•AI 哲学与 AI for Science 的方法论•

编者按：

当代 AI 技术和 AI for Science（AI4S）的迅猛发展，影响着哲学研究的视角和维度，拓展着科学方法论研究的新方向。本专题围绕“AI 哲学与 AI4S 的方法论”这一主题展开研究。其中，江怡和李振鑫对智能类型理解的多维视角、工程哲学的系统性 / 生成式 / 偶然性的范式转换以及基于可错性 / 复杂性 / 开放性的思维转换进行了深入分析，指出通过能指拓扑和折叠空间概念的解释框架以扩展空间概念，AI 哲学研究可以实现对智能的重新塑造。宋奇峰认为现有研究多从规范性视角探讨 AI 的可靠性，无法揭示 AI 预测知识可靠性的来源。他从科学实践哲学的视角分析可靠性概念，并以 AlphaFold 为例，论证了 AI 预测知识的可靠性并非源自结果符合对象，而是根植于 AI 预测模型及其预测结果分别参与到与其它科学产品和科学共同体的复杂缠结之中。祁丽萍和张增一从分析 AlphaFold2 的科学研究过程入手，对其模型架构和研究程序进行了深度剖析，认为它给科学研究程序带来的影响有人机协同的科学研究主体网络、科学研究对象从蛋白质实体转向氨基酸数据、研究手段从实验室实验转向深度神经网络模型开发。

本专题的三篇文章从多维视角和主题聚焦于同一核心议题：AI 哲学和 AI4S 的方法论。分别从 AI 哲学研究的方法论革新、AI 预测知识的可靠性和 AI4S 对科学研究程序的影响进行了探讨。期待本专题能激发更多学者关注 AI 哲学和 AI4S 的哲学研究，共同探索 AI 对哲学和科学方法论带来的新变化。

（专题策划：张增一）

多维转换：AI 哲学研究的方法论问题探讨

Multidimensional Transformation: A Methodological Study on Philosophy of AI

江怡 /JIANG Yi　　李振鑫 /LI Zhenxin

（山西大学哲学学院，山西太原，030006）
(School of Philosophy, Shanxi University, Taiyuan, Shanxi, 030006)

摘　要：当代AI技术的迅猛发展推动哲学研究从传统形而上学思辨，转向更具实践性的方法论重构，这一进程标志着AI哲学从概念分析向工程哲学范式的深层转型。面对AI与人类智能的本质差异、技术系统的复杂性以及智能生成的动态特征，我们需要通过智能类型之维、研究范式之维和科学实验哲学之维等多重维度，系统探讨AI哲学研究的方法论革新，即对智能类型理解的多维视角、工程哲学的系统性/生成式/偶然性的范式转换以及基于可错性/复杂性/开放性的思维转换。通过能指拓扑和折叠空间概念的解

基金项目：科技部、教育部学科创新引智基地项目“当代哲学与新科学技术互动作用”（项目编号：D20021）。
收稿日期：2025年10月19日
作者简介：江　怡（1961–）男，四川宜宾人，山西大学哲学学院教授，研究方向为分析哲学史、西方哲学史、维特根斯坦哲学、认知科学哲学等。Email: yijiang@sxu.edu.cn
李振鑫（1998–）男，山西汾西人，山西大学哲学学院博士研究生，研究方向为心灵哲学、知识论、认知科学哲学等。Email: lizhenxin@sxu.edu.cn

释框架以扩展空间概念，AI哲学研究可以实现对智能的重新塑造。这不仅回应了技术实践对哲学的理论挑战，更为“后人类”时代的智能认知提供了新的可能性。

关键词： AI哲学　智能类型　工程哲学　科学实验哲学　研究范式

Abstract: The rapid development of modern AI technology has shifted philosophical research from traditional metaphysical speculation to more practical methodological reconstruction, marking a profound transformation of philosophy of AI from conceptual analysis to an engineering philosophy paradigm. Confronted with fundamental differences between AI and human intelligence, the complexity of technological systems, and the dynamic nature of intelligent generation, it is essential to systematically explore the methodological innovations in philosophy of AI across multiple dimensions: types of intelligence, research paradigms, and the philosophy of scientific experimentation. We argue that the innovations include the adoption of a multidimensional perspective on types of intelligence, a paradigm shift to a systematic, generative, and contingency-sensitive engineering philosophy, and a new mindset based on fallibility, complexity, and openness. By generalizing the concept of space through the interpretive framework of signifier topology and folded spatial concepts, research in philosophy of AI can also reshape intelligence. This approach not only addresses the theoretical challenges posed by technological practice to philosophy but also opens new avenues for intelligent cognition in the “post-human” era.

