

科技评论

人工智能到底是一门什么学问？

李国杰

摘要 人工智能作为一门发展近 80 年的学问,其发展过程中边界持续扩大、研究方向不断丰富,学界对其学科属性、研究核心等问题仍存争议。对人工智能的学科属性进行了系统分析,剖析了人工智能的学科定位,分析了人工智能可以像人一样思考、像人一样行动、理性地思考和理性地行动这 4 个核心目标,指出了人工智能是一种方法而不仅是模拟人类,是求解复杂问题的新范式,人工智能已超越传统计算机科学分支的范畴。强调了人工智能是科学技术的“元级探索方法”,目前仍处于“前范式科学”阶段。

关键词 人工智能;学科属性;计算机科学;机器学习;科学范式

人工智能(artificial intelligence, AI)作为一门学问已经发展了近 80 年。现在人们常将 1956 年称为人工智能的元年,因为那一年召开的达特茅斯研讨会上,最先采用了“人工智能”这个术语,但实际上那次为期 2 个月的研讨会并没有留下里程碑式的文献。麦克洛克(Warren McCulloch)和皮茨(Walter Pitts)1943 年在《Bulletin of Mathematical Biophysics》期刊上,发表了第一篇关于神经网络模型的论文《A Logical Calculus of the Ideas Immanent in Nervous Activity》^[1],这个经典的 M-P(McCulloch-Pitts)模型对当今人工智能的发展仍然发挥巨大作用。因

此,人工智能这门学科从 1943 年算起,可能更加合理。

近 80 年来,人工智能的发展起伏,但边界一直在扩大,研究方向也在不断增加。人工智能的研究方向不仅包括传统的问题求解、知识表示、逻辑推理,自然语言理解、计算机视觉、机器学习、机器人等,还包括大量的 AI+X 的交叉融合内容^[2]。长期以来,人工智能被认为是计算机学科的一个分支,但人工智能的发展已经超出了传统的计算机学科的范围。教育部已经将“智能科学与技术”设立为一级交叉学科,但对人工智能到底是一门什么学问,还是有很多疑问。

从人工智能诞生起,就一直面临人工智能要研究什么的争论。虽然众说纷纭,但基本上是 2 条路

线之争,一条强调人工智能要模拟人的能力,目标是“像人”:从科学的意义上讲,就是要弄清楚“智能”的奥秘。另一条是把人工智能当成一种解决复杂问题的方法,目标是设计正确而且高效的人造“智能体”。这两条看起来完全不同的技术发展路线,为什么会长期纠缠,包含在一个学科中?从宇宙和人类进化史来看,包括人脑在内的“智能体”可能是进化至今的终极“产品”,因此探索智能的奥秘应当是人类需要解决的最重大的科学问题,但目前人工智能还处在科学范式形成前的百家争鸣阶段,没有形成有共识的基础理论。另一方面,人工智能的技术丰富多彩,已经走在科学的前面。作为一门学科,现在的人工智能的主流究竟是像数理化一样的科学,还是一

中国科学院计算技术研究所,北京 100190

收稿日期:2025-07-16;修回日期:2025-09-23

作者简介:李国杰,研究员,中国工程院院士,研究方向为计算机体系结构、并行算法、人工智能、大数据、计算机网络、信息技术发展战略等,电子邮箱:lig@ict.ac.cn

引用格式:李国杰. 人工智能到底是一门什么学问? [J]. 科技导报, 2025, 43(20): 17-24; doi:10.3981/j.issn.1000-7857.2025.07.00093

门工程技术,这也是需要澄清的问题。本文对人工智能究竟是一门什么学问发表一些看法,旨在抛砖引玉,引起讨论。

1 人工智能的 4 个目标

人工智能的定义五花八门,但各种各样的人工智能教科书也有一些共同的内容。其中斯图尔特·罗素(Stuart Russell)和彼得·诺维格(Peter Norvig)1995年出版的《Artificial Intelligence: A Modern Approach》(《人工智能:一种现代方法》)是一本影响较大的经典著作,已经发行第4版^[3],被全球1500多所大学用作人工智能的教科书,说明这本书讲的观点得到许多人工智能学者的认同。在这本经典著作中,人工智能的核心目标被归纳为4类:像人一样思考,像人一样行动,理性地思考和理性地行动。这4个目标构成了人工智能研究的多元视角即前两者关注与人类的相似性,后两者强调理性效能。

