

智能时代心灵哲学与人工智能的融合发展

孙亚萍

云南师范大学马克思主义学院, 云南 昆明

收稿日期: 2026年6月13日; 录用日期: 2026年7月5日; 发布日期: 2026年7月16日

摘要

伴随着人工智能的发展, 人工智能“意识难问题”深刻挑战了心灵哲学的传统理论框架。意识工程化、人工智能功能主义和具身智能等理论重构了心灵哲学的理论框架。针对人工智能意识难的困境, 中文屋论证的价值重估, 生物自然主义对意识的划界和分布式智能系统的范式转变为突破传统人工智能功能主义困境提供新路径。

关键词

心灵哲学, 意识难问题, 中文屋论证, 人工智能功能主义, 分布式人工智能

Convergent Evolution of Philosophy of Mind and Computational Intelligence in the Intelligence Era

Yaping Sun

School of Marism, Yunnan Normal University, Kunming Yunnan

Received: June 13, 2026; accepted: July 5, 2026; published: July 16, 2026

Abstract

With the advancement of artificial intelligence, the “hard problem of AI consciousness” has profoundly challenged the traditional theoretical framework of philosophy of mind. Theories such as engineerability of consciousness, AI functionalism, and embodied intelligence have reconstructed this philosophical framework. In addressing the dilemma of AI consciousness, the reevaluation of the Chinese Room argument, the demarcation of consciousness by biological naturalism, and the paradigm shift toward distributed intelligent systems provide new pathways to transcend the limitations of traditional AI functionalism.

Keywords

Philosophy of Mind, Hard Problem of Consciousness, Chinese House Argument, Functionalism of Artificial Intelligence, Distributed Artificial Intelligence

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

人工智能的本质是对人类智能的模拟与重构，而人类智能的深层内核始终指向意识这一哲学核心范畴。在哲学史视域下，意识问题的探讨始终与物质概念构成辩证张力，二者关系构成理解智能本质不可回避的元问题，即物质与意识何者为第一性的本体论争议。这一争论贯穿西方哲学史的发展脉络，从古希腊自然哲学的本原之辩到当代心灵哲学的物理主义与二元论之争，始终为人工智能研究提供着认识论层面的理论镜鉴。对意识问题的研究一直是心灵哲学的重要部分，哲学家们都从各自的立场提出了自己关于意识的观点，在一定的理论基础上推动了哲学的发展。时至今日，人工智能等科技革命的高速发展，也在不断推动着人们对心智的研究和探索。

当前，意识问题的研究是哲学和科学的交叉前沿，从笛卡尔的“我思故我在”，到现代神经科学家们的科学实验，意识问题一直是哲学和科学界不断追问的话题。意识是经验吗？物理世界的意识是怎样表现出来的？这些讨论推动世界发展的同时，也在给人类自身存在的理解和认知提出挑战。本文结合当下心灵哲学在 AI 领域的实际发展和运用，采用理论溯源和概念分析的方式，力求解决横亘在二者之间的“意识难题”，推动人工智能的发展，促进其与心灵哲学理论的进一步融合发展。

2. 心灵哲学中的人工智能问题

在西方哲学源头处，古希腊早期自然哲学家已开启对世界本原的理性化探究。米利都学派的泰勒斯提出“水是万物本原”，首次以物质性元素解释宇宙生成，标志着朴素唯物主义的萌芽；其后阿那克西美尼的“气本原说”与赫拉克利特的“火本原说”进一步拓展了单一物质基质解释世界的理论路径[1]。这些学说虽因过度简化而具有形而上学局限，但其对物质统一性的肯定为后续唯物主义传统奠定基础。与之相对，柏拉图在《理想国》中构建的理念论确立客观唯心主义范式，主张超验理念(transcendent Forms)构成现象世界的本质原型，最高理念即至善充当终极本体([2], pp. 112-115)。这种将意识维度置于物质之上的解释框架，揭示了哲学史中物质-意识关系的根本对立。