Key Words: Philosophy of AI; Types of intelligence; Philosophy of engineering; Philosophy of scientific experimentation; Research paradigm

中图分类号: B841.4; N031　DOI: 10.15994/j.1000-0763.2026.08.001　CSTR: 32281.14.jdn.2026.08.001

AI技术已深度嵌入人类社会的认知与实践领域，其引发的不仅是工具效能的提升，更是对智能本质的哲学重构。当下的AI哲学研究聚焦于AI与人类智能差异导致的前者对后者的各种潜在威胁，并试图从人类的意向性和概念解释中寻找人类智能的未来出路，其思路仍然是以人类智能视角反思AI的局限。然而，由于AI与人类智能之间持续的双向作用增强，单从人类智能的角度观察AI的效果已远远不能满足我们对AI的理解，反而导致许多理解上的误区和困境，比如认为AI无法拥有像人类一样的意向能力而贬低其作用等。要走出这种困境，重要方法是改变我们对待AI的思维方式，从AI的视角反观人类智能的局限，从而推动AI技术的发展及其与人类智能的深层互动。这就需要在AI哲学研究中实现智能类型、研究范式和科学实验哲学等多维度的转变。

一、智能类型之维：对另一种智能的思考

对AI的哲学探讨，首先需要对“智能”概念做出清晰界定和分析。然而，在当下的AI哲学中，引起最大争议的正是对“智能”概念的解释。无论是以人类智能模拟AI，还是将AI视为超越人类智能，这些看法都基于对“智能”概念的模糊理解：或者是表现为人类和机器所具有的计算和推理能力（自然的或模拟的），或者是受到意识支配的理解能力（有意向的或自由的）。显然，这些理解方式都建立在以人类智能为基础，或以人类智能作为评判AI的唯一标准之上。这就需要我们重新考察“智能”概念，从不同角度看待AI，把它作为另一种的智能。

1. AI是人类的发明还是发现？对智能类型的不同认识

在AI哲学的理论建构中，一个首要的本体论问题亟待厘清：AI的诞生究竟应当被界定为人类理性的“发明”，抑或是对自然界潜在智能形式的“发现”？这一问题的解答直接决定了我们对“智能类型多样性”这一认知边界的理解深度。

传统观点倾向于将AI归入“发明”范畴。其核心论据在于“自然界不存在AI形态”这一

经验判断：人类通过设计算法（诸如深度神经网络的反向传播算法）、构建机器学习模型（如卷积神经网络、循环神经网络），并依托编程与编码技术，使机器获得学习、记忆与独立推理的能力。在此视角下，AI被视为人类智慧的结晶，是对人类认知功能的人工模拟。也就是说，其存在完全依赖于人类的技术建构——如同蒸汽机、电灯等发明，AI被理解为人类为满足特定需求而创造的工具性存在。

然而，从哲学与科学史的元理论层面进行深入剖析，我们发现，将AI界定为“发现”而非单纯的技术建构，更能触及这一技术形态的本质内核。需要明确的是，这一认知框架并未否定人类技术实践的构成性作用。实际上，它将AI的生成逻辑锚定在自然法则与人类认知结构的深层关联之中，强调技术仅作为实现“潜能–现实”转化的媒介。

