

AI学习和人类学习的'原理黑箱'之共性解构

原创 王德生 华智慧泉 2025年03月19日 11:59 新加坡

——从认知过程到算法模型的多维探讨

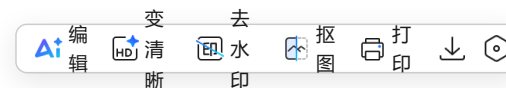
摘要 (Abstract)

人工智能（AI）技术与人类认知研究在近几十年来取得了长足进展，但对于“学习”本质的认识，却始终存在诸多谜团。无论是人脑的学习，还是深度神经网络的训练，在核心机理层面都面临一个“原理黑箱”（Black Box of Principles）的重大困境：我们能观察到输入与输出，却难以清晰、透明地呈现系统内部的运作机理与因果链条。

本论文旨在从多学科视角探讨这一问题：AI学习与人类学习在原理层面的黑箱性是否具有同源共性？若有，具体表现在哪些维度？为回答这一问题，论文系统回顾了人类学习与AI学习的主要发展脉络，对神经科学、心理学、计算机科学的相关理论进行梳理，重点分析大脑与深度神经网络的结构与功能类比，并结合哲学认识论和教育学的前沿观点，对“黑箱”的内涵及可解释性进行了深度探讨。论文在最后还展望了未来可能的研究方向，试图为跨学科的交融与互鉴提供一些思路。

关键词 (Keywords)

AI 学习；人类学习；原理黑箱；认知科学；深度学习；可解释性；神经网络；脑科学；哲学认识论





译

目 录

- 第一章 绪论
 - 1.1 研究背景与意义
 - 1.2 研究目的与方法
 - 1.3 论文结构与逻辑安排
- 第二章 “黑箱”概念溯源：哲学与科学的多重视角
 - 2.1 “黑箱”的早期使用：社会科学工程领域
 - 2.2 现代科学与技术中的“黑箱”问题
 - 2.3 从哲学认识论到认知科学：对“不可解释性”的思考
- 第三章 人类学习的图景：认知、神经与社会交互
 - 3.1 学习的心理学基础：行为主义、认知主义与建构主义
 - 3.2 神经科学视角：大脑结构与突触可塑性的奥秘
 - 3.3 社会与文化对学习过程的影响
 - 3.4 小结：人类学习“原理黑箱”的内在体现
- 第四章 AI 学习的发展脉络：从符号主义到深度学习
 - 4.1 符号主义与专家系统：可解释性的黄金时代？
 - 4.2 统计学习与连接主义：隐式表征与神经网络
 - 4.3 深度学习的兴起：海量数据、非线性与大模型
 - 4.4 强化学习、元学习与自我进化：AI 边界的不断拓展

- 第五章 AI 与人类学习的对比：形态、过程与难题
 - 5.1 神经元与人工神经元：结构与功能的类比
 - 5.2 学习算法VS.人类的经验积累：监督、无监督与潜意识学习
 - 5.3 环境与数据：人类感官输入与 AI 训练集的异同
 - 5.4 记忆与表征：如何在网络权重或突触连接中存储信息
 - 5.5 小结：相似与差异的交织
- 第六章 “原理黑箱”的提出：定义、内涵与核心问题
 - 6.1 人工智能中的“原理黑箱”：模型规模与可解释性冲突
 - 6.2 人类学习中的“原理黑箱”：心智活动难以透明化
 - 6.3 两种黑箱的交汇：是否源于同一认知困境？
- 第七章 神经科学视角的启示：人类学习的生物机制与其局限
 - 7.1 脑区分工与整合：从视觉皮层到前额叶
 - 7.2 可塑性与突触变化：学习如何改变大脑连接
 - 7.3 意识与潜意识：大量学习过程的不可言说性
 - 7.4 病例与实验研究：大脑损伤对学习机制的反证分析
 - 7.5 对“黑箱”问题的进一步佐证
- 第八章 AI 学习过程的深度剖析：算法、结构与非线性
 - 8.1 深度神经网络的层级化表征：卷积、循环与注意力机制
 - 8.2 训练范式：梯度下降与损失函数优化
 - 8.3 参数空间的高维度与局部最优：难以穷举的不可逆性
 - 8.4 模型可解释性技术：可视化、代理模型与注意力分析
 - 8.5 小结：从技术到哲学的多重维度
- 第九章 “原理黑箱”之共性探讨：同源性、同质性与差异性
 - 9.1 连接主义理论：人脑与人工神经网络的同构假说
 - 9.2 学习驱动力：生物进化 VS. 算法目标函数
 - 9.3 认知瓶颈与计算瓶颈：为什么都无法彻底透明？
 - 9.4 不可解释的边界：从无意识到超大规模网络
 - 9.5 小结：共性之下的多层次反思
- 第十章 认识论与本体论视角：为什么黑箱不可避免？
 - 10.1 主客二分的迷思：主体-客体-互动（SIO）在学习过程中的重构
 - 10.2 现象学与反思：体验的不可言说性与 AI 内部表示的抽象性
 - 10.3 知识生成的动态性：从涌现到去中心化
 - 10.4 黑箱是缺陷还是常态？多元视角的争议与融合
- 第十一章 教育学与实践启示：从人类教学到 AI 训练
 - 11.1 人类教学方法：启发式、引导式与探究式对机器学习的启示
 - 11.2 可解释 AI 在教育中的应用：智能辅导与自适应学习

- 11.3 黑箱透明化的必要与限度：教育实践中的平衡
- 11.4 未来展望：能否通过教育手段“打开”AI与人类的双重黑箱？
- 第十二章 技术与哲学的前沿交汇：走向下一代智能范式
 - 12.1 新型神经形态计算：在硬件层面模拟人脑
 - 12.2 多模态大模型与跨学科交叉：新的“学习革命”
 - 12.3 量子计算与脑机接口：黑箱尽头的曙光？
 - 12.4 伦理与社会影响：当黑箱成为决策主体
 - 12.5 小结：再思人类与 AI 的共同进化
- 第十三章 总结与展望
 - 13.1 研究结论与主要贡献
 - 13.2 局限性与未来工作展望
 - 13.3 对跨学科研究的呼吁：打开或理解黑箱的新思路
- 参考文献
- **附录：哲学的虚幻：认识世界**

译

第一章 绪论

本章主要探讨本研究的背景、意义，提出所要解决的**核心问题**，并概述文章结构与主要研究方法。

1.1 研究背景与意义

1.1.1 人工智能与人类学习：交汇与启示

人工智能（AI）在过去的几十年间经历了多次重大变革：从早期的**符号主义专家系统**，到**统计机器学习**时代，再到以深度学习（Deep Learning）为代表的**连接主义**浪潮，AI

技术的飞速发展引起了广泛关注，也在诸多领域产生了深刻的影响。与此同时，**人类对自身学习机制的研究也在不断深入：神经科学、认知心理学、教育学等领域的研究相继取得突破。**

人类与 AI 学习机制的对比，除了在方法层面引起关注，也在本体论和认识论层面带来新的思考：当我们在探究机器如何“学习”时，背后隐含的其实是对人类自我认知过程的反思。因为**许多深度学习网络的原理，正是对人类大脑神经学习特性的某种抽象模拟。**

1.1.2 “原理黑箱”问题：技术与哲学的双重困境

然而，无论是人类学习还是 AI 学习，都存在一个“原理黑箱”（Black Box of Principles）的问题：**我们无法充分、透明地说明整个学习的机理到底是如何在内部运作的。**对于人脑而言，人类可以进行主观反思，却难以详尽描述无意识、潜意识及海量神经元交互的全部细节。对于 AI 系统而言，我们能在技术层面查看网络架构和参数，但在面对数以亿计或数百亿计的网络权重时，也难以形成**直观、全面的“因果链式”的解释。**

这一现象不仅是工程层面的可解释性难题，更涉及到对认知、智能和**知识本质的思辨。**因此，本研究希望跳出单纯的技术分析视角，从哲学认识论、神经科学、教育学等多学科的视角综合探讨，力图在宏观层面解构这一黑箱，并讨论其必然性与可行的突破口。

1.1.3 跨学科对话的重要性

学习是一个复杂而多元的过程。若仅从算法设计或脑科学的单一角度去解释，都难以全面地触及学习的本质及其“黑箱性”。为此，本文将综合运用以下视角：

1. 认知科学

：考察人类学习的心理与神经机制；

2. 人工智能

：探讨 AI 学习的技术逻辑与方法；

3. 哲学认识论

：审视人与机器学习过程中“何为知识、何为可解释”；

4. 教育学

：从实践层面考量如何帮助个体和系统更好地学习；

5. 社会科学与伦理

：思考当“黑箱”成为社会决策的一部分时，应如何监管和规范。

1.2 研究目的与方法

1.2.1 研究目的

本研究旨在：

1. 系统梳理

人类学习与 AI 学习各自的核心理论、关键技术与难题；

2. 深度讨论

二者在“原理黑箱”层面的共性与差异，揭示其背后的认识论根源；

3. 提出启示

：从跨学科、跨领域的角度，为未来AI 研究及人类教育实践中的“可解释性”与“黑箱”问题提供若干思路。

1.2.2 研究方法

1. 文献综述与归纳

：对心理学、神经科学、AI 发展史、教育学等进行广泛而系统的文献回顾与提炼；

2. 跨学科比照与理论分析

：将多学科理论放在同一框架下对比，寻找相似点与不同点；

3. 案例研究与理论验证

：引用已有的实验研究（神经科学、心理实验）和 AI 模型案例，辅助理论分析；

4. 哲学思辨

：对不可解释性、不确定性、知识生成等议题给予哲学层面的剖析，以求在更高层次上理解“原理黑箱”现象。

1.3 论文结构与逻辑安排

本文共分十三章。第一章介绍研究背景与意义，并说明研究目的与方法。第二章对“黑箱”概念的起源和多领域应用进行梳理。第三章到第五章分别聚焦人类学习、AI 学习及其二者对比，为后续讨论奠定基础。第六章提出“原理黑箱”的核心概念，并探讨其关键内涵。第七、八两章分别从神经科学和 AI 技术角度进一步挖掘内部学习原理，形成对“黑箱”现象的整体认知。第九章深入对比二者在“黑箱”问题上的共性与差异。第十章聚焦认识论与哲学层面的阐述。第十一章和第十二章则分别讨论教育学与前沿技术对于这一问题的启迪与未来展望。最后，第十三章对全文进行总结与反思，并提出可能的未来研究方向。



