

深度学习的技术路径与哲学审视

杨 媛

华南师范大学马克思主义学院, 广东 广州

收稿日期: 2025年5月1日; 录用日期: 2025年5月23日; 发布日期: 2025年5月31日

摘 要

作为人工智能领域的核心技术之一, 深度学习在近年来的研究和应用层面都取得了显著进展。本文从技术演进、学习范式对深度学习进行哲学角度审视。首先梳理深度学习的发展脉络, 然后深入讨论了深度学习的学习范式, 探索了深度学习面临的现实问题。深度学习作为一种强大的机器学习方法, 引发了许多哲学问题的关注与思考。通过对有关深度学习哲学思考的文献进行梳理, 思考如何使用哲学对深度学习技术进行指导, 同时反思深度学习技术又如何影响现代哲学。

关键词

深度学习, 人工智能, 技术哲学

Technical Path and Philosophical Review of Deep Learning

Yuan Yang

School of Marxism, South China Normal University, Guangzhou Guangdong

Received: May 1st, 2025; accepted: May 23rd, 2025; published: May 31st, 2025

Abstract

As one of the core technologies in the field of artificial intelligence, deep learning has made significant progress in research and application in recent years. This paper aims to examine deep learning from the perspective of technology evolution and learning paradigm. Firstly, the development of deep learning is sorted out, then the learning paradigm of deep learning is discussed, and the practical problems faced by deep learning are explored. As a powerful machine learning method, deep learning has aroused the attention and thinking of many philosophical issues. By combing the literature on philosophical thinking of deep learning, this paper considers how to use philosophy

to guide deep learning technology and reflects on how deep learning technology affects modern philosophy.

Keywords

Deep Learning, Artificial Intelligence, Philosophy of Technology

Copyright © 2025 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

当前，深度学习正聚焦着全球目光，各国纷纷制定并推行人工智能发展规划，力图在未来科技竞争中占据优势地位。但是，随着深度学习的高速发展，它所面临的问题也越来越多。很多学者开始对深度学习进行反思，通过对相关哲学观点和最新研究成果的综合分析，展示深度学习与哲学的交叉点，为进一步探索深度学习的哲学意义提供启示。

2. 深度学习的技术演进与哲学意蕴

(一) 深度学习的发展脉络

深度学习的发展经历了感知机的失败、反向传播算法与万能近似定理的突破以及深度信念网络的提出。

深度学习的技术渊源可追溯至早期的感知机，它是一种可以模拟人类感知能力的机器。然而，以感知机为代表的第一次神经网络浪潮很快就被证明存在局限，导致了神经网络研究的第一次低谷期。后来，反向传播算法和万能近似定理被提出。反向传播算法是一种适用于多层感知器的学习算法，万能近似定理则证明了多层感知器的表达能力和图灵机是等价的，消除了对神经网络表达能力的质疑，推动了第二次神经网络研究热潮。然而，由于高维非线性模型的复杂性，神经网络在训练过程中遇到了梯度消失、运行速度慢和泛化困难等问题，再次陷入低谷。为突破这一困境，深度信念网络的快速学习算法应运而生。这种方法通过逐层预训练和最终微调的方式，成功实现了多层神经网络的有效训练。这一突破不仅打破了深度神经网络训练的难题，更推动了深度学习第三次热潮的兴起。

(二) 深度学习的学习范式与认识论取向

研究深度学习的哲学基础，首先要厘清深度学习在人工智能领域中的定位。目前，学界对人工智能的分类并没有达成一个统一的共识。依据 UCLA 统计学系与计算机系朱松纯教授的划分，人工智能主要包含计算机视觉、机器人学、机器学习等六大方向。在此分类体系中，深度学习是隶属于人工智能研究领域中的一种机器学习技术，而机器学习本身又可被视为人工智能的基础方法论。机器学习的学习技术是由数据驱动的，数据是机器学习的一个核心要素。根据所处理数据种类的不同，机器学习主要分为四种范式：监督学习、无监督学习、半监督学习以及强化学习。深度学习作为机器学习的重要分支，同样具有这些学习范式。

第一种主要学习范式为监督学习，即通过学习带有标签的数据来发现其中规律的机器学习策略。这种学习策略体现了实证主义认识论观点，即认为知识来源于可观察的经验证据。然而，监督学习高度依赖人工标注数据，这一过程不仅耗时费力，还可能存在主观偏差，暴露了“观察负载理论”的科学哲学

困境。数据的意义并非中立客观，而是由预先设定的标签体系所决定。

无监督学习是机器学习的另一重要范式。无监督学习与监督学习的区别在于数据采样过程中是否对样本进行了标签化处理。无监督学习是通过学习未标注数据来找到其中的隐藏规律，这与建构主义认识论相呼应。然而，无监督学习因缺乏明确标签的指导，难以保证学习结果的准确性。