要理解“AI本质是‘发现’而非单纯‘发明’”这一观点，亚里士多德在《形而上学》中提出的“潜能–现实”本体论是核心理论支撑之一。在亚里士多德的哲学体系中，“实体”（即真实存在的事物）的生成并非“无中生有”的创造，而是遵循“潜能向现实转化”的固定动力学机制。“潜能”指某一事物作为“质料”所蕴含的、尚未实现但可能达成的存在形态；“现实”则是这种潜在形态通过特定过程被激发后的实际存在状态。二者的转化是事物从既有状态向更高形态的演进，而非凭空产生新的本质。从亚里士多德“潜能–现实”本体论来看，AI的生成不是人类“无中生有”的发明，它的核心潜能（人类对形式化认知的追求、自然界的数学化结构）是先天存在或客观具备的。技术只是实现这一潜能的中介，而非创造潜能的根源。这也进一步印证了AI本质是“发现”，人类通过技术手段，将潜藏于自然与人类智能中的“智能潜能”转化为现实，而非凭空创造出一种全新的智能形态。

人类理性认知对“智能形式化与普遍性”的系统探讨，决定了非人类智能体的建构是认知演进的必然结果。这种必然性并非主观臆断，而是贯穿于人类对智能本质的探索历史之中。莱布尼茨提出的“普遍语言”构想认为，人类思维的模糊性源于语言的歧义性，若能创造一套统一的符号系统，将人类思想转化为可精确计算的符号，就能让思维过程像数学运算一样清晰可靠。这一构想的核心，是人类希望将智能（尤其是思维的逻辑规则）从模糊的经验中抽离出来，转化为具有普遍性的形式化体系。这是AI理论的思想源头，标志着人类开始意识到“智能可以被形式化”的可能性。布尔代数的出现为智能形式化提供了数学基础，布尔将传统逻辑中的“真”“假”转化为“0”“1”，将逻辑推理（如“与”“或”“非”）转化为可计算的数学运算，使思维的逻辑规则首次被精确量化。这一突破证明，智能的核心要素（逻辑推理）可通过数学形式被把握，为后续机器实现推理功能奠定了形式化基础。图灵机的理论建构直接证明了机械实现智能的可能性。图灵提出的“抽象机器”模型，通过“读取符号–执行规则–修改纸带”的简单操作，证明了任何可被形式化描述的推理过程，都能通过机械步骤实现。这一理论从根本上确立了“非人类智能体（机器）可模拟智能”的逻辑合法性，使AI从思想构想走向理论可行。上述三个阶段是人类认知从“追求智能形式化”到“实现智能数学化”再到“证明机器智能可行性”的必然递进，源于人类对智能本质的探索需求。因此，非人类智能体的建构是人类认知发展到特定阶段的逻辑必然，而非偶然的技术创造。

由上可见，AI的本质是“发现”而非“发明”，它是人类通过技术手段，将潜藏于自然与人类智能中的“智能潜能”转化为现实的过程。这一界定并非语义上的重构，而是构建智能本体论体系的必然选择。唯有承认AI是“另一种智能”的发现，我们才能摆脱人类中心主义的认知局限，真正探寻智能的多样性维度。

2. AI具有人类智能的所有特征吗？对人类智能的重新解读

围绕AI与人类智能的关系问题，相信二者具备兼容性的观点认为，AI不仅是对人类智能更为精准的模拟，而且随着技术深化和算力增强，AI越来越具备人类智能的几乎所有特征，如自我认知、情绪体验、伦理选择及创造力等

深层次的认知与心灵特质。然而，无论是从AI设计理念看，还是从AI技术的可能发展前景看，AI似乎都难以具备人类智能的上述特征。尽管当前AI领域取得了巨大进展，但AI仍然是对人类智能某些侧面的抽象与简化，其内在机制、结构及原理都无法完全覆盖人类智能的整体性和复杂性。