第二章 “黑箱”概念溯源：哲学与科学的多重视角

在讨论人类学习与AI 学习的“原理黑箱”之前，有必要先回顾“黑箱”这一术语在不同学科中的使用历程。此概念最初并不专门用于描述学习过程，而是广泛出现于社会学、工程学、科技史等领域，用来指称那些**只能观察其外在输入与输出，内部机制无从得知或难以理解**的系统或过程。本章将从历史、哲学与科学技术三个维度溯源“黑箱”一词的背景与演变。

2.1 “黑箱”的早期使用：社会科学与工程领域

2.1.1 社会学与科学技术史的“黑箱化”理论

在社会学与科学技术研究（STS）领域，布鲁诺·拉图尔（Bruno Latour）等人常用“黑箱化”（Blackboxing）来描述科学技术发展中的一种现象：当**某个技术或理论已被社会**

普遍接受，其内部争议和复杂过程就被“封装”起来，外界只看到其最终的功能或结论，而忽略了形成过程中的博弈与不确定性。

与此相似的是，现代社会中的许多大型系统（如互联网、金融系统）对大多数人而言也呈现“黑箱”特征：人们仅能使用其功能，却无法理解**内部的结构和机制**。即对系统的‘场’和‘波’模式不做研究，只是‘端对端’的检验，即粒子模式的考察。

2.1.2 工程与控制论中的“黑箱”范式

在控制论和工程学中，“黑箱”指我们只关注**系统的输入与输出行为**，而对其内部实现并不关心或无法得知。例如，早期的电路模型常用“黑箱”表达：只要我们知道输入信号与输出信号的规律关系，内部具体用什么元件、怎样排布，对使用者来说可以暂时不予考虑。

这种思路在一定程度上提高了工程效率，但也遮蔽了对系统内部**安全性、可靠性以及潜在缺陷**的关注。随着系统规模和复杂度增加，人们开始意识到“黑箱化”的风险。

2.2 现代科学与技术中的“黑箱”问题

2.2.1 高度复杂系统的不可解释性

不论是**经济模型**还是复杂**生态系统**，伴随规模的上升与要素的多样化，系统内部出现大量**非线性反馈与时变过程**，**传统的线性因果分析或简化模型很难给出完整解释**。在这些领域，研究者往往只能通过观察变量间的相关性来推测系统的整体行为，而难以做到对每个状态转变的透明化描述。

2.2.2 量子力学与“测不准”思维

在物理学层面，量子力学揭示了微观世界的测不准与波函数坍缩等现象，挑战了经典物理“可测与可预测”的范式。虽然我们可以用数学公式准确计算干涉结果，却无法从“经典直觉”出发解释电子为何以某种方式出现或消失。某种程度上，这也可被视作“黑箱”：在某些层面上，我们只能描述输入输出（测量设置与测量结果），无法真正穷尽中间的“真实”演化细节。

2.2.3 复杂算法与机器学习的现代“黑箱”

伴随信息技术的飞速发展，**机器学习和深度神经网络**日益成为“黑箱”讨论的核心：

1. 深度网络规模巨大

：数千万乃至数十亿个甚至目前的千亿参数；

译

2. 高维度非线性关系

：难以用简单的规则式、线性式方法剖析；

3. 数据驱动

：算法的特征提取和决策模式往往并非工程师提前设定，而是在训练中**自动涌现**。

这些特征使得“黑箱”成为机器学习领域的关键难题，并激发了可解释性（Explainable AI, XAI）的兴起。

2.3 从哲学认识论到认知科学：对“不可解释性”的思考

2.3.1 认识主体与客体：主客二分的困境

从笛卡尔到康德，认识论的一大主题在于主体如何把握客体“真理”，以及这种把握是否可能彻底透明。后来的现象学、后现代主义甚至进一步否定了认识过程中的客观性，一定程度上认为任何经验都打上了主体的烙印。正因此，当我们谈论“黑箱”时，也需要反思：

- 是否存在一个独立而客观的内部机制等待我们去揭示？
- 还是说我们总是难以将我们的观察者立场与对象本身严格区分？

2.3.2 人类认知的有限性与潜意识

当代认知科学承认，人类理性与意识只占整个认知过程的一小部分；潜意识、自动化认知在日常思维、判断与学习中发挥了至关重要的作用。换言之，**人类自己对自身的学习过程，也不具备绝对的透明度**：很多学习发生在无意识状态下，**没有形成可被语言捕捉的明确规则或逻辑链条**。

2.3.3 小结：黑箱作为不可避免的系统特征

综上，“黑箱”并非仅仅是一个贬义词或特殊的技术难题，它在一定程度上反映了***当系统足够复杂或以隐式方式运行时，外部观察者难以或无法完整重现其内部机制***的普遍现象。

后续关于人类学习与 AI 学习的讨论，正是在这一学术与现实背景下展开：两者或许都无法彻底摆脱“黑箱”性质，而这种性质并不是简单的技术方案或实验设计就能消除的。

译



第三章 人类学习的图景：认知、神经与社会交互

要理解人类学习与 AI 学习在原理黑箱上的共性，首先需要明确：**人类学习到底是一个什么样的过程？** 这一问题在心理学、神经科学、教育学等学科内的回答并不完全一致。本章将从多个角度概述人类学习的主要理论与研究现状，力图呈现一个综合且多层次的图景。

3.1 学习的心理学基础：行为主义、认知主义与建构主义

3.1.1 行为主义：刺激—反应的可观测链条

在二十世纪初期，行为主义（Behaviorism）学派提出：心理学应当研究外显、可观察的行为，而非不可观测的内在心理过程。巴甫洛夫的条件反射实验、斯金纳的操作性条件作用都表明，通过刺激—反应（S-R）机制可以塑造动物或人的行为模式。

• 优势

：强调客观、可测量，通过量化方式研究学习；

• 局限

：忽视了学习者的内部认知过程，比如思维、记忆、情感等主观要素。

在讨论“黑箱”时，行为主义往往选择将“内部精神活动”当作黑箱而不做过多探究，只关心输入（刺激）与输出（行为反应）的关联。

3.1.2 认知主义：信息加工与表征

认知主义（Cognitivism）在二十世纪中叶兴起，主张将学习视为信息加工过程。学习者并非被动接受刺激，而是主动构建心智模型，对信息进行选择、编码与整合。皮亚杰（Jean Piaget）的发生认识论从儿童认知发展角度证明，学习带有阶段性、建构性与自主性。

• 优势

：关注学习者的内部过程，如注意力、记忆、问题解决等；

• 挑战

：如何准确描述认知过程的运转机理仍是难题，许多细节停留在理论模型，缺乏全脑生物学证据。

3.1.3 建构主义：社会互动与情境体验

建构主义（Constructivism）进一步强调学习与社会环境、情境活动的关联。维果茨基（Lev Vygotsky）将学习置于文化与社会互动之中，认为个体是通过与他人交流、实践任务中不断修正自我认知结构而学习的。

这一视角尤其强调学习并非个人大脑中孤立的过程，而是主体与环境、他者持续交互的结果。可以说，行为主义和早期认知主义所忽视的“外部情境与社会因素”在建构主义框架中得到补足。

3.2 神经科学视角：大脑结构与突触可塑性的奥秘

3.2.1 脑区功能分工与整体协作

借助脑成像技术（fMRI、PET、EEG 等）以及对脑损伤患者的研究，科学家们逐渐勾勒出大脑功能区域：视觉皮层（Occipital Lobe）负责视觉加工，海马体（Hippocampus）与记忆形成密切相关，前额叶皮层（Prefrontal Cortex）与高级认



知和决策有关.....然而，这种分区并非绝对独立，各脑区之间存在广泛的交互与并行处理。

3.2.2 突触可塑性：学习的生物学基础

大脑拥有约数百亿神经元，每个神经元通过数千至数万个突触与其他神经元相连接。突触（Synapse）是神经元通信的关键部位，其强度可随经验和活动而改变，这被称作突触可塑性（Synaptic Plasticity）。**学习从某种意义上就是突触连接在模式与强度上的动态改变**，从而产生记忆或新的行为反应。

- ** 长期增强（Long-Term Potentiation, LTP）与长期抑制（Long-Term Depression, LTD）**是研究最多的突触可塑性机制，分别对应突触连接的加强与减弱。