“模拟人类”和“理性的思考/行动”这2条路径表面上目标不同,但本质上却深度相关,它们共享以下核心假设。(1)相同的信息处理观:不论是人脑还是机器,智能的本质都是对信息的接收、加工与利用,强调在信息不充分的条件下有效解决问题的能力。(2)一致的目标驱动:无论模拟人类思维还是追求理性效能,最终目标均为构建具有自主决策、学习和适应能力的智能系统。两者从不同角度切入,共同回答“什么是智能”这一根本问题。(3)应对不确定

性:认知行为和复杂任务求解都面临模糊、未知与动态变化的环境,人工智能研究必须面对复杂性和不确定性。

像人一样思考/行动的研究关注智能体的“人类兼容性”,从人类认知与行为中提取智能的表象特征(如语言、情感、社会互动),重视机器的智能行为能否被人类理解或接受。理性思考/行动的研究关注智能体的“效能”,从数学逻辑与优化中提炼智能的本质规律,如逻辑一致性、最优决策等,目标是在复杂环境中高效达成目标。人工智能的发展逐渐实现了从模拟人类到实际应用的融合:“像人一样思考”往往正是解决问题的最佳参考。模拟人类可为理性思考提供启发,人类思维(如直觉、常识)常包含非形式化但高效的决策模式,可为理性算法提供灵感。理性优化也可以为模拟人类提供工具,逻辑推理、概率模型等理性方法可提升模拟人类的精确性和泛化能力。两者通过方法互补和目标统一紧密关联,最终服务于AI学科的终极命题:创造既能理解人类,又能超越人类局限的智能系统。这种多元统一性是AI作为独立学科的核心凝聚力。

人工智能的4类目标如同“智能”的四维坐标,像人一样思考/行动定义了智能的“人文维度”(与人类兼容);理性思考/行动定义了智能的“科学维度”(与客观规律一致)。现代AI融合了四者,形成“人类启发+理性优化”的混合范式。在实际应用系统中,四者常需要结合。这样的案例很多,例如:大模型结合了人类语言数据与

自监督学习;在自动驾驶中,既需要模拟人类驾驶行为(像人一样行动),又需要理性规划路径(理性行动);在医疗诊断AI中,既需理解医生的决策逻辑(像人一样思考),又需通过数据提高诊断准确率(理性思考)。

2 人工智能是设计智能体的现代方法

对人工智能研究和应用而言,上述4个目标的重要性是不同的。模拟人类的思考和行动不是最终的目标。图灵提出的验证机器是否具有思考能力的方法,就是采用机器是否做出类人行为的测试(称为图灵测试),但这种测试实际上是关于人类易受骗性的测试,是否通过图灵测试并不重要。人工智能的研究应该发展科学技术,给人类提供更聪明更有用的工具,进一步走向人机协作,而不是为了让测试者上当。如果航空工程学也采取类似图灵测试的目标,应该是制造“能像鸽子一样飞行,甚至可以骗过真鸽子的机器”。好在飞机发明家和航空工程师没有模仿鸟类,转而研究空气动力学并使用风洞,人工飞行的探索才取得成功。

《人工智能:一种现代方法》这本经典教科书中,虽然提到人工智能的4类研究目标(或称为4种范式),但全书的重点是讲“理性行为”,书中明确指出,“在人工智能的大部分历史中,基于理性智能体的方法都占据了上风”,“人工智能专注于研究和构建做正确的事情的智能体”。智能体比传统的计算机程序能完成更多的任务,包

括感知环境、自主运行、自适应调控和实现目标。所谓“正确的事情”是人类提供给智能体的目标定义。为了区别于“模拟人类的思考和行为”，这本书的取名非常明确地告诉大家，人工智能是以机器理性为核心的一种“现代方法”。这种理性是“目标函数+算法策略”的理性，不一定像人那样思考，也不一定像人那样行动。人造智能体与人的最大区别是，其理性是被建模出来的，其本质不是“模仿人类”，而是设计出来能够在复杂环境中“合理行动”的机制。这种“现代方法”可以定义为一个统一的框架：一个理性智能体是感知其环境并做出行动选择的实体，其目标是最大化某种性能度量函数下的期望结果。无论任务多么复杂，都可以形式化为“输入感知→状态估计→行为选择→性能优化”这一链条。这个框架是一种通用范式，可以称之为人工智能的“标准模型”。