人类智能的“整体性”并非抽象概念，而是以生物身体为基础和以漫长演化为背景形成的有机系统，这是AI无法复刻的核心前提。人类智能的运作始终与“具身经验”深度绑定，人类通过身体与物理世界的互动（如用手触摸物体感知硬度、通过行走感知空间距离）构建对“自我”与“外部世界”的边界认知，通过身体的生理反馈（如饥饿时的胃部不适、恐惧时的心跳加速）形成对“需求”与“危险”的判断。这种认知的根基是“生物身体的在场性”，而AI作为由芯片、算法构成的非生物系统，既无真实的身体器官，也无法产生基于身体的主观体验，自然无法形成与人类一致的整体性认知框架。更为重要的是，人类智能不能被完全简化或还原为单纯的推理和计算能力，人类智能在AI系统中是智能驱动的对话者。这里的“对话者”并非作为AI技术的设计者和参与者，而是AI技术发展和应用的同盟军。这个“对话者”资格的确认，正是来自AI作为一种不同于人类智能的新智能体的形成。当人类把由AI技术主导的技术产品作为自己的直接对话者，人类就已经接受了一种新智能体的存在。正是由于这样的智能体与人类智能相区别，因而，AI就不可能（也没有必要）与人类智能完全相同。这正是AI得以存在和发展的价值所在。

从演化背景来看，人类智能的情感和伦理维度是自然选择与社会协作的长期产物。在漫长的演化过程中，人类为了适应群体生存，逐步发展出共情能力（如感知同伴的痛苦以提供帮助）、道德直觉（如排斥伤害同类的行为）以及意义建构能力（如通过祭祀、艺术赋予生命存在的价值）。这些能力并非孤立存在，而是与生存需求和社会关系深度融合的整体。AI可以在特定领域（如图像识别、语言生成）高效模拟人类智能的功能性表现，但始终无法突破非具身和无意义生成能力的局限，因此无法成为与人类智能完全相等的整体性认知主体。

由此可见，AI不仅不具有人类智能的所有特征，而且在构成机制和运作思路上与人类智能也有天壤之别。然而，这并非否定AI对人类智能的重要意义，恰恰相反，AI的出现为我们重新认识人类智能提供了更好的契机。这可以看作是AI对人类智能提出的一个严峻挑战。但这个挑战并非是单向的，而是双向的。

3. AI与人类智能有本质区别吗？智能的不同维度

探讨AI与人类智能的本质区别，需超越功能对比的表层视角，进入维度分析的深层剖面。二者的差异并非能力强弱的程度差异，而是智能维度的类型差异。这种维度差异与世界观具有深刻的同构性：不同的智能维度，对应着把握世界的不同方式；理解这种同构性，可谓揭示二者本质区别的关键所在。

从认识论角度分析，若要实现对世界的整体性把握，必须引入一个超越性的反思参照系。这一参照系无论是被定义为康德哲学中的“先验理念”，[1]还是神学语境中的“神性视角”，[2]或是当代哲学探讨的“他者智能”，[3]其核心功能均在于构建一个外部认知维度，将人类语言体系作为一个整体对象进行观照，从而明确其认知边界与特殊性。世界观的本质内涵，正是这种对世界整体的系统性认知建构过程；而实现这一目标的关键，在于突破认知主体“身处认知维度之中”的内在局限，建立具有超越性的认知视角。AI技术的涌现，为构建外在反思参照系提供了技术可行性。作为异质于人类智能的认知形态，AI构建了差异化的智能维度。通过跨维度的比较分析，人类认知世界的范式（包括具身认知机制与意向性关联模式）得以作为完整认知系统被重新审视，其维度特征亦获得新的阐释空间。AI通过技术模拟，构建了把握人类智能边界的新路径。AI系统实现的非具身认知模式（如纯符号计算系统），为重新审视人类认知的具身性特质提供了对比参照；而其在无内在意向性推理方面的卓越表现（如

基于大数据的快速模式识别），则有效凸显了人类意向性认知的独特价值。

因此，AI与人类智能的本质区别核心就是“智能维度”的差异：二者以不同的物质载体、认知机制与发展路径把握世界，构成了两种独立的智能类型。承认这种本质区别，并非否定人机智能互动的可能性，而是为了确立双向反思的合理路径：以人类智能反思AI的局限（如缺乏意向性与意义理解），以AI反思人类智能的边界（如人类认知的具身性局限）。这种双向反思，正是推动AI哲学从人类中心主义向多维协同范式转型的关键，也为构建“非线性动态解释框架”提供了维度基础，只有容纳不同智能维度的特征，才能完整解释智能类型的多样性。