3.2.3 对“黑箱”的神经解释：信息分布与多层次整合

尽管现代神经科学在实验与理论上不断进步，但**想要将数百亿神经元的放电模式与行为或认知内容一一对应**依然不现实。大脑的信息处理是分布式、并行化且动态涌现的：单独的神经元或突触无法简单地“代表”某个完整概念或记忆，而是要在网络整体活动的协调下才能呈现出学习效果。这意味着从神经层面理解学习的内部机制也存在高度的不可简化性。

3.3 社会与文化对学习过程的影响

3.3.1 学习的社会性：同伴互助与认同

人类学习并非仅依赖神经机制，也深受社会文化因素的影响。学校教育、家庭背景、社会互动等会塑造个体对知识的态度与习得方式。合作学习与同伴教学（Peer Learning）也被证明可大幅提升学习效率与深度。

3.3.2 “情境化”学习：真实环境与实践活动的重要性

建构主义强调学习要与真实情境相结合。人在解决现实问题时，需要综合利用已掌握的知识、工具和策略，往往在此过程中“学到更多”。这种体验式学习模式突出了人与环境、人与任务的双向交互，进一步说明**学习过程的复杂度**不只在脑内，也在于人与外部世界的耦合。

3.4 小结：人类学习“原理黑箱”的内在体现

从以上角度可见，人类学习绝非简单的线性过程，而是涉入多层机制：**神经层面**（突触网络）、**心理层面**（认知过程与记忆存储）、**社会层面**（文化环境与人际互动）。这种多维交互导致学习的高度复杂性与不可还原性。对于大多数日常学习过程，我们只能看到结果，而无法完整捕捉其中所有神经与心理细节，**这正是“原理黑箱”在类生物系统中的真实写照。**



这一结论也为后续关于 AI 学习的讨论打下基础：如果人类自身在学习上并无透明可见的内在机理，AI 学习在某种程度上必然也会出现类似的难题。



第四章 AI 学习的发展脉络：从符号主义到深度学习

在阐述了人类学习的丰富图景后，我们再将视线转向 AI 学习的发展历程。人工智能领域的研究范式经历了多次迭代，每一次都在某种程度上借鉴或模拟人类认知过程，却也都面临可解释性和“黑箱”性质的争议。本章将对 AI 学习的主要历史阶段与核心方法做一系统回顾。

4.1 符号主义与专家系统：可解释性的黄金时代？

4.1.1 早期 AI：符号操作与逻辑推理

二十世纪五六十年代，AI 研究以符号主义（Symbolism）为主流，认为智能的本质是对符号进行操纵和推理。此时的研究者往往采用逻辑规则或知识库来模拟人类的“思维过程”，典型系统如**专家系统（Expert Systems）**。

• 可解释性强

：专家系统可输出推理链条，如“若条件 A 则进行操作 B，否则执行 C”；

• 局限性

：规则需人工编写，难以应对大规模、不确定和模糊的任务情境。

4.1.2 可解释性与“黑箱”相对性的体现

在符号主义阶段，AI 很容易呈现给用户“一条一条的推理规则”，因此也被视为一种“白箱”或高度可解释的模型。然而这种规则式系统在面对图像识别、语言理解等需要**庞大特征空间与统计学习**的领域时往往力不从心，无法像人类那样灵活学习和适应新环境。这种**“可解释却缺乏深层学习能力”**的瓶颈也为后来的连接主义兴起埋下伏笔。

4.2 统计学习与连接主义：隐式表征与神经网络

4.2.1 统计学习：从感知机到支持向量机

随着计算机硬件与数据获取能力的提升，基于统计学原理的机器学习兴起。早期的感知机（Perceptron）已经初步模拟了生物神经元，但当时只具备一层或少量层次，表达能力有限。八九十年代出现的支持向量机（SVM）等算法在结构化数据上表现优异，但同样对于**高维感知任务**（如图像、语音）的泛化仍有局限。

4.2.2 连接主义浪潮：多层神经网络的再崛起

1990-2000 年代，随着算法（如反向传播）优化与硬件算力提高，连接主义（Connectionism）再度复兴。多层感知机、卷积神经网络（CNN）等在图像识别、语音识别上实现了突破。此时的 AI 学习更像生物大脑，通过大量样本训练而不是人工写入规则；权重更新和特征提取均在网络内隐式完成，**可解释性下降**，但**识别性能大幅提升**。

4.3 深度学习的兴起：海量数据、非线性与大模型

4.3.1 深度学习的关键要素

• 大数据

：在互联网、物联网时代，数据规模激增；

• 深层网络结构

：多层隐藏层的堆叠，具备强大的特征提取能力；

• 高性能计算

：GPU、TPU 等硬件加速使得训练百亿级参数模型成为可能。

4.3.2 关键模型与应用场景

• 卷积神经网络（CNN）

：图像分类、物体检测、医学影像；

• 循环神经网络（RNN）以及LSTM、GRU

：序列建模，如文本、语音；

• Transformer 结构

：注意力机制在机器翻译、自然语言处理上的重大进步；

• GAN、VAE

等生成模型：图像生成、风格转换；

• 大型预训练模型

（如 GPT 系列）：在语言理解、生成任务上展现出超越传统手段的性能。

4.3.3 “黑箱”再度凸显：深度网络的不可解释性难题

深度学习模型往往包含数百万、数亿甚至上千亿的参数，且大量特征是网络自发涌现出来的：

• 参数含义难以直观理解

：某个权重或某个神经元激活模式在多层网络中如何影响最终决策往往难以理清；

• 非线性映射不可逆

：我们无法简单地从输出追溯到哪几个输入特征导致了该结果；

• 隐式特征表示

：无需人工定义特征，模型自主学习，但也随之失去了清晰的可读性。

4.4 强化学习、元学习与自我进化：AI 边界的不断拓展

在深度学习的热潮中，强化学习（Reinforcement Learning）将决策与环境交互结合，打开了游戏智能、自动驾驶、机器人控制等新领域。而元学习（Meta-Learning）、**少样本学习（Few-Shot Learning）以及自我监督学习（Self-Supervised Learning）**等方法则进一步探索如何让 AI 像人类一样在少量经验或无标签数据中学习。

这些进展尽管在一定程度上接近人类的学习方式，却也加剧了模型结构的复杂度，使得“黑箱”现象更加突出。



第五章 AI 与人类学习的对比：形态、过程与难题

在上两章的铺陈下，我们已分别了解人类学习和 AI 学习的主要面貌与挑战。本章将通过逐点对比，对二者的异同进行梳理，以便更具体地认识它们在学习过程、结构和原理黑箱上的关系。

5.1 神经元与人工神经元：结构与功能的类比

5.1.1 生物神经元与突触模型

人类大脑由神经元及突触连接构成，突触可塑性负责调节神经元间通信效率，是学习和记忆的生物基础。一个神经元接收数千至数万个输入信号，**经综合后决定是否放电**并将信息传给下游神经元。

5.1.2 人工神经元与权重更新

在人工神经网络中，神经元的输入来自前一层输出，通过一组**可学习的权重（Weight）**和偏置（Bias）进行加权求和，再经过**激活函数**（如 ReLU、Sigmoid 等）产生输出。反向传播算法根据**损失函数梯度**对权重进行更新，这一过程类似**突触可塑性的动态调整**。

5.1.3 相似之处与局限

- **相似点**
：分布式处理、权重或突触可塑性、多层级结构；
- **局限性**
：生物大脑结构更复杂（存在胶质细胞、神经调质、空间拓扑差异），学习过程也更灵活，自主性更高；人工神经网络仍是对生物神经网络的理想化近似。

5.2 学习算法 VS. 人类的经验积累：监督、无监督与潜意识学习

5.2.1 监督学习：标注数据 VS. 明确目标

在机器学习中，大量应用属于监督学习：给定带标签的数据，让模型“知道”**正确答案并进行调优**。人类学习中，也有类似的场景，如接受教师反馈，练习题中的“正确答案”提供外部监督。但人类更常遭遇的情况是：**并非所有环境都有明确标签**，很多经验是在**模糊的探索或社会互动中获得**。

5.2.2 无监督与自我监督：模糊环境中的结构发现

无监督学习或自监督学习让模型在无标签或弱标签环境中自动挖掘数据结构，类似人类在感知世界时不一定有导师“指点”，依然可以建立对物体、事件的认知表征。此类学习更接近人类在婴儿期的“自发建构”过程。

5.2.3 潜意识学习：无法言说的隐式机理

人类还拥有大量潜意识或隐性学习，这部分学习不是通过语言或显性反馈来完成，而是在与环境、他人持续交互中渐渐习得（如学会骑自行车）。AI 中也存在类似的隐式表征学习，如大型语言模型在无监督文本海量训练中，学到了句法与语义模式，却无法明确“告诉”人类自己是如何形成这套知识的。

5.3 环境与数据：人类感官输入与 AI 训练集的异同

5.3.1 多模态感知 VS. 数据格式化

人类通过视觉、听觉、触觉、嗅觉、味觉等多模态渠道，同时接收混合且连续的感官信息；AI 则通常通过数码相机、传感器或数据文件（文本、图像）等离散化管道获取信息。

- **人类优势**

- ：感官融合与情境理解（Context Awareness）更自然；

- **AI 特点**

- ：可快速处理海量数据，但缺乏先天知识或常识 unless 通过训练数据弥补。

5.3.2 数据清洗与先验偏见

AI 模型的训练数据往往要经过人工或自动化清洗、标注；其中也可能包含人为偏见或不平衡分布而导致模型歧视或性能不稳。人类同样可能从社会文化中继承偏见和主观印象，但这一过程更难被精确量化，带有大量潜在的复杂性。