为什么要强调人工智能是一种方法而不仅仅是模拟人类？笔者认为至少是出于 2 方面的考虑。首先，模拟人类一定要先清楚人脑如何思考和行动，这实际上是脑科学和认知科学的任务。在人工智能发展的初期，学者不只满足于设计的程序能正确地求解问题，而是更关心将推理步骤与求解相同问题的人类进行比较，在一段时间内，人工智能和认知科学是混淆的。后来认识到脑科学和认知科学还需要作长期的基础研究，脑科学与认知科学就与人工智能相对分离地发展了。现在的人工智能基本上是一种技术，求解的问题不

限于人类的认知，更多的是考虑在科学研究和工程实践中传统科学技术解决不了的复杂问题。近几年 AI for Science 和各行业机器学习垂类模型蓬勃发展就反映了这个趋势。另一方面，如果只专注于模拟人类，人工智能可能会引向错误的方向。发展人工智能绝对不是为了替代人类，而是为了增强人类。模拟人类的所有能力，发展完全替代人类的“人类掘墓者”，不应该是人工智能的追求目标。因此，应当根据人类的需求，对模拟人类的研究方向做出取舍。未来，AI 可能不仅是一门技术，而是成为继演绎逻辑与实证科学之后的知识生产和问题求解的新范式，为人类应对客观世界的不确定性与复杂性开辟新的途径。

3 人工智能是求解复杂问题的新范式

为了应对客观环境，通过长期的经验积累，人类掌握了很多技术和诀窍。自从哥白尼、伽利略、牛顿等科学家开创第一次科学革命以来，人类在天文、数学、物理、化学、医学、工程等领域掌握了大量的科学技术，相对简单的问题用现有的科学技术都可以解决，但还有很多复杂的问题用现有的科学技术找不到满意的解决方案，必须寻求新的解决办法。通过近百年的努力，人类找到了一条解决复杂问题的新途径，这就是人工智能。

所谓复杂问题往往具有以下特征。(1) 多元性：问题中包含数量极多的相互作用的变量或因素；(2) 非线性：不满足线性叠加原

理，可能导致混沌等复杂行为；(3) 模糊性：问题没有明确定义，边界模糊；(4) 不确定性：结果难以预测；(5) 动态性：问题输入和边界条件随时变化。人工智能面对的复杂问题主要是以下 2 大类。

3.1 引起组合爆炸的复杂问题

组合爆炸在数学界被称为“维度灾难”(curse of dimensionality)，是指随着问题规模(维度)增加，求解难度和所需数据量呈指数级增长的现象，反过来说，就是高维空间中数据极端稀疏、距离度量失效的现象。例如，若一维空间用 10 个点可以充分采样，那么在 100 维空间中，达到类似的采样密度就需要 10^{100} 个样本。科学研究和工程实现中常常需要在指数爆炸的解空间中进行组合搜索，寻找满足某种要求的答案或多目标最优解。做蛋白质结构预测，1 个包含 100 个氨基酸的蛋白质三维结构，理论解空间大小约为 $9^{100} \approx 10^{95}$ 。集成电路设计的理论解空间高达 10^{540} ，几乎是无限大。

在计算机科学中，这一类计算时间随问题规模指数级增长的问题，称为 NP 困难问题，寻找 NP 困难问题的满意解是计算机科学的重大挑战，几十年来虽然有些进展，但不令人满意。基于人工神经网络的机器学习将高维数据的无结构搜索，转化为有结构的函数逼近，不用遍历所有维度组合，而是只关注局部相关、共享的模式和少数潜在特征，巧妙地减少了有效维度，极大地缓解了维数爆炸带来的参数暴涨和数据需求。这种“以深度换宽度”的策略，使得深度网络能够用合理数量的参数表

达某些本来需要几乎无穷多的参数才能表达的函数,提供了绕开维度爆炸的有效途径。近几年,机器学习在蛋白质结构预测、新材料发现、集成电路全自动设计等领域的突破表明,人工智能以组合搜索为切入口,有望成为一种发现自然规律、设计物理结构、合成新材料、发明数学算法的新工具。这不只是工具的更新,而是一种新型“组合智能”的诞生,未来潜力可能比深度学习本身更为深远。

3.2 定义不明确,边界模糊的“脏问题”