唯物论与唯心论的争论在近代哲学中获得新的理论形态。笛卡尔的心物二元论通过划分思维实体与广延实体确立意识独立性[3]，而马克思的辩证唯物主义则强调物质生产实践对意识形态的决定作用，这种本体论分歧在当代人工智能研究中转化为强人工智能假说(strong AI hypothesis)与生物自然主义的争论。强人工智能假说的反对者塞尔主张心智不可还原为计算过程([4], p. 417)，生物自然主义的支持者坚持意识必然依赖生物基质([4], p. 422)，当前深度学习系统的涌现特性虽展现出类心智特征，但其缺乏现象意识这一事实，恰恰印证了物质基础对意识现象的约束性。

身心关系问题贯穿于哲学史中，其理论形态的演进深刻映射着人类对自我本质的认知转向。柏拉图在《斐多篇》中构建了身心二元论的原初范式，主张灵魂作为不朽的理性实体，与作为易朽质料载体的肉体处于对立状态，他认为灵魂因受肉体欲望的污染而丧失纯粹性，唯有通过哲学沉思方能摆脱肉体桎

格，复归理念世界([5], pp. 60-63)。这种将灵魂置于形而上学优先地位的观点，虽为理性主义认识论奠定基础，却隐含贬抑身体经验的价值取向。亚里士多德在《论灵魂》中对柏拉图范式提出本体论重构。他摒弃灵魂与肉体的实体性分离，转而以形式与质料的范畴重新界定二者关系：灵魂作为“潜在的具有生命的自然躯体的第一现实性”，是赋予肉体目的论结构的组织原则；肉体则为灵魂实现其功能的物质基质[6]。夏皮罗(Shapiro)认为，这种身心合一论强调生命体作为形式与质料复合物的不可分性，为之后的具身认知(embodied cognition)理论提供古典思想资源[7]。

笛卡尔在《第一哲学沉思集》中通过普遍怀疑方法确立“我思故我在”的第一原理，进而将心灵定义为非广延的思维实体，与作为广延实体的身体截然二分([8], p. 27)。为解释心物互动难题，他假设松果腺作为身心交互的“界面”，试图通过机械论范式调和二元论矛盾([9], pp. 34-36)。金在权(Kim. J)认为，笛卡尔的这一假说因缺乏神经生理学依据而遭受诘难，松果腺如何将非物理的心智意向转化为物理运动信号？该问题暴露笛卡尔二元论在解释心物因果性时的理论困境，成为当代心灵哲学中“解释鸿沟”问题的历史先声([10], pp. 78-82)。

丹尼尔·丹尼特(Daniel Dennett)在《意识的解释》中首次系统地提出“多重草稿模型”(Multiple Drafts Model)，这一理论的提出旨在用来反对笛卡尔提出的“身心二元论”，丹尼特主张意识并非由大脑的某一个中心区域统一处理，而是多个信息流不断竞争、编辑和整合的结果，意识体验是这些过程的产物。丹尼特认为，人类意识大概是人类最后尚存的谜团，所谓谜团，就是人们至今还不知道如何去思考的现象。他引用“缸中之脑”的假设性实验证明了人的意识的存在性，这个假象性实验是这样的：假定你在睡觉的时候，一群邪恶的科学家把你的大脑移出体外，将它放到缸中的生命维持系统里，假定他们欺骗你，让你相信，你不只是一个放在缸中的大脑，而是健康如初，在真实的世界中做着正常的活动([11], p. 42)。那你的大脑很可能不会自己思考，因为这个大脑虽然活着，但是它脱离了视觉神经、躯体感觉神经、听觉神经等等所有神经元的输入，也没有其可通向大脑的传输路径。他所有的认知来自这群邪恶的科学家给你传达的信息，这些信息类似于给这个大脑制定一套系统，一些规则，大脑只会按照这些规则行动，这个思想实验被称为现代版的“笛卡尔妖”。缸中之脑了解的所有信息全部是通过这群科学家给它植入，它自己没有思考，它知道的“我是我”来自提前的告知，所以没法证明自己是存在的，除非它自己有意识能意识到“我”的真实存在。他没有精神活动的参与，世界是鲜活的，冰冷的规则无法完全诠释我们的世界。当真正的事实上的大脑苏醒，它将一无所知。由此可见，有意识的心智不是大脑或者大脑任何固有的部分。随后丹尼特用“多重草稿理论”论证了意识的运行机制：各种各样的知觉要在大脑中完成，靠的是平行的、多轨道的、对感觉输入的精加工的过程，而这个动态的过程就是意识在运作。