5.4 记忆与表征：如何在网络权重或突触连接中存储信息

5.4.1 海马体与长期记忆：人类记忆的“中转站”

人类大脑的海马体在记忆巩固中扮演关键角色，但长时记忆被认为分散储存在大脑皮层的不同区域。对于“如何分散存储和提取”的机制，目前仅部分澄清，仍有许多未知。

5.4.2 AI 权重矩阵的“内隐记忆”

深度神经网络的记忆主要体现在权重矩阵的数值分布中。当网络完成训练后，这些数值搭配特定结构，决定了模型对输入数据的处理方式。然而网络并不会像人类一样“自省”或“自陈述”，其知识完全以数字化参数的形式隐式存在。我们若想“解读”这些权重意味着什么，往往困难重重。

5.5 小结：相似与差异的交织

从结构到过程再到外部环境需求，我们发现**人类学习与 AI 学习**既存在“连接主义”意义下的深层相似（如分布式处理、隐式学习）也存在**在生物约束与社会交互层面的诸多差异**（如自主意识、多模态融合、情感因素）。

这种复杂的异同为“原理黑箱”提供了一个丰富的图景：二者的内部机制在某些方面确有可类比之处，但都没有现成、完美的方式让我们彻底“读出”或“透明化”其全部运作原理。这为下一章正式提出“原理黑箱”问题埋下了伏笔。

第六章 “原理黑箱”的提出：定义、内涵与核心问题

基于前面对人类学习和 AI 学习的多维探索，本章将聚焦于所谓“原理黑箱（Black Box of Principles）”这一核心概念，提出其定义、内涵及关键问题，从而为后文深入讨论奠定理论基础。

6.1 人工智能中的“原理黑箱”：模型规模与可解释性冲突

6.1.1 传统可解释性与现代大规模模型的矛盾

如前所述，早期符号主义 AI 系统可提供较为清晰的推理链条，但在复杂任务上表现不足；而现代深度学习在性能方面占优，却在可解释性上显著退步。****这里的“黑箱”**不**

仅指“我们不知道某个预测结果到底基于哪些特征”，更深层地，**指我们无法在原理层面（如神经网络为何能从数据中涌现出高阶抽象概念）给出完备的描述。**

6.1.2 大模型时代：黑箱问题更加突出

如今的语言模型（如 GPT 系列）或多模态模型动辄拥有数十亿甚至万亿级的参数，不仅在推理和生成任务上取得惊艳表现，也带来了**近似人类“直觉思维”的行为。但这种行为背后是否与人类的认知过程相似？我们几乎无法仅凭阅读海量参数来得出确切的答案。**

6.2 人类学习中的“原理黑箱”：心智活动难以透明化

6.2.1 自我意识并不等于自我解释

尽管人类具有思维和意识的主观体验，但神经科学和心理研究表明，很多决定性认知活动是在潜意识或自动化层面完成，我们仅能在事后进行“合理化”解释。也就是说，**对自我行为的解释常常并非真实的内在推理过程的直接再现，而是一种语言化的叙述性产物。**

换句话说，人类大脑同样有无法外显的“信息处理子系统”，对其细节我们亦无从得知。

6.2.2 无法追溯的成长与经验

回顾个人成长过程，我们大多数技能与知识都在日常生活、社会环境、兴趣驱动或随机机遇下习得，根本无法用类似“算法日志”或“训练日志”的方式完整记录。于是，**学习的“原理”**就“藏”在脑中极度复杂的神经连结变化之中。这与 AI 在训练时自动学习到的内在表征极其相似，**都难以在事后给出线性、因果式的自我说明。**

6.3 两种黑箱的交汇：是否源于同一认知困境？

6.3.1 同源论：人类与 AI 分享同一分布式学习原理？

有些学者（尤其是连接主义与神经形态计算支持者）认为，**AI 学习与人类学习之所以都难以解释，正因为二者在本质上都属于高度分布式的神经或类神经信息加工。**从这一角度讲，“原理黑箱”就是分布式学习的一种不可分割特征。

6.3.2 异质论：人类与 AI 的根本区别仍然存在

也有学者强调**人类的自我意识、情感系统与社会性**等高阶特征是当前 AI 尚未实现或模拟到位的。这部分特征可能赋予人类某些“理解”或“元认知”的能力，使得人类的黑箱与 AI 的黑箱并不完全相同。但同时，这些“高级特征”自身也难以被解释或外显，因此并不能彻底消除“原理黑箱”问题。

6.3.3 小结：原理黑箱的跨物种普适性与多重争议

总体而言，从机器到人脑，“原理黑箱”都指向了对学习背后原理机制的不可透明性。这种现象或许与分布式信息处理、非线性涌现、社会文化介入等多因素共同作用有关。后续章节将从神经科学与技术细节两大视角继续追寻这一黑箱的深层结构。



第七章 神经科学视角的启示：人类学习的生物机制与其局限

要理解“原理黑箱”在生物层面如何呈现，必须深入到神经科学的实证研究。现代脑科学在实验技术上已经取得显著进展，包括使用电生理记录、功能性磁共振成像（fMRI）、正电子发射断层扫描（PET）等方法观测脑区活动。然而，这些手段依然远远不足以让我们彻底解码大脑内部的所有学习过程。本章将详述一些关键发现，并进一步阐释其与“黑箱”问题的关系。

7.1 脑区分工与整合：从视觉皮层到前额叶

7.1.1 视觉系统的典型研究

研究大脑学习常常从视觉系统切入，因为视觉输入路径相对可控且研究手段成熟。经典发现包括**初级视觉皮层（V1）**中存在对边缘、方向等低级特征的选择性神经元；**高级视觉皮层（如 IT 皮层）**则对更复杂形状或甚至脸部有选择性。

然而，这种分区并不代表信息是“单向”或“分块”处理的，大量反馈连接与侧向连接让视觉加工呈现动态交互模式。若要追溯某个对象识别的具体神经放电序列，几乎不可能在全脑范围完整还原。

7.1.2 前额叶与决策：高级认知的执行控制

前额叶皮层（尤其是背外侧前额叶、前部扣带回等区域）被视为人类高级执行功能和决策控制的重要枢纽，对注意、计划、工作记忆等起关键作用。研究发现，前额叶的功能在学习过程中帮助个体形成目标导向行为。但前额叶也需要与其他感知、情感脑区实时交互，真正的决策往往是多脑区协同涌现的结果。

这说明，即使在“大脑最高层”，对某一决策或学习的内部机理也是“离散且并行”的活动，难以在单一区域找到一个“解释节点”。

7.2 可塑性与突触变化：学习如何改变大脑连接

7.2.1 LTP 与 LTD：突触可塑性的分子基础

长期增强（LTP）意味着突触传递效率的增加，被视为学习记忆形成的重要机制；长期抑制（LTD）则是突触传递效率的降低，为遗忘或抑制无用连接提供机制。研究者通过电生理刺激和分子生物学手段确认了这些现象的存在与化学信号通路的对应关系（如 NMDA 受体等）。

但在大脑宏观层面，**数千亿突触共同作用，形成极其庞大的可塑连接网络**，任何局部的 LTP/LTD 改变都可能对整体产生连锁影响，造成时空尺度上的极度复杂性。

7.2.2 从突触到网络涌现：理解学习的关键谜题

试图从单个突触或单个神经元的角度去解释整个人类学习，是不可行的。信息是通过分布式模式编码、神经回路循环与多级反馈等机制共同实现的。正因如此，**大脑的学习原理很难被线性、还原式地描述**。研究者们通常使用网络模型或大规模模拟（如 Blue Brain Project）去探索大脑局部电路如何涌现更高层功能，但离完整解码仍有非常遥远的距离。

7.3 意识与潜意识：大量学习过程的不可言说性

7.3.1 潜意识学习的神经基础

日常行为中，我们经常“不知不觉”地学会了某些技能或知识，如语法习得、骑车平衡等。脑成像研究提示，皮层—基底节环路和小脑等结构在隐式学习中发挥重要作用，而前额叶皮层可能仅在需要意识控制或反思时才明显活跃。

潜意识学习本身就证明，大脑并不需要时刻都以有意识或可言说的方式来编码学习，**这为“原理黑箱”提供了又一个实证支撑**：许多关键的学习在意识之外完成，我们很难自己向自己解释它的详细原理。

7.3.2 自我报告与主观反思的局限

在人类学习研究中，问卷、访谈等主观自我报告方法广泛使用，但这些方法只能捕捉到个体自觉层面的内容。若个体并未意识到学习或无法用语言表达自己对某概念的认知结构，自我报告就失效了。
换言之，**人类并不是自己大脑机制的透明观察者**，因此从第一人称视角进行解释同样难以穷尽所有内部过程。

7.4 病例与实验研究：大脑损伤对学习机制的反证分析

7.4.1 失忆症患者与记忆区隔

临床上，有些失忆症患者如著名的 H.M. 病例，因双侧海马切除而无法形成新的显性记忆，但保留了习惯技能（如画画、弹琴）的学习能力。说明了**脑内不同类型记忆的分离**，同时也强调技能学习常常是无意识或自动化的，与显性记忆无关。这再一次表明学习的复杂性超越了单一机制。