第三种学习范式为半监督学习。半监督学习的特点在于学习的数据中仅有部分数据具有标签而其他数据缺少标记。这是一种将监督学习与无监督学习相结合的机器学习策略，展示了辩证法中矛盾的统一与转化的哲学思想。针对监督学习在标注数据稀缺时泛化能力受限的问题，半监督学习能显著提升对未见过的新数据预测准确性；针对无监督学习因缺乏样本标签引导而导致准确率不高的缺陷，半监督学习通过引入部分标注信息加以改善。半监督学习利用有限的标注和通过未标注数据的内在结构来增强学习，既克服了纯经验主义的局限性，又避免了纯粹理性主义的空泛性，实现了经验归纳与理性推理的辩证统一。

第四种学习范式为强化学习。强化学习主要通过“交互 - 试错”机制，与环境不断进行交互，根据环境的反馈来修正行为，进而提升机器学习水平的学习策略。^[1]受到心理学中的行为主义理论的启发，该理论认为个体在奖励或惩罚的刺激下会逐渐调整行为以获取最大利益。强化学习包含智能体、环境、状态、动作和奖励五大核心要素。智能体通过与环境的交互，在试错过程中依据奖励信号指导行为，实现奖励的最大化。当智能体执行某个动作后，会进入新状态，环境随之给出正负奖励信号，智能体据此修正行为并执行新动作。鉴于人类与环境的交互也遵循类似机制，强化学习具有广泛的适用性，被认为有望解决通用人工智能问题。

3. 深度学习的进展困境与哲学反思

(一) 深度学习技术突破引发的哲学议题

深度学习技术的突破性进展引发了深刻的哲学思考。从 2016 年 DeepMind 基于深度学习开发的 AlphaGo 战胜国际围棋冠军李世石，到 2017 年 AlphaGo Zero 通过自我对弈实现超越，再到 2022 年 ChatGPT 的全球风靡。这一系列标志性事件展现了人工智能从专用领域向通用领域拓展的演进路径。AlphaGo Zero 不依赖人类先验知识和棋谱的自主学习机制，挑战了传统知识获取的观念，促使我们重新思考“经验”的本质。

ChatGPT 的崛起则将这一哲学讨论推向新高度。ChatGPT 的上下文理解能力和记忆机制实现了连贯对话。同时，ChatGPT 通过人类反馈强化学习技术显著提升了生成内容的创造性和多样性。系统通过算法实现了对人类语言、常识和价值观等方面的高度模拟，一方面展现了技术的突破性进展，另一方面也挑战了人类对“理解”和“意识”等核心概念的独占性地位，引发了关于智能本质的新一轮哲学论辩。

(二) 深度学习技术局限的哲学分析

尽管深度学习已在多个领域取得突破性进展，但现有研究表明其仍存在局限性。通过对国内现有文献的梳理，可归纳出两大核心问题：

第一，深度学习存在“封闭性”问题。当前深度学习在围棋、电子游戏等领域的卓越表现，本质上依赖于封闭性场景。封闭性场景具有三个关键属性：降射不是无尽的、场景不含难解变元、应用不存在致命性失误。而唯有在封闭性场景中，深度学习可以取得显著的应用成效^[2]。在这样的环境中，深度学习能够高效建模输入与输出之间的映射关系，并通过明确的评估标准优化模型性能。然而，一旦应用于开放、动态的现实世界，面对不确定性、信息不完全或突发干扰因素时，深度学习的表现往往会显著下降。封闭性问题的哲学根源在于形式系统与真实世界的差距。深度学习在封闭场景中的成功依赖于明确的规则和奖励信号，这反映了逻辑实证主义对确定性和可计算性的追求。然而，现实世界的复杂性和开放性

要求我们超越这种刚性框架，转向更具适应性和包容性的认知范式。

另一个局限在于模型的鲁棒性缺陷。深度学习的性能表现高度依赖于训练数据的完备性和质量，当面临有异常点入侵、类别不均衡等数据异常情况时，其可靠性会显著下降，甚至产生远低于人类认知水平的错误结果。这就导致深度学习的脆弱性与不安全性[3]。这种鲁棒性缺陷不仅制约了深度学习在医疗诊断、自动驾驶等安全敏感领域的可靠应用，更暴露了纯粹数据驱动方法的固有局限。这反映了归纳法基于有限观察的概括难以应对异常情况的固有局限，因而需引入因果推理等演绎机制，实现归纳与演绎的辩证统一。

尽管深度学习取得了显著进展，但仍然受制于上述两大核心问题。这使得现有深度学习在可靠性、可解释性、可信度和安全性等方面面临严峻挑战，难以满足关键领域对人工智能系统的严格要求。

4. 深度学习的范式突破与哲学探讨

目前，人工智能领域尚未就其核心概念、定义及研究范畴形成普遍共识。一方面，概念定义的弹性空间有利于人工智能的发展。因为人工智能的未来取决于现在的研究成果，充分发挥创新精神，不断尝试新的研究方向，才能推动人工智能更好、更快地发展。另一方面，许多专家认为与其耗费精力界定“智能”或“人工智能”的精确含义，不如聚焦于技术突破，主张认为概念的明晰化将随研究深入自然显现。但实际上，概念的混乱降低了学术对话的质量，导致研究成果难以有效整合。由于人工智能在基本理论框架上的模糊性，它必须保持高度的学科开放性。而哲学作为最具包容性的知识体系之一，恰恰能够为人工智能提供必要的概念反思。