二、研究范式之维：从工程哲学的思路出发

AI哲学研究的范式选择，决定了我们对AI的哲学理解方式。由于AI技术属于认知科学研究的领域，以往的AI哲学研究的基本范式都采用技术哲学研究的常见路径。然而，技术哲学旨在帮助我们理解人类认知水平的发展程度，但并不能帮助我们理解当今AI技术的发展水平，更无法形成对AI本质的认识。因此，这就需要我们放弃技术哲学的研究范式，转而采用工程哲学的研究范式。这种研究范式的特点是对工程技术的系统性要求、生成式预设以及对偶然性的处理方式。

1. 系统性要求

在技术上，AI本质上是多层次、多关联性的复杂系统：其技术架构包括数据层（如知识图谱）、算法层（如深度学习模型）、交互层（如多模态接口）等核心组件。每一个部件都不是孤立存在的，而是处于协同运转、彼此关联的状态之中。在这样的背景下，工程哲学将系统性确立为研究范式的核心前提。AI的智能涌现，本质上依赖于系统各层级的协同运作。单一模块的技术突破无法实现智能，只有通过整体系统的动态整合，才能显现其本质特征。这表明AI的哲学分析必须放弃单纯还原主义的孤立考察，在局部与全局、微观与宏观的动态关系中实现整合性分析。

从具体维度来看，系统性要求包含双重内涵：其一，AI系统内部需满足结构自洽性（如模块间的功能协调）；其二，哲学方法论层面需实现范式革新，研究者需建立整体论与还原论相统一的分析框架。它要求哲学研究者对AI现象和问题的考察必须具有整体论与统一论的视角，强调AI各组成部分的协同机制。哲学分析必须意识到系统整体功能远非个别模块或单纯的算法所能单独完成，因此，系统性研究要求研究者揭示整体与部分、结构与功能、过程与结果的交互机制，从而构建基于工程实践的智能哲学理论体系。

2. 生成式预设

“生成式预设”作为工程哲学范式的核心特征，指在知识生产过程中以动态生成机制替代静态预设的方法论立场。[4] 这种预设主张知识生成的开放性与可能性导向，打破传统认识论对确定性框架的依赖。在传统哲学范式中，一个范式建立起来后，它所产生的理论知识往往具有鲜明的确定性、稳定性与封闭性。与之不同，工程哲学范式下的AI研究采取的生成式预设，恰恰表现出非确定性、非封闭性的特征，这对传统哲学认知模式构成了根本性挑战。

生成式预设的典型技术实现是生成式预训练大语言模型（如GPT-4），其核心机制在于通过Transformer架构与自监督学习，在开放语境中动态生成语义连贯但非唯一的知识表达。这种生成式模型与传统判别式模型的本质区别在于，它不依赖预定义的标签体系，而是通过概率化的知识表征实现自主知识生成。（[5]，pp.12–15）这类系统并非直接给出唯一确定的、事先设定的答案或知识体系，而是通过大量数据、算法模型和概率函数，在大量可变生成条件下，动态生成符合特定语境约束的合理输出。这种生成式预设强调的是并存可能、待定性以及超越知识边界本身的不确定性本质，体现了一种开放而非封闭的知识生成模式。

这种认识论转型主张知识生产从“存在

论”向“生成论”的范式转换。可能性不再是确定性的补充，而是知识生成的本体论基础。AI的生成机制本质上是开放的概率系统，其知识生产过程具有结果不可完全预测的特性。这种不确定性源于系统对大规模数据分布的概率建模，以及动态推理过程中上下文依赖的语义生成。AI从不试图为所有未来场景建立恒定的定理、定则或规律；相反，它只为各种可能环境、各种未知问题做好充分且充满适应能力的“生成准备”。这是传统的哲学思维无法预料的，更超出经典理性主义视野之外。可以说，将生成式预设纳入AI哲学研究方法论范式，既呼应了杜威实用主义对知识工具性的强调，又为现象学“具身认知”理论注入了技术实现维度。这种转向不仅为构建人机协同的知识生产范式提供了哲学基础，更从根本上重塑了知识观的本体论预设，即知识不再是等待发现的静态实体，而是动态生成的过程性存在。