7.4.2 语言功能障碍与偏侧脑损伤

某些左半球损伤者可能出现严重的语言功能丧失，却仍保留部分语言理解或语义推断能力，表明**语言处理在脑中分布于多个区域**，某些区域间或许存在功能冗余或替代。因而，简单将语言理解归因于“一个区域”或“一个通路”并不靠谱。这同样暗示了人的学习在神经基质上具有更大弹性与分散度，增大了解释困难。

7.5 对“黑箱”问题的进一步佐证

通过神经科学的视角，我们更清晰地看到：**人类大脑学习的高度复杂与分布性是“原理黑箱”的根本根源之一**。海量神经元、无处不在的突触可塑性、多脑区并行交互以及潜意识参与，都让“学习”显得错综复杂。**对人脑而言，“原理黑箱”几乎是一种系统属性，而非仅仅是暂时的技术难题。**

这一结论也为我们理解 AI 学习黑箱提供了重要类比：当 AI 模型规模达到亿级参数量级，并在隐式方式下自动学习特征时，或许与人脑所面临的复杂程度存在某种意义上的同构性。





第八章 AI 学习过程的深度剖析：
算法、结构与非线性

第七章从神经科学证明人类学习难以完全解释。本章则聚焦于 AI 领域的实际算法与网络结构，详尽剖析其在训练过程、参数空间与非线性涌现方面的特征。通过这样的细节层面对比，我们可更深刻地体会到**AI 黑箱与人类黑箱的同构性或相似性**。

8.1 深度神经网络的层级化表征：卷积、循环与注意力机制
8.1.1 卷积神经网络（CNN）：层层抽象与特征涌现

CNN 在图像处理上大显身手，其分层结构通常呈现**从低级边缘检测到高级语义特征**的渐进过程。研究者发现，在训练好的 CNN 中，越靠近输入层的滤波器倾向于检测边缘或简单纹理，而越靠近输出层的卷积核会响应更高层的概念（如“猫脸”、“车轮”）。但这种“层级化表征”并非简单的线性组合，而是多层卷积和池化的交织，最终激活模式难以一一解析。

8.1.2 循环神经网络 (RNN)、LSTM、GRU：序列依赖与记忆门控

在处理语言或时序数据时，循环结构通过隐藏状态将过去的信息传递到下一时刻。LSTM (Long Short-Term Memory) 与 GRU (Gated Recurrent Unit) 等改进结构利用“门控机制”来缓解梯度消失问题，提升长期依赖的捕捉能力。这些门控机制在一定程度上模仿人类大脑的“记忆选择”过程，但也带来更复杂的内部状态。

8.1.3 注意力机制 (Attention) 与 Transformer 架构

Transformer 的兴起彻底改变了自然语言处理和多模态学习。基于注意力 (Attention) 的结构可以在任意位置间建立关联，提升了并行计算效率和对长距离依赖的处理。然而，网络内会学习到庞大的“注意力权重分布”，这些权重可能在语义层面具有一定可解释度，也可能只是对上下文的一种数学分配，**其含义并非总能与人类语言规则完全对应。**

8.2 训练范式：梯度下降与损失函数优化

8.2.1 反向传播 (Backpropagation) 算法

反向传播通过计算输出误差相对于各层参数的梯度，逐层更新权重。其工作原理类似于**梯度下降在高维损失空间里搜索最优解，但损失地形往往极度复杂**，存在无数局部极小值或鞍点，训练过程也具有路径依赖性。

8.2.2 批量归一化、正则化与数据增强

为改善训练稳定性与模型泛化能力，业界引入批量归一化 (Batch Normalization)、L2 正则化、Dropout 等手段。但这些技术使得网络对数据分布、初始化、超参数设置的敏感度仍然存在，**训练过程在微观层面充满随机性与不可逆性。**

8.3 参数空间的高维度与局部最优：难以穷举的不可逆性

深度网络的参数维度常以百万、亿计，形成高维度的误差表面。研究表明，在高维空间里，局部极小值并不如低维空间那般明显，且鞍点、峡谷等地形特征使得优化路径充满了复杂动态。**我们难以用简单的因果链来回溯每个梯度更新为何会导致某些表征的形成。**

8.4 模型可解释性技术：可视化、代理模型与注意力分析

8.4.1 可视化手段：卷积核可视化与类激活映射

例如**Grad-CAM**等技术可以在图像分类中突出模型关注的区域，帮助人类理解网络的“注意点”。但这种可视化往往仅能呈现结果的一部分，并不能全面解析网络的每个神经元或参数的功能。

8.4.2 代理模型与局部解释（LIME、SHAP）

LIME 等方法通过在某个输入附近进行扰动测试，学习一个线性或可解释的代理模型，以此解释黑箱模型在该局部的行为。虽然对局部区域的解释往往有一定准确度，但无法保证对整个模型的全局解释。

8.4.3 注意力权重分析：透过门控看“思考”过程？

Attention 机制的可视化被许多人认为是提升可解释性的关键之一。然而，学界也指出注意力权重并不一定等于因果推理，它只是计算分配的结果，并非真正代表人类语义理解的过程。

8.5 小结：从技术到哲学的多重维度

通过对深度学习框架与训练过程的剖析，我们见证了**庞大参数维度、非线性涌现、数据驱动特性**共同构成了 AI 学习的内在黑箱：

1. 量级庞大的参数

不具备单一语义；

2. 训练路径的随机性与不可逆性

限制了对全过程的重演和分析；

3. 自动化特征提取

带来对人类语言和意识层面的“失联”。

这与人类大脑内分布式、隐式与潜意识的学习机制具有某些平行相似之处，进一步佐证了二者“原理黑箱”可能存在的**共性根源**。





第九章 “原理黑箱”之共性探讨： 同源性、同质性与差异性

经过前八章的铺垫，本章将集中对“AI 学习与人类学习在原理黑箱上的共性”进行深入探讨。这也是本论文的核心问题之一：**为什么人类与 AI 都存在难以深入解析的学习原理层黑箱？它们的共性是否意味着二者本质一致？** 我们也将指出二者的一些关键差异点，以避免过度类比的偏差。

9.1 连接主义理论：人脑与人工神经网络的同构假说

9.1.1 同构核心：分布式并行计算与突触/权重更新

连接主义（Connectionism）理论认为，**智能或学习并非源于高级逻辑推理，而是通过大量简单单元在网络结构中的并行活动涌现。**人脑中，神经元的发放与突触可塑性；AI

中，模拟神经元的激活与权重调整。
两者都利用多层次的分布式表征，从而在大规模数据或感知输入中逐渐“学会”新的模式。这些内在表征往往不是先天固定的，而是在训练与经验积累中形成，**具体形成细节依赖全局网络活动**，难以局部或单点追溯。

9.1.2 同构并不意味着完全相同

即使在连接主义框架下，人脑的神经元结构比人工网络更复杂，且有生化信号、膜电位、神经调质等多种机制共同作用。AI 网络一般只模拟突触连接强度的改变，对于突触亚基结构或神经递质调制并无对应机制。但宏观原理上，两者都是**分布式学习 + 并行表征**，**黑箱**自然在此结构中滋生。

9.2 学习驱动力：生物进化 VS. 算法目标函数

人类和动物的学习从生物进化的角度看，旨在**生存与繁衍**；AI 学习则是人类工程师设计的算法，在“最小化损失函数”“最大化奖励”或其他目标函数下进行迭代。

- **共同点**
：都存在一个“外部目标或选择压力”，推动系统不断调整内部结构；
- **差异点**
：人类进化所带来的情感、动机与社会学习等因素，令学习具备更多维度；AI 的目标函数往往更单一或由人类定义。

9.3 认知瓶颈与计算瓶颈：为什么都无法彻底透明？

9.3.1 人类认知瓶颈：大脑资源有限与进化交易

人类大脑在结构和能耗上已经非常精妙，但依然有容量与资源限制，不可能对所有外部信息进行细粒度处理，更无法对自身内部活动进行全时段监控和显性表达。
由此带来的结果是：大量潜意识运算留存在“内部”，个人对“为什么会学会某项技能”往往只能给出高度概括的事后回顾。

9.3.2 AI 的计算瓶颈：高维参与与不可逆训练路径

AI 算法在面对海量参数和数据时，也只能使用近似或启发式的优化手段。那些训练过程中出现的随机微小扰动，都会在最终的权重分布中留下痕迹。我们很难通过某种“回放机制”把所有梯度更新都完全可视化或可追溯。

9.4 不可解释的边界：从无意识到超大规模网络

9.4.1 人类的无意识机制：无法语言化的经验积累

如同第三、七章所述，人体有大量学习并不经过自觉控制，语言化只能覆盖一部分意识层面；在无意识或潜意识层面进行的信息整合和记忆关联难以被个体自己察觉，也就谈



不上“解释给他人听”。

9.4.2 AI 超大规模网络：内在表征远超人类范畴

对于超大模型（例如 GPT-4 及之后的超大模型），亿级、千亿级参数形成复杂的向量空间。即使我们能访问所有参数，也难以理解每个参数在整体功能中的作用，更遑论从一一对应的角度去拆解“模型到底为何能回答这些问题”。因此，模型自身的高维性构成了一道不可逾越的理解障碍。