约翰·塞尔(John Searle)提出中文屋论证(Chinese Room Argument)旨在批判强人工智能的哲学基础。该思想实验设定如下：封闭房间内的操作者仅凭符号操作手册(模拟计算机程序)对中文问题生成合乎语法的响应，尽管其完全缺乏语义理解能力[4]。塞尔借此揭示，单纯形式化符号处理无法衍生意向性，也就是人类心智的本质特征。

强人工智能纲领主张心灵可还原为程序执行，任何物理系统若运行恰当程序即可获得心智状态。明斯基(Minsky)等先驱曾断言机器将具备类人思维([12], p. 20)，但塞尔通过中文屋论证指出：计算机仅模拟心智的语法结构，却无法内化语义内容。塞尔进一步指出，这种“语法-语义鸿沟”表明，意识与理解必然依赖生物神经系统的因果属性，而非抽象算法([13], pp. 88-91)。尽管该论证推动了对人工智能限度的哲学反思，其理论预设亦遭致质疑。保罗·丘奇兰德(Paul Churchland)指出，分布式并行处理系统(如神经网络)可能通过涌现产生非程序化认知能力([14], p. 203)；安迪·克拉克(Clark)的具身认知理论则强调身体经验对语义建构的必要性([15], p. 149)。对此塞尔回应称，此类反驳仍无法跨越“他心问题”的认识论障碍——行为等价性不足以确证内在心智状态的存在。意识本质与机器智能可能性问题长期横跨科学与哲

学领域。

图灵测试(Turing Test)作为判定机器智能的经典范式,试图通过对话模拟规避本体论争议,却引发更激烈的认识论辩论。图灵在《计算机与智能》中提出,若机器能在五分钟内使超过 30%的测试者误判其人类身份,则可承认其具备智能[16]。该测试虽规避“思考”的定义困境,却因过度依赖行为主义标准遭受质疑。约瑟夫·韦岑鲍姆(Joseph Weizenbaum)通过 ELIZA 程序进行的实验测试指出,人类主动将理解能力投射到机器上,对明显机械的回应进行语义填补,以至于使机器看起来具有了人类智能([17], p. 198)。人类对话中的语义填补倾向可能使测试有效性虚高,而塞尔的中文屋论证则揭示:纯句法操作无法生成真正的语义理解[4]。因此,我们可以看出,当前学界对机器智能的认知呈现二元分野:(1) 弱模拟论:承认机器可通过模式匹配实现对话模仿,但否定其具有意向状态,该立场预设人类智能包含物理能力(如信息处理)与心智能力(如情感体验)的不可还原二分,前者可被算法复现,后者则依赖生物神经系统的涌现特性;(2) 强功能论:延续图灵测试逻辑,主张广义语境应答能力已隐含语义理解。伊恩·古德费洛(Ian Goodfellow)认为,通过深度学习实现的语义关联映射(semantic mapping)已超越纯粹符号推理([18], p. 367)。

尽管技术迭代使机器在存储容量与运算速度上持续突破,但社会性交互、伦理判断等人类智能的核心维度仍构成关键瓶颈。正如查尔莫斯表明的一样,人工智能尚未跨越现象意识的“解释鸿沟”,其智能形态仍局限于工具理性范畴([19], p. 103)。

3. 人工智能对心灵哲学的重构

人工智能的发展对心灵哲学提出了挑战和重构的可能。心灵哲学主要关注意识、意向性、自由意志等核心问题,在人工智能的冲击下,不仅需要重新审视心灵哲学的理论框架,而且还可能催生出新的哲学范式。