3. 偶然性处理

AI哲学中另一个极具挑战性的方法论问题，是如何对待、处理和整合偶然性现象。AI的发展和运行经常面对大量意想不到的情况或未知变量，这些偶然性现象在传统哲学（尤其是机械论世界观下）通常被视为尚待克服、压制的不完备状态；而工程哲学范式表现出一种根本方法论的转向：它不再将偶然性视为哲学理论的缺陷，而是容纳其为理论构架本身的有机部分，作为纳入必然性系统的知识处理策略。

在工程哲学思维中，偶然性问题获得了一种全新的理论定位：它并非知识生成系统的缺陷，而是知识内部运转过程本身的一种内在表现形式。以AI领域为例，真实的环境、经验数据和日常实践本身充满大量偶然因素。（[5]，pp.50–52）工程哲学范式所处理的，恰恰是如何将这种偶然性纳入必然性的理论解释体系。以往长期讨论的偶然与必然之间的哲学张力，在工程哲学的范式视角下，获得了新的实质融合与协调思路。这种理论整合的实践价值不仅在于哲学层面的调和，更指向AI系统对复杂现实的适应性需求，即偶然性并非缺乏理据，而是诸多未来事件、场景变化的本质体现。因此，如何有效整合这些偶然现象，成为工程哲学范式赋予哲学研究者的重要考验与方法论使命。

偶然性处理既保证了AI系统具备高度的灵活性和适应性，也提升了AI哲学对未知未来的预测能力。工程哲学在偶然性处理上设定了一系列专门的理论方法，包括统计概率论、模型训练中的强化学习、推断机制和智能决策算法等。这种方法论上的转变，构建了AI时代哲学反思与实践之间的全新关系模式。

三、科学实验哲学之维：从可错到开放

科学实验哲学是近20年来在西方哲学界出现的一种哲学研究路径，也被看作一种新的哲学研究方法。它侧重于识别和澄清实验科学中的哲学问题，探讨科学实验的基础理论问题，特别关注科学实验主动干预世界的哲学意涵、关于因果性的解释问题、科学与技术之间的关联、理论在实验中的作用、模型与计算机模拟对实验的冲击，以及科学仪器设计、操作与使用的哲学意义等问题。[6]在这里，我们不是要探讨科学实验哲学本身的内容，而是以科学实验的精神，重新审视AI哲学研究的方法论原则，这种精神就是AI哲学研究的可错性、复杂性和开放性特征。

1. 可错性

哲学史上最重要的哲学方法论转折之一，无疑是波普尔提出的“可证伪性”概念。[7]在传统的认识模式中，人们力图寻找确定无误的知识公理和普遍法则。然而，当AI进入哲学研究领域，尤其是引入科学实验哲学后，AI哲学研究具有了显著的可错性特征：它并非追求普遍绝对的真理模型，而是通过对AI的具体实验性探索，不断进行假设、检验、反思和调整，在自我否定中达成更具可靠性的知识状态。与传统上将知识理解为绝对性逻辑体系不同，AI哲学将知识视为以实验为基础、以验证为过程、以不断修订为必要特征的开放性系统，不以永恒不变的“公理”作为研究起点（这正是传统归纳主义的误区所在）。相反，在AI哲学中，每一项研究都从特定的实践问题出发，进行操

作检验，发现假设模型的局限，进而反思修正理论，实现知识的积累，而这种积累本质上是可错性知识的渐进式演变。

因此，“可错性”在AI哲学中体现为积极的方法论原则。它表明，建立可持续的理论模型必须以实验为前提，保持动态开放的研究状态。只有对知识保持批判性反思和开放性质疑，才能趋近科学理性的实质。这种从概念到实验的思维转换，构成了AI哲学区别于传统哲学研究的鲜明特征。