9.5 小结：共性之下的多层次反思

综上，无论是人脑还是 AI，在学习过程中都呈现了高度并行、分布式和动态涌现的属性，导致内部原理不易还原。两者虽然在具体生物或工程实现上存在巨大差异，但在**“黑箱”层面却表现出某种同源性**：即**都是在高维、多因素交互下发展出一种隐式表征体系**，外部观察者难以穷尽描述该体系如何运行与演化。

这为我们理解本文提出的核心命题奠定了事实依据：AI 学习和人类学习的“原理黑箱”**确实有共性**。至于是否完全等同，不同学者会有不同见解，但至少能说明二者共享某些**不可解释性**的根本机制。



第十章 认识论与本体论视角：为什么黑箱不可避免？

在明确了人类与 AI 学习的多维黑箱特征后，本章将上升到更高的哲学层面，借助认识论与本体论的分析框架，探讨为何“黑箱”会成为一种不可避免的普遍现象，以及这对我们理解认知与智能意味着什么。

10.1 主客二分的迷思：主体-客体-互动（SIO）在学习过程中的重构

10.1.1 传统认识论：主体认识客体的线性图景

笛卡尔以来的西方哲学大都将“认识”描述为一个主体（认知者）对客体（被认知之物）进行观测与分析的过程，潜在假设是：**主体与客体在认知之前已相对独立存在**，认识只

是把客体的性质“复制”到主体的心灵中。

10.1.2 SIO 本体论：主体、互动、客体交织的复合体

现代的科学哲学出现了对主客二分的反思，提出了**SIO（主体-互动-客体）**（发明人王德生）的观点：

- **主体**
并非先天固定，而是在与环境及其他主体的持续互动中形成；
- **客体**
也并非绝对独立，而是在被认识或被测量时才“显现”；
- **互动**
是一个关键要素，**知识**本质上是主体与客体在互动过程中形成的SIO模式或表征。

将其应用于学习过程，我们可发现：**人类或 AI 在与环境（数据、任务）互动时，才构建起对外部信息的内在表征**；此时，若要彻底呈现这套表征形成的历程，就必须重现所有交互细节，这在实践中几乎不可能。

10.1.3 对黑箱的启示：不可还原的动态网络

SIO 本体论说明，任何认知或学习都不是静态、孤立完成，而是**主体（学习者）与客体（环境/数据）经由互动**所生产的涌现结果。要想完全解析，需要追溯并模拟整个过程及上下文，而在实际中此过程通常高度复杂、涉及海量微观变动，**因此“黑箱”即是这套动态网络不可被低维度线性描述的体现。**

10.2 现象学与反思：体验的不可言说性与 AI 内部表示的抽象性

10.2.1 现象学对主观体验的强调

胡塞尔（Edmund Husserl）等现象学家聚焦于个体对事物的直接经验，指出**任何经验都带有先行的意向性**，并不可能完全客观化。对“自身体验”本身的把握也需要不断的反思（Epoches），难以在语言中完整描述。

10.2.2 AI 内部表示与数值符号的抽象性

AI 的表征多以矩阵、向量等数值符号形式存在，人类若要解读，往往只能在高层将其映射为视觉化或统计分析。不论是人的主观体验，还是机器的数值表征，都具有**“跨不可见鸿沟”的性质**：体验无法彻底转换成外部语言，数值表征也无法简单地映射为现实语义。

10.2.3 “黑箱”作为转换鸿沟：主观—客观与数值—语义

如果说现象学强调的是“主观体验向客观描述的鸿沟”，那么在 AI 学中也可说有一条“数值表示向人类语义描述”的鸿沟。这种鸿沟就是“黑箱”的核心：**任何复杂认知系统的内在状态都难以被完全投射到外在语言或公式中。**

10.3 知识生成的动态性：从涌现到去中心化

10.3.1 涌现（Emergence）的必然性

无论人脑还是深度神经网络，都展现出**自下而上的涌现**特性：大量低层次单元（神经元或人工神经元）的活动在宏观层面形成新的模式与功能。涌现理论指出，这种整体功能不等于部分功能之和，且不易被还原。

10.3.2 去中心化与无全知者

在分布式系统中，没有一个全知全能的“中央控制”节点，任何看似全局的行为都要依赖庞大的局部交互。对于人类大脑而言，没有一个“元老神经元”可支配全部认知；对于 AI 模型，也不存在一个“管理参数”或“逻辑规则库”来明确记录所有知识。

10.3.3 黑箱是涌现的副产品？

由涌现性与去中心化可推知：无论是大脑还是 AI，都无法轻易“说明自己”，因为根本没有单一控制器或全视角观察者。**黑箱**因此成为涌现系统的一种“副产品”，**当系统功能超过一定复杂度，就不再能被个体成分或单点决策链轻松解读。**

10.4 黑箱是缺陷还是常态？多元视角的争议与融合

10.4.1 工程视角：希望提升可解释性

在实际应用中，工程师与社会各界都希望减少“黑箱”带来的不确定性和不可控性，鼓励发展可解释 AI。法律与伦理层面亦要求重大决策需具备可追责与可理解性，以保护公共利益。

10.4.2 哲学视角：或许黑箱并非完全消极

从哲学层面看，“黑箱”意味着系统的创造性与开放性。人类大脑的不可还原性也使我们拥有灵感与创造力；AI 在深度网络的隐式学习中也可能产生意料之外的创新结果。因此，过度追求绝对的白箱化，反而可能扼杀系统的涌现潜能。

10.4.3 综合：在理解与创造之间寻求平衡

理想状态是：**在尽量保留系统创造力的同时，通过一定手段（如可解释性模块、可视化工具、审计和调试机制）确保安全与可信。**但要在完全理解系统内部与绝对的黑箱之间寻求中道，始终是一个艰难的平衡。

译

第十一章 教育学与实践启示：从人类教学到 AI 训练

“学习”一词本身具有极强的教育学与实践背景。本章将把前述对学习原理黑箱的探讨，引向教育学应用与实践层面：人类是怎样教学的？AI 的训练过程与人类教学能否互相启发？对于“黑箱”我们又该采取什么策略去应对或利用？

11.1 人类教学方法：启发式、引导式与探究式对机器学习的启示

11.1.1 启发式和反馈在学习中的作用

传统上，人类教学依赖课堂讲授，但**启发式教学**和**探究式学习**表明：当学生通过问题驱动与反馈循环主动建构知识，其学习效果更佳。类似地，在机器学习中，**强化学习**（利

用奖励反馈）、**主动学习**（模型主动请求标签）等方法都展现出类似的启发式机制。

11.1.2 引导式学习 VS. 自主学习

教育学强调教师可根据学生水平提供适度“支架”，帮助学生在“最近发展区”内完成挑战，既不过难也不过易。AI 训练中的**渐进式学习**或**课程学习（Curriculum Learning）**正是类似理念：先从简单任务开始，逐渐过渡到复杂任务，使网络更平稳地学习。

11.2 可解释 AI 在教育中的应用：智能辅导与自适应学习

11.2.1 个性化学习与“可解释反馈”

在教育技术中，AI 已被用来实现个性化学习路径、智能测评与反馈。例如“AI 助教”能根据学生答题情况判断其知识薄弱环节并给予针对性练习。但若系统是个不可解释的黑箱，则学生和教师可能会质疑系统推荐的可靠性。可解释 AI 可以让系统输出简要的“理由”或“提示”，从而增强用户信任与配合。

**11.2.2 学生对 AI 的认知：从被动使用到共同“学习”

如果 AI 能够在一定程度上“展示自己的思路”，学生与 AI 的互动就不仅是“接受答案”，而可以是对 AI 推理过程的探究。这种教育方式或可培养学生的批判性思维，让他们了解智能系统的工作方式和局限，同时促进对自身学习过程的反观。

11.3 黑箱透明化的必要与限度：教育实践中的平衡

11.3.1 教学过程中何时需要拆解黑箱？

并非所有学习或教学场景都需要彻底解释内部机理。就像学生学习某些技能时，也不一定需要知道大脑的神经元如何放电，关键在于能否有效掌握知识应用。对于 AI 辅助教学而言，同理：如果系统能可靠指导学生，是否需要将内部权重分布全盘公开？未必。

11.3.2 当黑箱导致错误或偏见时怎么办？

不过，一旦出现明显错误或潜在偏见，教师和学生就需要一个可审查和纠正的通道。这意味着在教育实践中，仍然要构建**审计与干预**机制，以便排查模型错误并迭代优化。这和教师纠正学生偏误的过程有相似之处。

11.4 未来展望：能否通过教育手段“打开”AI与人类的双重黑箱？

11.4.1 互相学习的可能性：人机协同教育

随着人机交互的深入，“机器教人”和“人教机器”有望互相启发：人类学生或许能在与可解释 AI 的对话中更了解自己的学习过程，AI 模型也可以从人类教学反馈中获得更丰富的样本和启发式信息。

或许某一天，我们会看到真正的“人—AI 协同学习共同体”，在教育领域中既能借助机器之长，又保留人的创造性和价值判断。

11.4.2 教育理念与技术手段的深度融合

要走向这样一个未来，还需教育学、认知科学和 AI 工程技术的紧密结合。借鉴人类教学方法优化 AI 训练范式，利用 AI 学习原理反思教育模式，这种双向交融也有望进一步打破“黑箱”，为下一代学习系统奠定坚实基础。



第十二章 技术与哲学的前沿交汇：
走向下一代智能范式

在揭示人类与 AI 学习黑箱共性的基础上，我们不禁要问：有没有新的理论或技术路径，能让我们更接近对学习本质的洞察？本章将对一些前沿思路进行简要探讨，涵盖神经形态计算、多模态大模型、量子计算、脑机接口等领域，同时也包括潜在的社会伦理影响。

