陈小平，中国人工智能学会人工智能伦理与治理工委会主任，中国科学技术大学机器人实验室主任，广东省科学院人工智能首席科学家。曾任2015世界人工智能联合大会（IJCAI2015）机器人领域主席、2008和2015机器人世界杯及学术大会（RoboCup2008, 2015）主席、Journal of Artificial Intelligence Research和Knowledge Engineering Review编委。获中科大“杰出研究校长奖”、机器人世界杯冠军、最佳论文奖、行业年度十大科技进展及其他国内外学术荣誉20余项。

卷首语 Foreword

人工智能诺奖热中的反思与展望

陈小平

2024年度两个诺贝尔科学奖均颁给了人工智能及相关成果，引发了人工智能诺奖热。值此之际，人工智能研究者更需冷静反思本学科的发展状况、面临的挑战以及背后的基本问题。

获得本年度化学奖的成果AlphaFold2从氨基酸序列预测蛋白质的三维结构，能够预测几乎所有已确定的2亿种蛋白质结构，解决了化学界50多年来持续探索的一个重大问题。一般情况下，诺贝尔科学奖的两个必要条件是作出重要创新和解决本领域的重大问题。每个领域都有大量创新成果，但能解决重大问题的则如凤毛麟角。经70余年努力，用人工智能解决重大科学问题的可能性首次变成了现实，今后将出现更多类似成果。这是一个分水岭，标志着人工智能进入了一个新的发展阶段。

原理模拟观认为，人工智能是用人工方法模拟人类智能的工作原理。然而迄今对人类记忆、学习、推理、决策、理解等认知机制的认识远不足以支撑诺贝尔科学奖的成果，AlphaFold2的成功主要不是依靠对人类认知机制的模拟。图灵1948年提出的机器智能观认为，机器智能的工作原理与人类智能的工作原理可以不同。这一观点可谓石破天惊，开创了机器智能超越人类智能之思想先河。与人类相比，AlphaFold2用更短的时间、更低的成本和更高的精度，在更大范围内实现了蛋白质结构预测，是图灵机器智能观在科学研究领域中取得的一场重大胜利，具有划时代意义。

那么，如此强大的机器智能到底是按什么原理工作的？这些原理是否突破了“智能”的传统边界？能否、如何科学地把握这些原理？这是人工智能面临的科学基础挑战。这一挑战在大模型研究中体现得尤为明显，大量深度测试揭示了大模型诸多无法解释的奇异表现。我在《智能系统学报》2023年第4期和《中国人工智能学会通讯》2024年第1期的文章中，提出了刻画大模型底层原理的一种科学假说——类L_C理论，得到了越来越多深度测试的验证。根据类L_C理论，大模型隐含着两项原理性突破——实例性和弱共识性，而传统的自然科学理论、数学、计算机科学和人工智能强力法都基于概括性和强共识性。这表明，人工智能已经突破了科学技术的传统边界，进入了无限广阔、亘古未有的“无人区”。

人工智能的下一个战略目标是解决产业的重大问题，比如为某个行业的智能化升级提供关键共性技术。在工业、农业、交通运输等所有实体经济部门，人工智能应用通常不是资料性应用，而是技术成果在物理世界中自主或半自主的运行，因此不再局限于信息处理，相应的人工智能研究也不能囿于信息空间，从而为人工智能思想的新一轮深刻变革提供历史性机遇。中国人工智能学会名誉理事长李德毅院士等中外学者积极发展的“认知物理学”，为打通信息空间和物理空间中的人工智能提供了一条新思路。

机器智能相对于人类智能的另类性决定了一定条件下相对于人类智能的超越性，于是也难免带来伦理和安全的潜在风险。Hinton表示：授予我物理学奖的最大作用是，让人们重视我对潜在风险的警示。但愿他的警示能引起更大关注。但与风险防范相比，更重要的是驾驭机器智能对经济、社会和文化具有的巨大重塑力，确保人工智能为人类塑造一个更加美好的未来。这是人类面临的最大课题，是全社会的共同责任，更是人工智能工作者的历史使命。