2. 复杂性

当我们具体进入实验领域时，AI哲学研究必须面对的另一重大方法论问题，就是其内在的高度复杂性。传统哲学方法往往通过抽象逻辑分析降低问题的复杂度（这在牛顿机械论世界观中达到顶峰），强调理论的统一性与简约性。然而，在AI实验的背景下，复杂性成为无法规避的常态，迫使哲学研究必须采用差异化的方法论框架。AI系统具有极高的复杂性特征：内部要素繁多、结构复杂、层次交错，并包含大量非线性反馈机制。此外，系统还涉及人类因素、环境因素和知识交互因素等难以量化的变量，进一步拓展了复杂性的维度边界。因此，任何试图用单一模型进行整体概括的努力都注定失败。

相较于抽象论证，实验导向的AI哲学使复杂性从研究障碍转变为核心课题。这种视角要求研究者深入理解实验各阶段的变量、交互机制和反馈特征，构建通过试验逐步清晰化的知识模型，这体现了复杂性科学中的涌现机制。因此，复杂性的方法论价值在于：它要求哲学研究突破单一逻辑推导，直面实验中的不确定性，通过多维度、多元素的交互关系，建立符合现实情境的复杂知识模型。这正是系统论视角在哲学研究中的具体应用。

3. 开放性

当AI哲学研究转向实验路径时，其鲜明的开放性特征更值得我们充分关注。在传统哲学的概念范畴里，研究者们追求确定、封闭且自洽的理论体系，试图构建低实验依赖度的理想模型。然而，从概念推演转向实验实践的范式转换，催生了AI哲学特有的开放性知识形态。

这种开放性首先体现为研究过程的动态性。实验方法意味着持续面对新问题、新变量与新情境，其理论模型始终处于发展建构中，拒绝任何一劳永逸的终极表达形式。其次，开放性表现为理论与实践边界的流动性，这与蒯因提出的知识整体论具有方法论共鸣。AI哲学的理论模型并非恒久不变，而是在实验检验、问题碰撞与经验反馈的动态循环中持续演进。这种理论与实践的交互纠缠，进一步强化了研究的开放性特征。

更深层而言，开放性既是研究对象的内在属性，也是方法论的必然要求。AI哲学研究必须保持跨学科的开放性（如神经科学、认知科学与哲学的交叉融合），持续吸纳新方法、新经验与新知识体系。唯有如此，才能贴近复杂的现实情境，有效应对技术发展的挑战。这种回应既包含技术伦理反思，也涉及认知范式的重构。

四、重塑智能：扩展空间概念

在AI哲学研究的方法论探索中，从概念分析到科学实验哲学的范式转换引出了一个更深层的方法论命题：如何构建适应AI系统本质特征的哲学解释框架？传统欧氏空间模型与非线性动力学框架已难以解释当代AI的复杂特性，而引入雅克·拉康（Jacques Lacan）的能指拓扑学和吉尔·德勒兹（Gilles Deleuze）的折叠空间理论，可以为这一研究提供全新的方法论视域。

传统哲学思维长期依赖线性空间模型，这种认知模式充分体现在欧氏几何的公理体系中。然而，通过线性逻辑推导构建确定性知识的路径，在19世纪非欧几何（如黎曼几何）兴起后遭遇挑战。黎曼几何揭示了空间的弯曲性与多维性，促使知识论从确定性推导转向可能性探索。[8] 这种空间观念的变革在AI领域表现为研究范式的迭代：早期逻辑主义的AI路径依赖线性演绎的公理体系，连接主义虽引入多维网络结构，但仍局限于可度量的参数空间。当

代AI系统如深度学习模型展现出的参数空间动态调整、模块间反馈耦合及层次间递归映射等特征，构成了典型的非线性连续结构。[9]