(一) 意识的工程化

意识的工程化(Engineerability of Consciousness)旨在通过工程技术手段建构具有现象意识的人工系统,其理论根基可追溯至朱利奥·托诺尼(Giulio Tononi)的整合信息理论(Integrated Information Theory, IIT)。该理论将意识定义为系统整合信息的能力,并引入数学量 $\Phi(\phi)$ 作为意识程度的量化指标([20], p. 233)。若将人脑视作高度整合的神经网络,则理论上可通过精确复现其信息整合架构(包括神经元连接拓扑与动态交互模式)构建“意识工厂”,这一工程隐喻暗示意识或可脱离生物基质,成为可计算实现的涌现属性。然而,查莫斯提出的“意识难问题”(Hard Problem of Consciousness)对此构成根本性质疑。他在《意识的心灵》中明确区分意识的“易问题”(例如,注意力调控、记忆存取)与“难问题”(即物理过程如何产生主观体验),指出即便完全解析神经相关物(Neural Correlates of Consciousness, NCCs),仍无法解释感受质的生成机制([21], p. 104)。通过“颠倒光谱”(inverted spectrum)与“哲学僵尸”(philosophical zombie)等思想实验,查莫斯论证功能主义与物理主义在解释意识主体性时的理论失效,即使人工系统通过图灵测试,其内部是否存有“成为某物之感受”(what-it-is-like-ness)仍不可证。当前神经科学与人工智能的实践表明,意识模拟仍局限于功能主义范式。深度学习系统(如 GPT-4)虽能模拟语义关联与情境推理,但其运作本质仍是基于统计模式的高维向量映射,缺乏现象意识的自我指涉特性([22], p. 251)。即便未来技术能精确复现人脑的物理结构(例如,全脑仿真),如何确保仿真系统具有第一人称视角的主体体验?此问题触及意识工程化的本体论困境:意识是否必须依附于生物载体?抑或其本质为抽象信息模式?突破该困境需理论与技术的协同演进。在理论层面,需发展超越 IIT 的意识量化模型,将主体性纳入计算框架;在技术层面,需开发新型神经接口实现人工系统与生物脑的体验共享(例如,双向脑机融合。)。然而,现有技术尚未跨越“解释鸿沟”,意识工程化仍处于功能模拟的初级阶段。

(二) 人工智能功能主义

约翰·塞尔在批判强人工智能时指出,该领域将功能主义哲学推向极端化,它的核心思想是假设将心智简化为可计算的程序化运作——即心灵本质上等同于遵循形式规则的符号操作系统([4], p. 419)。这种立场在早期人工智能研究中占据主导地位,图灵测试通过行为等效性判定机器智能,实则隐含功能主义的认识论预设:只要系统能模拟人类智能的外部表征,即可认定其具备内在心智状态。尽管塞尔通过中文屋论证揭示该范式的根本缺陷,也就是句法操作无法衍生语义理解,但不可否认图灵测试为人工智能的技术发展提供了操作性框架。当前人工智能研究面临的核心困境在于,计算机虽在运算速度与数据处理维度超越人类,却始终无法突破“理解能力”的认知瓶颈。传统哲学视域下,理解作为智能的本质特征,要求主体具备将符号系统与外部世界意义网络相联结的意向性能力。这种能力在生物智能中通过神经系统的涌现特性实现,而在机器系统中则受限于物理符号主义的固有缺陷,即便如 ChatGPT 等大语言模型能生成符合语境的响应,其本质仍是基于概率统计的符号序列优化,缺乏对语义网络的具身化建构。