科学发展史中,所谓科学问题常常被形象地比喻为“钟表”——结构精密、因果明确、规则清晰。牛顿力学、电磁场理论等经典科学,多以这种可预测、可解析、可验证的“钟表式”知识为基础。拉普拉斯认为宇宙运行完全由初始条件和物理定律决定,宇宙的过去、现在和未来是一个连续因果链。若已知某一时刻的完整信息,理论上可通过数学计算推演出宇宙的全部历史与未来轨迹,甚至能“像回顾历史一样预见未来”。然而,随着科学技术问题的复杂化,科学家和工程师们越来越多地面对一些边界模糊的“脏问题”(wicked problems),不再具有钟表式的知识。

所谓“脏问题”这一概念,最早由霍斯特·雷特尔(Horst Rittel)和梅尔文·韦伯(Melvin Webber)在1973年发表的论文《Dilemmas in a general theory of planning》^[4]中提出,学术界也翻译成“棘手问题”或“抗解问题”。“脏问题”往往具有以下特征:(1)问题无明确

定义,问题边界模糊,不同人对“问题是什么”有不同理解;(2)解决方案没有绝对的“对”或“错”,只有“更好”或“更差”,解法具有主观性,常取决于价值观和判断标准;(3)问题具有独特性,没有通用解决方法,无法套用通用公式或模板来解决问题;(4)问题之间相互交织,难以切割为小问题单独处理;(5)许多解决方案是一次性尝试,一旦尝试过某种策略,情况可能发生改变,不能“重来”,有些问题尝试求解后可能带来不可逆的后果,难以通过试验来学习;等。对人类的感知和认知的模拟也属于这类问题。例如机器翻译,将一本中文古典名著翻译成英文,翻译是否正确并没有公认的判断标准。在这个背景下,人工智能作为一种新兴的通用技术,对“脏问题”处理能力的表现令人瞩目。有学者将人工智能比喻为“抹布”——能把现实世界中各种“脏”的表面擦干净。这一比喻生动形象,突出了AI处理模糊问题的能力。从“解空间已知+目标明确”到“解空间未知+目标模糊”的转变,正是AI方法本质性的特点。

人工智能目前的处境和量子力学诞生时非常相似,当时大多数科学家不接受量子力学模棱两可的表示。测不准原理的提出者沃纳·卡尔·海森堡(Werner Karl Heisenberg)认为,量子力学用“骑墙者”的语言描述微观世界可能是一个“语言学问题”,而不是“物理学问题”,将来经典力学和量子力学也许会共享同一个话语体系,彼此和谐共处。量子力学哥本哈根学派的奠基人尼尔斯·玻尔(Niels

Bohr)一口回绝了这种可能,他用“抹布”比喻来反驳海森堡:语言其实就是一块“抹布”,你无法要求存在一块“绝对干净的抹布”,在擦拭桌子之后依然保持“干净”。

AlphaGo战胜世界围棋冠军李世石时,许多人下结论:人工智能只适合于解决像围棋一样边界清楚、规则明确的问题。但是,深度学习在文本图像生成、机器翻译等领域的成功改变了人们对人工智能的认识。人工智能不但在解决边界清晰、规则明确的问题上表现出色,目前正在获得“模糊任务理解”“自我建模边界”的能力^[5]。它不是只适合封闭系统,而是从封闭系统出发,向开放世界的泛化智能系统演进。人们不再要求必须先明确定义问题和规则,而是通过大量实例数据学习输入与输出之间的关系,从中找到规律和问题的满意解。人工智能不是钟表,也不只是抹布,而是融合模式识别、策略演化、语言表达与逻辑推理的全新的“智能生成范式”。它既能在混沌中找秩序,也能在严谨中找突破;既能辅助人类,也在重构人类的知识定义方式。未来的AI可能是一种“超级认知系统”,将人类的经验、理性与机器的“认知”密切结合,成为人类面对复杂世界不可或缺的合作伙伴。

4 人工智能已经不只是计算机科学的一个分支

人工智能促进协会(AAAI,原称美国人工智能学会)对人工智能的定义曾经是“人工智能是计算机科学的一个分支”。笔者过去的

文章也强调人工智能是计算机学科的分支。但是,近几年人工智能的发展让我改变了看法。笔者现在认为人工智能已经不只是计算机学科的一个分支,它研究的领域和研究方法,都与传统的计算机学科大不一样了。