(一) 深度学习与因果学习融合推动因果科学哲学新一轮模式转换

ChatGPT 是一种基于深度学习技术语言生成模型，由美国人工智能研究公司 OpenAI 于 2022 年 11 月底发布后迅速火遍全球。但其存在着可解释性缺陷，主要表现为难以判断因果关系、推理受限等。为了解决这一问题，需要因果科学哲学的新一轮范式转换，即深度学习与因果学习的融合[4]，以建立可解释人工智能。此时哲学的角色就变得至关重要，因为哲学研究的正是因果关系和解释性。因此，建立可解释性人工智能需要哲学与深度学习的交叉融合，这将为深度学习算法提供更加准确的解释和理解，也将为人工智能的发展开辟一条新路。

(二) 通过深度学习为机器人道德能力的构建提供新技术路径

在自动驾驶、工业自动化和智能医疗等关键应用场景中，人机交互引发的伦理问题日益凸显。如自动驾驶中的事故责任界定和机器人伤人事件的道德归因等。正如人类的道德直觉在日常道德实践中具有举足轻重的地位，人工智能机器人中的道德直觉问题始终关乎人工智能机器人的道德地位及其与人类关系的未来发展方向。[5] AlphaGo 通过深度学习复制了人类对弈的直觉网络，而索菲亚机器人获得公民身份的案例则表明，基于深度学习的人工智能系统可能具备获得道德主体资格的技术基础。要谈论人工智能的道德直觉问题，就必须在哲学上对道德直觉进行清晰的分析与表述。未来的人工智能哲学研究中，“人工智能机器人的道德直觉”将成为热点议题之一。

(三) 深度学习的发展对传统“经验”概念提出了深刻的哲学反思

深度学习以“从数据学习”和“从经验学习”为特点，“数据 - 经验”被视为智能体经学习算法创造的新经验类型，它并不是知识论意义上的经验类型，而是具有记忆特性的经验类型[6]。深度学习中的数据与经验并不等同于人类的经验。人类经验是在真实世界中通过感官和思维活动获得的，包含了情感、意识和主观体验等方面的内容。而深度学习模型获取的经验主要源于数据，这意味着数据成为经验的新来源。从本体论视角审视，“数据 - 经验”的生成机制挑战了“生命存在作为经验前提”的传统命题。该命题预设了经验积累必须依附于生物性生命载体，而深度学习系统通过算法架构实现了对数据经验的自

主建构。如果人工智能能够从数据中学习，是否意味着某种新型“生命”形式的出现？哲学视域中经验概念的边界得以拓展，形成生物经验与数据经验并存的认知图景，这要求我们重新审视经验生成机制与生命形态之间的本质关联。

(四) 深度学习追问学习本质的哲学叙事，既是对人工智能技术逻辑的哲学理解，又是对人工智能时代认识论的哲学探讨

学习作为人类智能的核心发展机制，不仅是个体认知能力的形成途径，更是智能系统自我演进的内在动力。受生物智能体学习方式的启发，深度学习结合了系统论、概率论、神经科学以及数字神经网络等理论，实现了人工智能领域的突破性进展。深度学习的关键技术逻辑包括层次化的概念、反向传播算法以及大规模数据的泛化处理。这些技术逻辑不仅实现了从无理到非线性的理性转化，更以技术现象学的方式揭示了如何通过学习实现对象世界的概念化、实践干预的原则化以及感性经验的科学化。深度学习在探索学习本质的哲学逻辑方面，既是直面人工智能存在论难题的前提性工作，也是理解学习构建智能本质的基础[7]。

当前，深度学习作为第三次人工智能热潮的主流研究内容仍处于蓬勃发展期，而第四次人工智能热潮初现端倪，对深度学习展开哲学层面的反思与探讨显得尤为重要且恰逢其时。

参考文献

- [1] 孙永丹. 从康德哲学反思深度学习的发展方向[D]: [硕士学位论文] 重庆: 西南大学, 2022.
- [2] 陈小平. 人工智能中的封闭性和强封闭性——现有成果的能力边界、应用条件和伦理风险[J]. 智能系统学报, 2020, 15(1): 114-120.
- [3] 张钹, 朱军, 苏航. 迈向第三代人工智能[J]. 中国科学: 信息科学, 2020, 50(9): 1281-1302.
- [4] 任晓明, 邢万成. 从可解释人工智能视角看因果科学革命[J]. 武汉科技大学学报(社会科学版), 2023, 25(5): 530-536.
- [5] 何瑶. 具有道德直觉能力的机器人是否可能[J]. 道德与文明, 2020(6): 132-139.
- [6] 杨庆峰. 论作为人类经验新形式的“数据-经验”[J]. 南京大学学报(哲学·人文科学·社会科学), 2020, 57(1): 78-87, 159-160.
- [7] 涂良川. 深度学习追问学习本质的哲学叙事[J]. 学术交流, 2022(11): 27-38, 191.