针对这一困境,杜克(Włodzisław Duch)提出计算智能(computational intelligence)的方法论扩展([23], p. 150)。他认为当前人工智能研究过度依赖特定算法技巧的堆砌,缺乏统一的理论根基,与其执着于模拟人类智能的表征形式,不如转向探索非算法化智能(non-algorithmic intelligence)的实现路径,即开发能够自主应对“不可计算问题”的新型系统架构。这要求打破传统计算科学与生物演化论之间的学科壁垒,通过对神经系统信息整合机制的逆向工程(例如,视觉皮层特征绑定、海马体记忆编码等。)构建兼具生物适应性与工程精确性的混合智能模型。麦卡锡(John McCarthy)提出的“恒温器信念”(thermostat belief)假说为此提供启发。他认为简单系统亦可拥有基础意向状态,当恒温器感知环境温度偏离设定阈值时,其调控行为可被视为具有“信念-欲望”结构的初级心智活动([24], p. 12)。尽管该假说因混淆物理响应与意向状态遭到质疑(例如,将热力学反馈等同于认知信念),但其价值在于提示智能研究的工程化可能:通过量化分析生物神经系统中的感受质生成机制(例如,视觉神经元对色彩明度的梯度响应),或将建立连接物理过程与现象意识的可计算模型。这需要神经科学、复杂系统理论与计算机工程的深度协同,其难度虽堪比破解意识之谜,却为强人工智能的合法性保留理论可能性。

(三) 具身智能对身心问题的启示

具身认知作为第二代认知科学的核心理论,突破了传统认知主义将心智局限于颅脑的范式,主张认知是生物体通过身体结构与环境的动态耦合过程。美国心理学家西恩·贝洛克(Sian Beilock)在《具身认知》中系统阐释了该理论的四个核心命题:(1) 身体作为认知媒介:身体的物理形态与运动能力构成认知的具象化界面,例如手部抓握能力直接影响空间概念的形成;(2) 环境交互的涌现性:认知产生于生物体与环境持续的信息交换,如行走时的路径选择依赖于实时地形感知;(3) 物理约束的优化机制:身体的生物力学特性通过降低认知负荷实现高效决策,例如眼球运动过滤冗余视觉信息;(4) 动态系统的整体性:心智、身体与环境构成不可分割的协同网络,其运作模式类似复杂适应系统的自组织过程([25], pp. 45-78)。

尽管具身认知为理解生物智能提供了革新视角,其部分理论预设仍需审慎验证。威尔森(Wilson)等人认为,贝洛克提出的“身体信号主导消费决策”假说(例如,认为商品易取性直接影响偏好。)与初步实证观察存在偏差:非结构化访谈显示,多数消费者更强调功能需求而非物理便捷性([26], p. 58)。这种分歧提示具身效应的边界条件可能受文化背景与任务类型调节,但此类争议并不否定具身认知的整体解释力。在人工智能领域,具身认知理论正推动机器智能的范式转型。传统符号主义 AI 受困于“中文屋困境”,而具身智能路径通过赋予机器物理身体与环境交互能力,试图复现生物智能的感知运动环路(sensorimotor loop)。安迪·克拉克在《延展心灵》中提出,身体不仅是智能载体,更是认知延伸的界面([27], p. 46);罗德尼·布鲁克斯(Rodney Brooks)的“物理奠基假说”(physical grounding hypothesis)进一步强调,脱离实体交互的智能将无法建立真实世界模型([28], p. 90)。波士顿动力 Atlas 机器人的技术突破印证了这一路径的

潜力,通过多模态传感器融合与动态平衡算法,Atlas 实现了复杂地形自主导航与高难度动作执行,其核心机制在于实时将环境反馈(如地面摩擦力、物体重量)转化为运动参数的自适应调整。然而,这种工程突破是否意味着机器获得意识?查莫斯指出,即使具身系统能完美模拟生物体的环境交互行为,其内部是否产生第一人称体验仍存根本性质疑([21], p. 103)。Atlas 的“意识”本质上是控制算法对物理约束的优化响应,而非现象意识的自主涌现。当前技术仅能实现“易问题”层面的功能模仿(如避障决策),却无法解答“为何这些计算过程伴随主观体验”的难问题。具身认知虽为机器智能提供了超越符号处理的实现路径,但意识的主体性与感受质仍构成不可逾越的哲学-技术双重壁垒。