在传统计算机科学中,问题通常具有以下特点:(1)求解的问题是形式逻辑、算法性问题;(2)解空间由离散、可枚举的解构成;(3)目标是求正确解或最优解;(4)解决范式强调推理,主要采用规则推导、算法计算、逻辑演绎,传统编程是“人写规则→计算机执行”;(5)输入、输出、约束、目标函数都有精确、清晰的形式化定义;(6)可验证性强,是否通常有唯一标准答案。

而对于人工智能,问题边界常常是模糊的,评价标准主观性强。机器翻译、图像识别、对话系统的“好坏”往往依赖于人类评价或任务上下文。解决方式具有非确定性,往往依赖概率模型、统计推断、启发式搜索、机器学习而非精确推导。求解目标不一定是最优解,往往是满意解。从科学范式上讲,当前的 AI 与计算机科学主要的区别,是从“规则驱动”走向“数据驱动”,从“用已知算法求解”到“通过机器学习求解”,这是一种范式转变。表 1 归纳了传统计算机科学和人工智能的区别。

从历史渊源看,人工智能最初确实是计算机科学的一个子领域。但随着 AI 的发展,它在研究目标、方法论、评价体系等方面,逐渐展现出与计算机科学其他分支(如编译原理、操作系统、计算

复杂性理论等)显著不同的学科特征。人工智能越来越多地借鉴了其他学科的知识和方法,正在实现跨学科的融合。计算机科学关注“如何构建计算系统解决精确定义的问题”,而 AI 关注“如何构造能够自己学习解决问题的机器”,包括那些人类不能完全形式化定义的问题。人工智能用概率建模代替精确定义,在图像识别中,不是问“某张图像是不是猫”,而是问“它是猫的概率是多少?”用目标函数代替显式规则,优化问题变为最小化某种损失函数,让模型自己学习达到最优。人工智能用数据驱动的“经验空间”逼近问题空间,无法明确定义“猫”,就收集数百万“猫图像”的样本,建模其高维特征分布。这些方法本质上是用逼近策略来形式化模糊问题,使

其可训练、可优化、可评估。现在的 AI 更像是一个跨学科的新学科,而非计算机科学的一个分支。表 2 归纳了人工智能借鉴其他学科的内容。

在传统的计算机科学中,是对一个问题类中的所有待计算的问题特例(problem instance),都采用同一个明确的固定的程序来计算。例如排序算法就是针对所有的序列进行排序,即使序列趋向无穷长,也可以用同一个方法求解。而在人工智能中,往往没有精确定义的问题类,实际上是对问题的特例求解。这是人们经常忽略的一个重要区别,因为面对给定的特例,就有可能利用此特例中一切可以利用的特点,如对称、规律、局部模式等,使得消耗的资源尽可能减少。通用人工智能(AGI)的倡

表 1 人工智能和传统计算机科学的区别

角度	传统计算机科学	人工智能
问题定义	明确、可形式化	模糊、不完备
问题类型	形式逻辑、算法性问题	感知/认知,复杂问题
主要求解方法	算法+逻辑推理	数据建模+统计学习
解空间	离散、可枚举	连续、高维、不可完全枚举
求解目标	正确解、最优解	满意性、近似解
可解释性	强	弱
可验证性	严格	近似或统计性
执行模式	人写规则→计算机执行	人给数据→机器自动归纳
鲁棒性与泛化性	泛化能力较弱	追求较强的泛化能力
当前趋势	保守稳定	高速演化、跨学科融合

表 2 人工智能借鉴其他学科的内容

学科	借鉴内容
数学	概率论、线性代数、求解反问题
统计学	贝叶斯推断、假设检验、回归模型
神经科学	生物启发的学习模型
语言学	自然语言语义分析
哲学	意识、推理、认知框架

导者王培对智能的定义是:智能是在有限资源的情况下,做出最佳适应的能力。这个定义实际上是把智能和计算复杂性联系起来了。克服复杂性是驱动智能产生和发展的原动力,智能正是为了超越计算复杂性而形成的能力。目前的大模型大量采用猜测加验证的求解方法。所谓猜测就是一种试错,通过试错可以获得反馈,通过反馈可以更好地寻找搜索方向。智能体与过去的程序不同之处在于能够和给定的特例互动。针对特例的计算和非确定图灵机有非常密切的关系,非确定图灵机既是计算理论的重大问题,也是人工智能的重大理论问题。过去只有理论界关心非确定图灵机,针对问题特例的猜测加验证求解方法推动非确定性图灵机的研究走向实用。这是近几年人工智能取得重大突破的重要原因。

