W. Connectionism and cognitive architecture: a critical analysis[J]. Cognition, 1988, 28(1): 3-71.
- [11] TURING A M. Computing machinery and intelligence[J]. Mind, 1950, 59(236): 433-460.
- [12] SEARLE J R. Minds, brains, and programs[J]. The Behavioral and Brain Sciences, 1980, 3(3): 417-457.

Philosophical Reflections on the “Imitation Game” of Artificial Intelligence: Two Paradigms

GUO Pengkun¹, HAN Sheng², WANG Jingjing¹

(1. School of Marxism, Henan Institute of Technology, Xinxiang 453003, China; 2. School of Marxism, Shandong University, Jinan 250100, China)

Abstract: In terms of cognitive AI, artificial intelligence is a simulation of human intelligence, with two dominant paradigms for its realization: symbolism and connectionism. These two paradigms respectively correspond to the epistemological philosophies of modern rationalism and empiricism. Both dominant paradigms have achieved a certain degree of success, but they both have certain presuppositions and inherent flaws. From Turing's perspective of the imitation game, the intelligence realized by both paradigms is functional and belongs to physicalism, with a gap existing between them and human intelligence.

Key words: artificial Intelligence; symbolism; connectionism; imitation game

(责任编辑 陈秀娟)