12.1 新型神经形态计算：在硬件层面模拟人脑

12.1.1 神经形态芯片与类突触器件

传统计算机架构在处理神经网络时并非最优，神经形态芯片尝试在硬件电路层面模拟突触和神经元的放电行为，可能在能耗和并行效率上更接近生物大脑。

如果这一方向成功发展，未来的 AI 学习或许能够更加逼近人脑工作模式，从而带来更逼真的“原理黑箱”——不仅算法难以解析，硬件层次也可能成为新的难以拆解的维度。

12.1.2 脑机协同计算：融合人类大脑与人工网络

另一个方向是通过脑机接口（Brain-Computer Interface, BCI），将人脑与 AI 网络直接联通，实现联合作业或信息交互。这一设想若成真，或许能让 AI 更“直观”地获取人类大脑的状态，反之也能让人类更深入地调控 AI 的内部机制。但这也带来伦理与安全上的巨大争议。

12.2 多模态大模型与跨学科交叉：新的“学习革命”

12.2.1 多模态融合：视觉、语音、语言、动作合流

将图像、文本、语音、视频、传感器数据等多种模态整合到同一模型中，使得 AI 能在更加丰富的场景下进行理解与生成，被认为是“下一代”AI 的重要趋势。例如 CLIP、Flamingo、BLIP 等多模态模型就展现了强大的跨模态理解和推理能力。

人类的学习本来就多模态，若 AI 实现类似的多模态融合，**黑箱问题**或许也会更加复杂，因为网络内部需要同时处理多种异构信息，而出现更多层面的涌现表征。

12.2.2 跨学科交叉：脑科学、认知心理学与 AI 的深化结合

随着脑科学、认知心理学与 AI 领域的对话愈发密切，我们可能发现更多“学习策略”和“神经规律”，反过来指导新的网络结构设计。如**仿生视觉系统**、**仿生听觉系统**、**仿生脑机交互**等，进一步模糊人类与 AI 的界限，也有望揭示某些隐藏的学习原理。

12.3 量子计算与脑机接口：黑箱尽头的曙光？

12.3.1 量子计算：并非万能的解决方案

量子计算能够在某些问题上带来指数级加速，如因子分解或量子模拟，但是否能彻底消除 AI 或人脑中的“黑箱”？仍无定论。多数量子算法同样需要庞大的量子态空间，测量后坍缩的结果也无法完全复现其中的全部信息。本质上，“信息不可完全外显”的难题依旧存在。

12.3.2 脑机接口：直接读取脑信号是否可行？

一些科幻式的畅想希望通过脑机接口实现对脑信号的精准解读，似乎就能破解人类的“原理黑箱”。然而，目前技术只能获取部分皮层电或血氧变化，距离“实时高分辨读取每个神经元”依旧遥远；且即便能读取神经元放电，还需要庞大的解码模型将其翻译为语义，这又会产生新的“算法黑箱”。

12.4 伦理与社会影响：当黑箱成为决策主体

12.4.1 AI 决策的不可解释性：风险与责任认定

当 AI 被应用于医疗诊断、金融审查、司法量刑等高风险领域，其“黑箱”特性会引发严重道德与法律问题：一个被黑箱决策拒绝贷款或判刑的当事人，如何去质疑或上诉？这需要从制度与法规层面强制一定程度的可解释性或可审计机制。

12.4.2 人类本身的“黑箱”决策也曾被容忍

有趣的是，人类法官或医生在做决策时，同样带有个人偏见、情绪、潜意识等“黑箱因素”。我们通常要求他们在法律与职业操守范围内，至少提供一个合理的解释。或许对 AI 也可采用类似思路：不要求其展示全部原理，但需要能给出“可检验的理由”。

12.5 小结：再思人类与 AI 的共同进化

随着技术的升级与交叉学科的融合，**人类学习和 AI 学习的边界**愈发模糊。尽管黑箱问题不会被轻易消除，但也正是这份“神秘感”激励着科学家持续研究与创造。也许未来我们会发现新的范式，让人类大脑与人工神经网络在更深层次上交融，从而开辟学习与智能的新纪元。



第十三章 总结与展望

经过前十二章的系统论述，我们对“AI 学习和人类学习的‘原理黑箱’之共性”有了较为全面的了解。本章将对全文进行回顾与总结，提炼出主要结论，并指出潜在的研究方向与应用前景。

13.1 研究结论与主要贡献

1. 人类学习与 AI 学习均存在深层次的“原理黑箱”：

- 人类大脑的学习过程因多脑区并行、突触可塑性及潜意识机制而高度复杂，个体自身也无法完全自我解释；
- AI 深度学习模型在高维参数空间中自动涌现特征，训练过程不可逆且难以追溯，导致内部机理难以透明化。

2. 二者“原理黑箱”的共性源于分布式、隐式与非线性涌现：

- 都是通过分布式结构来表征信息；
- 学习中大量特征或规则是隐式形成，不经过线性推理或人工设定；
- 涌现性使得局部机制难以单独解释全局行为。

3. 人类与 AI 学习在生物与社会约束、目标驱动等方面存在差异，但并不否定其在“原理黑箱”上的高相似性。一些哲学视角（如 SIO、本体论和涌现论）更能囊括二者的共同特点。

4. 教育学与可解释 AI 实践为应对“黑箱”提供了部分策略，但绝对透明或白箱化几乎不现实。要在应用中兼顾性能与解释，需要社会、技术、伦理多方协作。

5. 前沿技术与跨学科融合（神经形态计算、脑机接口、量子计算、多模态等）或许进一步加深“黑箱”现象，也可能带来新的突破口。黑箱与涌现的矛盾或将一直伴随人类对智能的探索历程。

13.2 局限性与未来工作展望

1. 局限性

- 本文以理论和文献综述为主，尚缺乏针对具体模型或具体神经回路的实验数据进行精细对比；
- 对哲学、教育学、认知心理学等领域的引介仍属概括式，未做过于深度的个案探讨；
- 由于篇幅原因，无法兼顾所有学派观点，可能存在取舍偏颇。

2. 未来工作方向

- **实验层面**
：开展更系统的跨学科实验，如脑成像与 AI 模型训练日志的联合分析，对同一类学习任务进行人机对照研究；
- **理论层面**

：进一步完善连接主义与涌现论在“不可解释性”方面的建模，探究能否提出新的可部分解释的中层框架；

◦ 应用层面

：在教育、医疗、司法等关键领域，探索“黑箱”下可行的审计与监管机制；

◦ 跨学科交融

：继续深化脑机接口、生物计算等，使人机协作成为一种新的学习与研究范式。

13.3 对跨学科研究的呼吁：打开或理解黑箱的新思路

透过本论文的系统探讨可以看出，学习的“原理黑箱”既是科学前沿的重大挑战，也是一种指引我们向更深层理解迈进的动力。它背后并非一味的“无知”，而是“复杂性”在认知系统中的必然体现。若能在未来的跨学科研究中，引入更多神经科学实验、哲学思辨和技术创新，或许我们能部分开启这一“黑箱”，或者至少学会在黑箱中更好地驾驭和创造。

参考文献

1. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.
2. Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. Wiley.
3. Piaget, J. (1977). *The Development of Thought: Equilibration of Cognitive Structures*. Viking Press.
4. Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.
5. Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, 350(6266), 1332–1338.
6. Olah, C., Mordvintsev, A., & Schubert, L. (2017). Feature visualization. *Distill*, 2(11).
7. Hassabis, D., Kumaran, D., & Summerfield, C. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95(2), 245–258.
8. Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux.
9. Craver, C. F., & Darden, L. (2013). *In Search of Mechanisms: Discoveries across the Life Sciences*. University of Chicago Press.
0. Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.
1. Dehaene, S. (2020). *How We Learn: Why Brains Learn Better Than Any Machine... for Now*. Viking.

2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

3. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117.

4. Marcus, G. (2018). Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*.

5. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.