5 人工智能是科学技术的“元级探索方法”,还处在“前范式科学”阶段

人工智能不只是模拟人类认知的特定工具集,而是一种用于探索未知、发现规律、生成模型的“元级探索方法”。人工智能是一种“通用问题求解框架”,能够在

知识不完备、资源不充分的条件下获得满意解,可广泛用于数学、物理、化学、生物等领域的基础科学研究,也能在各个领域的技术开发、工程实施中发挥重大作用。对处理政治、经济和社会管理问题,人工智能也是一把利器。因此,不能简单地把人工智能当成一个学科专业,它是驱动知识前沿不断扩展的新兴技术科学范式。可以把人工智能比作“科学技术的望远镜”与“认知的发动机”。

人工智能的巨大作用体现在它引起的认知革命,在人类的理性之外,人工智能创造了一种“机器理性”,或者称“机器认知”。机器理性不同于人类理性,两者区别见表3。人类理性是“自然界演化出来的有限智慧”,机器理性是“设计出来的最优化程序”。AI的最终目标不是看起来像人类,也不是思考方式模仿人,而是在特定环境下,采取最大化效用的策略。这也是强化学习、智能体规划、多目标优化等技术成为AI核心方法的原因。人类理性强调“适应性与意义”,机器理性强调“可控性与效率”。下一代AI,也许不是更像人,而是更理解人。人类的身边多了一个能“思考”的“协作伙伴”,这是人类从未遇到过的科技革命和认知革命,对其影响绝不能低

估。人类理性不容易形式化,机器理性也不易类人化。理解它们的差异,是确保AI向安全可控方向发展的前提。只有“工具理性”的AI可能给人类带来重大的风险,如果目标函数设置错误或者AI完全自主地设置目标,AI可能高效完成“有害任务”。

AI对人类社会的影响远远超过一门普通的学科,必须充分估计它的潜力。经过几十年的努力,AI已经不是像炼金术和占星术一样停留在假说阶段的“前科学”,而是一个可以广泛应用的通用技术,已包含较深厚的科学基础和较成熟的技术栈(各种学习算法、推理引擎、知识表示方法等)。它不仅仅是“准备成为科学”,它本身就是一个正在快速发展的、融合了科学、工程和应用技术的混合体。但是了解“智能”的奥秘是一个极难攻克科学目标,总体来讲,人工智能目前还处在“前范式科学”阶段。科学哲学家托马斯·库恩(Thomas Samuel Kuhn)指出“前范式科学”阶段的特点是:(1)没有统一的研究范式,主要依赖试错、经验和直觉进行探索;(2)存在多个互相竞争的理论体系;(3)基本概念、研究方法和标准尚未统一;(4)学者之间对“什么是重要的问题”“什么是好的解释”还没有达成共识。目前的AI明显具有上述特征。近几年虽然Transformer架构主导了“深度学习”子领域,但还没有在符号推理、因果学习、具身智能等方向起到关键作用,并未成为整个人工智能领域的统一研究范式。当前的基准测试虽带来一定共识,但仍无

表3 人类理性和机器理性的区别

比较维度	人类理性	机器理性
形成机制	生物演化、社会文化	工程设计、算法优化
决策基础	经验、直觉、情感、道德	目标函数、输入映射
推理风格	模糊、灵活、举一反三	数据驱动、可计算
适应能力	有较强适应能力	在已知目标下优化
可解释性	重视因果关系	基于相关性,可解释性较差
自主性	自主性强	有限可控的自主性

法全面评价“智能体是否真正具有智能”。因此,从AI领域整体来看,尚未进入库恩定义的“常规科学”阶段。目前的AI是在尚无统一理论指导的前提下,帮助人类进行认知探索与问题求解的“通用前置方法”,这是一种用于探索“尚未被系统理解的领域”的通用方法。

1962年库恩在他的经典著作《The Structure of Scientific revolutions》(《科学革命的结构》)中提出“科学范式”概念,至今已经60多年了。随着科技的发展,科研范式的形式可能也会发生变化。人工智能可能会形成不同于库恩阐述的范式演进的新途径,人机协作就是人工智能带来的新科研范式。这种范式的特点是:在不确定的环境中,基于数据和模型,通过机器学习先找到某种潜在结构或经验规律;再由科学家解释、验证,构建普适理论;最后形成工程规范或学科定律。AlphaFold 2通过深度学习预测蛋白质结构,它本身并未提出生物化学新理论,但却为理论建构提供了前所未有的“结构猜测”。这种模式可能会成为今后科学研究的常态。需要指出的是,对机器学习的认知局限性要有清醒的认识,机器理性不可能完全替代人类理性。发现新规律需要灵感和顿悟,发现新材料、新药物等仍需要人类做大量艰苦细致的测试和临床实验。应当由人类完成的事不可能完全交给机器,自媒体上关于完全替代甚至征