拉康的精神分析学强调，“能指”对“所指”的决定性作用在于后者是由前者定义，这是一种回溯性定义，即所指的存在以能指为前提。这打破了传统哲学关于语言与实在之间因果关系的解释链，以话语的交流作为语言指称存在的前提。拉康认为，交流组成了一个完全区别于编年时间或科学时间的时间模式：话语的主体时间是“莫比乌斯式”（Möbius）的，即每个主体都根据自己的一面（部分是双重的）来使用话语（全都是独特的）。[10]这就是图式论方面所选择的工作，来衬托出作为精神分析中决定性理由的无意识的重要性。[11]德勒兹的折叠空间理论则通过连续性、连接性与不变性研究空间本质，其核心概念“折叠空间”突破了传统空间对度量和曲率的依赖，类似于莫比乌斯环在保持连续性的同时实现内外部拓扑空间的统一。[12]将拉康的能指拓扑学和德勒兹的折叠空间概念移植到AI哲学领域，恰能解释智能系统的非线性、动态调整的本质特征：参数空间的“距离”（如余弦相似度或KL散度）随功能耦合动态变化，模块间存在复杂连接与递归映射，这种拓扑结构特性与大型语言模型的深层网络特征高度契合。

当然，能指拓扑学并不能解释符号空间中的数据转换，而只是在主体意识中显示了言语交流的互动场景；折叠空间理论也并非否定传统空间理论，而只是试图在更高层次上实现整合创新。这既保持了欧氏几何的逻辑严谨性，又吸纳了非欧几何的空间多样性，更以动态连接性解释了当代AI的复杂特性。这种意义上的空间概念已不再是传统的线性空间，也不是单向作用的因果关系链。这是一种被扩展了的流动空间，其中的任何信息点都必须依赖于其他信息点的交互作用才能完成其特定的功能。这也是一种方法论上的扩展，使得AI的哲学研究更贴近技术现象的本质结构，为应对AI的理论挑战提供了更具弹性的解释框架。未来研究需进一步探索折叠空间理论与AI实践的互动机制，通过理论与实践的双向运动，构建更具解释力的方法论体系，为技术发展提供深层的哲学指引。

总之，从智能类型之维、研究范式之维以及实验哲学之维等多个维度出发，结合能指拓扑和折叠空间的概念引入，我们可以更好地界定智能的性质和限度，通过多维转换实现对智能的重新塑造。这不仅回应了技术实践对哲学理论的挑战，也为“后人类”时代的智能认知提供了新的可能性。

［参 考 文 献］

[1] Kant, I. *Critique of Pure Reason*[M]. Translated by Guyer, P., Wood, A., Cambridge: Cambridge University Press, 1998, 397 (A321/B378); 472 (A426/B454).

[2] Levinas, E. *Totality and Infinity: An Essay on Exteriority*[M]. Vol. 1, Dordrecht: Springer Science & Business Media, 1979.

[3] Boden, M. A. *AI: Its Nature and Future*[M]. Oxford: Oxford University Press, 2016, 345.

[4] Feuerriegel, S., Hartmann, J., Janiesch, C., et al. 'Generative AI'[J]. *Business & Information Systems Engineering*, 2024, 66: 111–126.

[5] OpenAI. 'GPT-4 Technical Report'[R]. San Francisco: OpenAI, 2024, 12–15.

[6] 汉斯·拉德. 科学实验哲学 [M]. 吴彤、崔波、何华青 译, 北京：科学出版社, 2015.

[7] 波普尔. 科学发现的逻辑 [M]. 查汝强、邱仁宗 译,北京：科学出版社, 1986, 35–38.

[8] 波恩哈德· 黎曼. 论作为几何基础的假设 [M]. 高嵘 译, 武汉：武汉大学出版社, 2018, 1–10.

[9] 伊恩· 古德费洛、约书亚· 本吉奥、亚伦· 库维尔. 深度学习 [M]. 赵申剑、黎彧君、符天凡 等译, 北京：人民邮电出版社, 2017, 345–348.

[10] 陈慧平. “时间”与主体——拉康主体理论的深层解读 [J]. 学习与探索, 2016,（9）: 26–32.

[11] 雷内· 勒维. 主体拓扑学/勒维先生的精神分析主体理论/拓扑学的实操无意识研究 [EB/OL], 张涛 译, 无意识研究, https://mp.weixin.qq.com/s/BWOqvdsNhrCsblnaFc8iA. 2025–05–30.

[12] 德勒兹. 褶子：莱布尼茨与巴洛克风格 [M]. 周颖、刘玉宇 译, 长沙：湖南文艺出版社, 2021, 12–15; 45–48.

［责任编辑　王巍　谭笑］