=====

附录：哲学的虚幻：认识世界（约3000字）

在整个人类思想史中，“认识世界”一直是哲学家、科学家乃至普通人都极为关心的问题。然而，当我们谈及“认识”时，不可回避的是主客体之间那条看似无法跨越、却又在不断被挑战的鸿沟。无论是古希腊哲人对“本原”的探求，还是东方智慧对“空”与“幻”的阐述，都在折射出一个核心命题：****我们认识的世界，究竟在多大程度上是真实的，抑或只是一种幻象？****这种幻象是彻底的虚妄，还是我们探究世界的必经之途？以下内容尝试从不同角度探讨这一“哲学的虚幻”，并在反思中寻求更为明晰的认识之径。

一、从古希腊到现代：世界观的转变与幻象之源

在西方哲学的早期，古希腊思想家就已经注意到感官经验与世界本身之间可能存在的距离。赫拉克利特的“万物流变”宣称我们所观察到的世界永远在变化，无常的表象难以直观反映一个恒定不变的“真实”。而巴门尼德则主张存在只能是“一”，运动与变化都是感觉上的错觉。这种对“表象一实在”二元性的分裂，正是对“幻象”最早的讨论之一。

到了柏拉图，关于“洞穴隐喻”的著名论述，更是将“幻象”推到了哲学思维的核心位置。他认为，普通人的感官之“见”好比在洞穴中只能看到火光投射的影子；只有当人走出洞穴，方能见到阳光下的“真正实体”。洞穴之影即是感官经验与偏见的束缚，也是使人陷于幻象的根源。柏拉图实际上将“理念世界”与“现象世界”区隔，宣称我们在日常经验中所认知的并不是真实的“本质”，而只是近似于影子的模糊表征。

然而，随着近代以来科学的崛起，笛卡尔通过“方法论怀疑”来检验何谓真理时，也不得不先承认一切都可能是“幻觉”，只有“我思故我在”才是不可动摇的。这里的“思”并非对世界本体的直接把握，而是一种对自我意识存在的确认。这进一步指出：**在认识论的最初起点上，人类就面对着对外界幻象无法彻底证伪的窘境。**

到康德时代，“物自体”与“现象”之分，再一次强调了我们永远无法知晓物自体本身；我们所能掌握的，只是感官与先天形式（时空、范畴）相互作用后形成的“表象”。按照他的



说法，世界在我们眼中始终是经过认知结构加工后的“现象”，而那真实本体或许永远处于一层“隔膜”背后。由此观之，**世界对我们而言或许本就是一场被先验范畴修饰过的“虚幻”**。但这并非完全否定认识的价值，只是提醒人类有限性是不可逾越的。

二、东方智慧：空性与幻相

与西方对“实体”与“理念”的执着相比，东方思想在很早就以另一种方式讨论了幻象问题。譬如在印度的佛教哲学中，“空”是极为重要的概念。缘起性空并非绝对虚无，而是一切事物皆依因缘而生、无自性可得；执着事物之“真实不变”只会带来痛苦与偏见。在此意义上，我们所看到的一切现象都是“因缘和合”的暂时显现，故而其本质是“无常”和“幻相”。人若执著于表面现象不放，即落入幻觉与烦恼之中。

中国哲学中的老庄学说也呈现出类似对实与虚的辩证思考。老子说“道可道，非常道”，当我们试图用语言和概念为“道”下定义时，往往已经偏离了“道”的本然状态。庄子更是以许多寓言故事来揭示：人与天地万物常处于化变之间，我们对“物”与“我”的划分常常只是一种约定俗成的幻象。庄子笔下那只“庄周梦蝶”，更是生动地质疑了主客体边界的绝对性，抑或说我们究竟是梦蝶还是蝶梦我们？

在这些东方哲学的讨论中，“虚幻”似乎并不总是负面的贬义，而是一种让人类保持自省、放下执著之道。对万物之虚幻的体认，不仅是对世界本身不可捉摸的映射，也是对人类“心智”或“妄念”之执着的消解。东西方虽然路径不同，但都试图回答：**在我们以为“真实”的表象之下，是否还藏着一个更无可言说的层面？我们能否抓住那永恒不变的实在？**

三、近现代科学：从客观实证到量子迷雾

近现代科学从实验、数学推演等角度大大加深了人类对自然世界的理解。然而，随着物理学尤其是量子力学的出现，科学对世界的认知又一次遭遇了“幻象”的冲击。海森堡的测不准原理、波函数坍缩、多世界诠释等概念都在挑战经典物理时期的“客观、可确定”图景。微观世界中，粒子与波的二象性似乎昭示着：**我们对物理实体的认识也许只是特定观测情景下的某种投影**，而不再是具有绝对客观性的“实在”。

再者，爱因斯坦的相对论指出时空并非绝对独立存在；观测者的状态会影响对时空的测量结果。因果与同时性的问题也在打破我们日常的直觉。这些革命性理论告诉我们：**即便在科学最“严谨”的领域，人类对世界的图景也难免带有“局部性”和“框架依赖”的烙印**。人类往往将本来复杂的事物建构成一个可以验证和重复的模型，然后以此模型去推知世界，却忽视了模型所忽视或简化的一切成分。正如凯恩斯所言，“如果你把经济问题只化简成两三个变量，你就能得到一个漂亮的数学模型，但那很可能是对真实经济的极大扭曲”。

因此，现代科学并没有彻底驱散幻象的阴影，而是以更加精密的方式说明“我们总是栖居在一些理论假设、数学框架和观测限制之中”，从中获得对自然世界愈发精细的“近似把握”，但距离终极真相也许还有无限遥远的距离。

四、认识中的建构主义：幻象如何塑造现实？

从心理学和教育学的角度看，建构主义（Constructivism）主张知识并非纯粹地映射外部世界，而是个体在与环境交互中“主动建构”的结果。皮亚杰、维果茨基等人指出，学习本身是一种内外互动；所谓“真实”并不是通过被动吸收外在信息而获得，而是取决于个体已有的认知结构、社群语言、文化符号的共同作用。

这意味着：**有时我们以为看到的是客观世界，但它却是由我们大脑或社会语境塑造、过滤后才显现出来的“幻象”**。当我们的知识框架改变，世界也“随之改变”。某些研究表明，不同文化背景的人对同一视觉刺激的理解、对同一自然现象的命名都各异，这再次说明了世界与个体之间并无所谓的“零距离接触”。

然而，这种“建构主义”并不等同于宣扬一切都是主观虚构，也并非否定科学研究的客观性追求。它强调的是：**人类对世界的理解在任何时刻都受制于认知结构与社会文化**。这在某种程度上与康德的“先天范畴”相呼应，又与东方的“诸法因缘生”相印证。

五、信息时代的虚拟现实：当幻象越发逼真

随着数字技术、VR/AR（虚拟/增强现实）技术的兴盛，“认识世界”的命题面临新一波冲击。人们通过VR头盔进入一个虚拟构建的环境，感知与交互皆可被高度仿真，甚至可能超越现实世界中某些物理限制。数字世界中的“幻象”不再是简单的假设或批评，而是真切地在视觉、听觉、触觉多模态层面影响着用户。

在此背景下，哲学家与社会学家开始深刻思考：当我们越来越多地将生活搬入数字平台，人与人之间、人与物之间的“关系实在”会不会被取代或淡化？在沉迷于虚拟角色、虚拟场景时，我们是否会进一步迷失于双重甚至多重幻象？或许我们本能追求的是一种最舒适、最能满足欲望的交互，而非对“真实世界”的回返。由此可见，我们对真实的把握更显脆弱，对虚幻的沉溺或许不再只是古人笔下的隽永隐喻，而是切身的时代问题。

六、虚幻并非绝望：走向觉知与谦卑

面对如此多角度的“幻象”论述，我们不免会产生悲观情绪：难道我们自始至终都活在幻影之中？难道付出的努力与探寻毫无意义？事实上，哲学史和思想史的发展也展示了一条更为理性的道路：**承认世界在某种意义上带有幻象成分，未必等同于放弃对真理的追求**。

其一，承认幻象是承认认识的有限性，这会让我们在方法论上保持谦卑与开放。有时我们高举绝对主义、排他性的世界观，恰恰源于对“真实掌控”或“客观把握”的过度自信。若我们知道任何理论都有它的适用边界，任何观测都可能受制于仪器与框架，那我们便会自觉地留出调整与修正的空间。

其二，幻象并不意味着所有东西都无凭无据，而是提示我们应该注重过程与整体。许多科学理论即便是近似或不完整的，也能在特定范围内给予我们有效预测与应用。正如爱因斯坦的广义相对论在宏观尺度上比牛顿力学更“正确”，但在日常生活的低速弱引力条



件下，牛顿力学也不失为“足够好的模型”。我们运用它来落地工程、发射卫星或建造桥梁就已经足矣。

其三，明白“世界可能是幻象”的同时，我们也能更清晰地自省自己的立场和价值观。当意识到我们对世界的认识不可避免地包含主观投射时，会更愿意倾听他人的观点或文化差异。这种对多元价值与视角的尊重，或许正是人类在社会与历史波折中渐渐领悟到的重要一课。

译

七、结语：承认幻象，通向丰富与未知

“认识世界”永远是魅力与挑战并存的主题。哲学家们思考了两千多年，所触及到的仍是冰山一角，但这“冰山一角”已足以让我们对世界之神秘和认识之艰难报以敬畏。从柏拉图的洞穴之影、佛教的“空”、康德的“物自体”、量子力学的测不准原理，再到VR世界的数字幻象——这一路的思想演进，都在警示我们：**我们所拥有的认知常常是对一切“更大真相”的局限性投影。**

然而，正因为我们无法“彻底认识世界”，因此才有无数先贤和后继者在探寻的旅途上上下下求索——哲学家、科学家、艺术家以及普罗大众，都在各自的领域尝试拨开层层迷雾，哪怕最后只能发现新的迷雾，却依旧不断前进。这种探索的过程本身，或许就是人类“认识”最迷人也最根本的意义所在：**幻象并不等同于骗局，而是我们在有限的主体与无限的客观现实之间构筑的“过渡带”，帮助我们在不断修正与拓展中，走向更广阔的未知。**

也许终有一天，某种新的理论或新的智慧会涌现，再度改写我们对“认识世界”的所有想象——或许让世界看起来更加虚幻，又或许让我们对幻象有了更灵活与理智的驾驭方式。站在当下，我们能做的，是以谦卑与求真的态度去拥抱这份未竟之路：**在承认幻象的同时，用更丰富、更多元、更深邃的思维和实践，去拥抱每一寸真实。**

SIO 10 学习 3 AI 2

SIO · 目录

上一篇 · ChatGPT+DeepSeek协作形成创新迭代模型