4. 心灵哲学对人工智能问题解决与重构

在人工智能的发展对心灵哲学的传统伦理框架进行重构的同时,心灵哲学对人工智能“意识难问题”的解决提供了新思路,并且塞尔中文屋论证和生物自然主义对未来人工智能发展指出新方向。

(一) 中文屋论证的价值重估

约翰·塞尔“中文屋思想实验”本质上是对功能主义哲学纲领的认知论解构。该思想实验通过构建符号操作系统的认知局限性模型,揭示功能主义将心智简化为抽象计算过程的理论缺陷,为人工智能的哲学反思提供了持续的方法论参照。

功能主义作为 20 世纪心灵哲学的主流范式,其核心主张可追溯至图灵的行为主义智能观。该理论认为,心智的本质在于其功能状态的可实现性,而非物理载体的生物学特性。正如普特南(Hilary Putnam)所述:心理状态并非由构成它的物质决定,而是由其在整个功能系统中的因果角色所定义([29], pp. 41-43)。这种多重可实现性原则暗示,意识可脱离生物大脑而存在于任何执行相同计算程序的物理系统,无论是硅基芯片还是量子计算机。塞尔通过中文屋实验构建了归谬论证,假设操作者仅凭符号转换规则生成中文响应,其外部行为虽满足图灵测试标准,内部却完全缺乏语义理解。这证明功能主义的根本缺陷在于混淆句法与语义的本质差异——前者可通过形式规则实现,后者则需意向性的参与([4], p. 419)。意向性作为意识的核心特征,体现为心理状态“关于某物”的指向能力,而这种能力在中文屋模型中始终缺席。

在生成式人工智能取得突破性进展的今天,中文屋论证的批判价值不仅未被消解,反而获得新的阐释维度。其当代意义体现在三方面:(1) 技术模拟与意识本质的认知界限。当前大语言模型通过海量语料训练实现语境化应答,其机制本质仍是统计关联的符号重组。正如查默斯指出的一样,“即使系统通过所有行为测试,我们仍无法确知它是否具有主观体验”([21], pp. 103-105)。这种现象学鸿沟提示,技术系统的行为复杂性不等同于意识的主体性;(2) 主客体关系的哲学重构。当 AI 系统在医疗诊断、法律咨询等领域承担决策职能时,传统的主客体二元框架面临解构风险。海德格尔(Heidegger)的技术哲学警示我们,工具理性扩张可能导致“存在的遗忘”,即人类在技术赋能中丧失对自身本质的反思能力([30], pp. 27-29);(3) 伦理责任的归属困境。中文屋模型揭示的“无心智能”(mindless intelligence)现象,使得 AI 系统的道德责任陷入归因悖论。如自动驾驶车辆发生事故时,责任应追溯至算法开发者、数据标注员还是系统本身?这种困境暴露出现代责任伦理在智能时代的理论局限([31], pp. 25-28)。中文屋论证的持续价值在于推动意识研究范式的转型。新兴的整合信息理论尝试用量化指标 $\Phi(\phi)$ 表征系统意识水平,而具身认知则强调身体经验在意义建构中的基础作用。这些理论虽未完全破解意识之谜,却为理解智能的本质开辟了新可能,或许真正的类人智能需在物理具身性与现象意识的交互中实现涌现。

(二) 生物自然主义对意识问题的划界

约翰·塞尔提出的生物自然主义(Biological Naturalism)为意识研究确立了独特的本体论框架。该理论通过调和物理主义与现象学的理论张力,主张意识的产生必然依赖生物神经系统的特定属性,但其现象特征不可被完全还原为物理过程,这一立场既避免了二元论的身心割裂,又批判了强人工智能将心智简