服人类的所谓“硅基新物种”的种种“预言”,目前还只是一种科学幻想,无需过分关注。

长期以来人们普遍认为,采用演绎推理和简洁的公式表达知识是科学研究进入常规阶段的必要条件,试错性的搜索和统计推理是科学范式不成熟的表现。但是,人工智能技术的发展表明,数据驱动和概率推理可能会在科研工作中发挥越来越重要的作用,成为今后普遍采用的科研模式。人工智能的核心是应对不确定性、复杂性和知识不充分性,这一特点决定了AI研究是一种与以往任何科学研究不同的范式。现在的语言大模型有千亿级以上的权重参数,这些参数的统计分布可以看作人工智能的通用语言,结构相同的大模型有可能通过直接交换权重参数进行交流和学习,成为机器理性的一种新范式。之所以认为当前的人工智能还处在前科学范式阶段,是因为人工智能要走的路还很长,今后的人工智能是不是还是以统计推理为核心技术,现在还说不准。人工智能的范式仍在快速演进中,尚未完全固化。但有一点可以肯定,在相当长的时间内,人工智能是一个范式快速演进、技术不断拓展的学科,不要期望它很快成为像物理、化学一样稳定的“常规科学”。

人工智能还有一个与众不同的“自我消解”的特点,所谓“自我消解”是指,当一项人工智能技术被深入研究,形成了可靠、系统化

的规则、算法或知识库,并且其求解过程变得透明、可预测甚至“平凡化”之后,人们往往就不再认为它是“人工智能”了,而是将其归入某具体应用领域的常规技术或软件。图像处理、数据库、光学字符识别(OCR)和专家系统等,都曾经历或正在经历这个过程。这就是说,AI不是理论的终点,而是理论发现的起点。AI可能永远进不了库恩定义的“常规科学”的大门,而是永不毕业的“前范式科学伴侣”。

“人工智能究竟是一门什么学问”是一个永远开放的问题,已经问了几十年,今后还会问下去。老的疑惑解决了,新的疑惑又会冒出来。“不识庐山真面目,只缘身在此山中”,也许只有走出这座大山以后,才能真正看清人工智能的真面目。

参考文献(References)

- [1] McCulloch W S, Pitts W. A logical Calculus of the ideas immanent in nervous activity[J]. Bulletin of Mathematical Biology, 1990, 52(1/2): 99-115.
- [2] 李国杰. 智能化科研(AI4S)里程碑式重大成果综述[J]. 计算, 2025, 1(4): 6-15.
- [3] Russell S, Norvig P. Artificial intelligence: A modern approach[M]. 4th edition. London: Pearson, 2020.
- [4] Rittel H W J, Webber M M. Dilemmas in a general theory of planning [J]. Policy Sciences, 1973, 4(2): 155-169.
- [5] 李国杰. DeepSeek引发的AI发展路径思考[J]. 科技导报, 2025, 43(3): 14-19.

What kind of discipline is artificial intelligence exactly?

LI Guojie

Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

Abstract As a discipline with nearly 80 years of development, artificial intelligence (AI) has continuously expanded its boundaries and enriched its research directions. However, there remain controversies in the academic community regarding its disciplinary attributes, research core, and other fundamental issues. This paper conducted a systematic analysis of the disciplinary attributes of AI and examines its disciplinary positioning from multiple dimensions. It analyzed AI's four core goals—thinking like a human, acting like a human, thinking rationally, and acting rationally—and pointed out that AI is a methodology rather than merely a simulation of humans, serving as a new paradigm for solving complex problems. Moreover, AI has transcended the scope of traditional computer science subdisciplines. The paper emphasized that AI is a "meta-level exploration method" for science and technology, and it is currently still in the stage of "pre-paradigm science".

Keywords artificial intelligence (AI); disciplinary attributes; computer science; machine learning; scientific paradigm ●



(责任编辑 王丽娜)