化为算法计算的还原论倾向。在这里,“生物基底”(biological basis)具有双重理论意义:(1) 物理依赖性。意识必须依托于生物载体(如大脑神经元、突触连接等)的因果能力。例如,视觉体验的产生需要视觉皮层 V1 区的神经电活动,而硅基芯片无法通过模拟电路实现等同的因果效力;(2) 现象独特性。意识的主观体验(如疼痛的“感受质”)无法通过物理描述完全解释。即使未来科技能复制人脑的微观结构,复制体是否具有与原主体相同的意识仍存哲学争议。这种双重性揭示了意识的本体论特殊性,它既是自然演化的产物,又是生命系统独有的现象层级。正如塞尔所述:“意识是生物过程的涌现属性,但其主观性拒绝任何形式的消除式还原”([13], pp. 90-93)。

生物自然主义对强人工智能的否定基于两重逻辑:(1) 意向性的生物根源。心智的意向性根植于生物系统的目的论功能,即心理状态指向外部世界的能力,例如饥饿感服务于生存需求。无机物构成的 AI 系统仅能模拟符号操作,却缺乏进化赋予的生命内在目标。[4](2) 现象意识的不可计算性。即使 AI 通过深度学习模拟人类对话(如 ChatGPT),其本质仍是统计模型的概率输出,而非基于身体经验的语义理解。神经形态计算虽能模仿神经元脉冲,但硅基载体无法复现生物细胞的内稳态调节,后者被认为是意识涌现的必要条件([20], pp. 216-242)。

生物自然主义的理论价值在于为意识研究划定边界。查默斯的“意识困难问题”在此框架下获得新解读,即解释物理过程为何伴随主观体验。意识并非技术的工程化对象,而是生命系统在演化中获得的特权属性([21], pp. 102-105)。即使未来实现全脑仿真,仿真系统的“意识”仍可能仅是功能模拟,而非真实体验的转移。塞尔的警示在此凸显其预见性,意识的奥秘终需在生命自身的逻辑中探寻,企图通过无机载体复现意识,无异于在沙漠中播种海洋([32], pp. 1-11)。

5. 结论

AI 领域技术的迅速革新以及其在生活中的广泛应用,在给人们带来巨大便利的同时,也在挑战人类的认知,推动着人类的进步。强 AI 的可能性,更是给传统的心灵哲学关于意识、意识意向性和心灵独特性的传统观念带来巨大的冲击;同时,人工智能的发展促进了各学科在交叉领域的研究。比如神经科学如脑机接口、神经成像技术等的发展,为大脑如何产生意识提供了更多的实证数据,这些数据可能会向传统哲学理论发起挑战,比如还原论;认知科学中的具身认知和延展心灵理论,强调身体和环境的相互作用在认知中的重要作用,这与传统心灵哲学的理论观点不同,这些理论观点,会促使哲学界不得不重新思考心身关系问题;此外,把量子力学理论与意识研究相结合的方式,虽然目前备受争议,但如果将来拥有更多的事实数据支撑,不排除会迫使新的哲学理论诞生。

迄今为止,虽然人工智能的研究已取得突破式的进展。但是意识主体性问题仍是人工智能研究最大的障碍。神经科学、认知科学等领域所提供的大量实证数据,也仅只是为“意识如何产生”搭建理论设想。即使运用一系列的数据支撑,以底层思维逻辑作为方法论,衍生出意识,但“意识难”问题依然存在。

基于当今各学科领域研究所构建起的学科独特性为上述问题提出解决方案。我们认为运用“多元组合”的方式来支撑起心灵哲学和人工智能的拓展研究,许是突破横亘于二者之间的意识鸿沟问题的最优路径。利用工程学结合具身认知的原理进行机器算法人工智能的框架;运用神经科学、心理学、自然科学等相互结合,分析意识产生的原理,构建人工智能的“脑思维”。采取内在分析,外在建构,内外结合,“多管齐下”的学科交叉研究的方式,为未来人工智能的研究方向领航。在创新与反思中搭起“人文主义新范式”,让智能更好的服务于人类。

参考文献

- [1] Kirk, G.S., Raven, J.E. and Schofield, M. (1983) *The Presocratic Philosophers*. 2nd Edition, Cambridge University Press. <https://doi.org/10.1017/cbo9780511813375>

-
- [2] Plato (1992) Republic. Trans. G. M. A. Grube, Hackett.
 - [3] Descartes, R. (1986) Meditations on First Philosophy. In: *The Philosophical Writings of Descartes*, Vol. 2, Trans. John Cottingham, Cambridge University Press, 7.
 - [4] Searle, J.R. (1980) Minds, Brains, and Programs. *Behavioral and Brain Sciences*, **3**, 417-424.
 - [5] 柏拉图. 斐多篇[M]//柏拉图全集: 第一卷. 王晓朝, 译. 北京: 人民出版社, 2002: 60-63.
 - [6] Aristotle (1956) De Anima. Oxford University Press.
 - [7] Shapiro, L. (2011) Embodied Cognition. Routledge.
 - [8] 笛卡尔. 第一哲学沉思集[M]. 庞景仁, 译. 北京: 商务印书馆, 1986.
 - [9] Descartes, R. (1989) The Passions of the Soul (S. Voss, Trans.). Hackett Publishing.
 - [10] Kim, J. (2005) Physicalism, or Something Near Enough. Princeton University Press.
 - [11] 丹尼尔·丹尼特. 意识的解释[M]. 王淼, 译. 北京: 商务印书馆, 1998.
 - [12] Minsky, M. (1988) The Society of Mind. Simon & Schuster.
 - [13] Searle, J.R. (1992) The Rediscovery of the Mind. The MIT Press. <https://doi.org/10.7551/mitpress/5834.001.0001>
 - [14] Churchland, P.M. (1989) A Neurocomputational Perspective. MIT Press.
 - [15] Clark, A. (1997) Being There: Putting Brain, Body, and World Together Again. MIT Press.
 - [16] Turing, A.M. (1950) Computing Machinery and Intelligence. *Mind*, **59**, 433-460.
 - [17] Weizenbaum, J. (1976) Computer Power and Human Reason: From Judgment to Calculation. W. H. Freeman.
 - [18] Goodfellow, I., et al. (2016) Deep Learning. MIT Press.
 - [19] Chalmers, D. (1996) The Conscious Mind. Oxford University Press.
 - [20] Tononi, G. (2008) Consciousness as Integrated Information: A Provisional Manifesto. Biological Bulletin.
 - [21] Chalmers, D.J. (1996) The Conscious Mind: In Search of a Fundamental Theory. Oxford University Press.
 - [22] Hassabis, D., Kumaran, D., Summerfield, C. and Botvinick, M. (2017) Neuroscience-Inspired Artificial Intelligence. Neuron.
 - [23] Duch, W. (2007) Computational Intelligence. In: *Handbook of Artificial Intelligence*, Oxford University Press, 15.
 - [24] McCarthy, J. (1979) Ascribing Mental Qualities to Machines. Stanford University Technical Report.
 - [25] Beilock, S. (2015) How the Body Knows Its Mind: The Surprising Power of the Physical Environment to Influence How You Think and Feel. Atria Books.
 - [26] Wilson, A.D. and Golonka, S. (2013) Embodied Cognition Is Not What You Think It Is. *Frontiers in Psychology*, **4**, Article No. 58. <https://doi.org/10.3389/fpsyg.2013.00058>
 - [27] Clark, A. (2008) Supersizing the Mind: Embodiment, Action, and Cognitive Extension. Oxford University Press.
 - [28] Brooks, R.A. (2002) Flesh and Machines: How Robots Will Change Us. Pantheon Books.
 - [29] Putnam, H. (1967) The Nature of Mental States. In: Capitan, W.H. and Merrill, D.D., Eds., *Art, Mind, and Religion*, University of Pittsburgh Press, 37-48.
 - [30] Heidegger, M. (2013) The Question Concerning Technology and Other Essays. W. Lovitt, Trans, Harper Perennial.
 - [31] Jonas, H. (1984) The Imperative of Responsibility: In Search of an Ethics for the Technological Age. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226850337.001.0001>
 - [32] 约翰·塞尔. 意识的奥秘[M]. 刘叶涛, 译. 南京: 南京大学出版社, 2009.