

生成式人工智能时代认识主体观的三重类型及其互补制衡

李昊, 刘正平

昆明理工大学马克思主义学院, 云南 昆明

收稿日期: 2026年4月13日; 录用日期: 2026年5月5日; 发布日期: 2026年5月18日

摘要

生成式人工智能的迅速发展, 使主体问题重新成为认识论、科学技术哲学与技术治理研究中的焦点。与传统人工智能主要承担识别、检索和自动化处理功能不同, 生成式人工智能已经深度介入文本生产、知识组织、问题生成、论证展开和意义建构等环节, 从而使谁在认识如何认识何为认识等经典哲学问题在新的技术背景下被重新激活。围绕这一现实变化, 现有研究逐渐形成三种具有代表性的认识主体观: 意向性认识主体观、交互性认识主体观与分布性认识主体观。意向性认识主体观强调主体是否具有关于世界的意向性结构, 并试图以理解或语义指向作为主体资格的核心判准; 交互性认识主体观将主体理解为在具体协同实践中形成的角色结构, 强调人机互动中的认识代理、责任分工与规范安排; 分布性认识主体观则把主体问题推进到平台、制度与网络结构层面, 突出生成式人工智能嵌入分布式认知网络后所形成的系统性主体效应。三种主体观分别把握了生成式人工智能时代认识活动的不同维度, 但任何单一立场都难以完整说明现实中的主体重构。基于此, 本文主张以互补制衡的方式理解生成式人工智能时代的认识主体问题: 既拒绝把人工智能神秘化为具有独立意志的新主体, 也反对把它简单还原为完全中性的工具, 而是在坚持人的主体地位与责任地位的前提下, 承认算法系统、平台结构与制度安排对认知活动的深度塑造作用。这样的综合视角, 有助于突破人工智能是不是主体的抽象二元争论, 更准确地把握当代知识生产方式变化中的主体位置、责任边界与规范含义。

关键词

生成式人工智能, 认识主体, 意向性, 交互性, 分布性, 互补制衡

The Threefold Typology of the Epistemic Subject in the Era of Generative AI: Complementarity and Checks and Balances

Hao Li, Zhengping Liu

School of Marxism, Kunming University of Science and Technology, Kunming Yunnan

Abstract

The rapid development of generative artificial intelligence (AI) has brought the question of the “subject” back into the spotlight within the fields of epistemology, the philosophy of science and technology, and technology governance studies. Unlike traditional AI—which primarily performs functions related to recognition, retrieval, and automated processing—generative AI has become deeply involved in processes such as text generation, knowledge organization, problem formulation, argumentation, and meaning construction. Consequently, classic philosophical inquiries—such as “who knows?”, “how is knowing achieved?”, and “what constitutes knowing?”—have been reanimated within this new technological landscape. In response to this shifting reality, existing scholarship has gradually coalesced around three representative conceptions of the epistemic subject: the intentional subject, the interactive subject, and the distributed subject. The intentional view emphasizes whether a subject possesses an intentional structure regarding the world, attempting to establish “understanding” or “semantic directedness” as the core criterion for subject status. The interactive view conceptualizes the subject as a role structure forged within specific collaborative practices, highlighting the dynamics of epistemic agency, the division of responsibility, and normative arrangements within human-machine interactions. The distributed view, conversely, elevates the question of the subject to the level of platforms, institutions, and network structures, underscoring the systemic subject-effects that emerge when generative AI is embedded within distributed cognitive networks. While each of these three perspectives captures a distinct dimension of epistemic activity in the era of generative AI, no single standpoint suffices to fully account for the actual reconstruction of the subject occurring in reality. Accordingly, this paper advocates for an approach to the epistemic subject in the generative AI era characterized by “complementary checks and balances”: it rejects both the mystification of AI as a novel subject endowed with independent will, and its simplistic reduction to a completely neutral tool; instead—while firmly upholding the primacy of human agency and responsibility—it acknowledges the profound influence that algorithmic systems, platform architectures, and institutional frameworks exert upon cognitive activities. Such a synthetic perspective helps transcend the abstract, binary debate over whether AI constitutes a “subject,” thereby enabling a more precise grasp of the subject’s positioning, the boundaries of responsibility, and the normative implications inherent in the evolving modes of contemporary knowledge production.

Keywords

Generative AI, Epistemic Subject, Intentionality, Interactivity, Distribution, Complementary Checks and Balances

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. 问题的提出

主体问题历来是哲学中最具基础性和争议性的议题之一。就观念史脉络而言, 主体并不是一个自始就等同于自我意识的概念。海德格尔在考察主体概念的词源时指出, 希腊语 *hypokeimenon* 的字面意思是躺在下面的东西, 意即放在下面和放在基底的东西[1]。由此可见, 主体最初并非近代哲学意义上的理性自我, 而更接近某种承载属性、支撑存在的基底。近代以来, 笛卡尔以“我思故我”在为现代主体哲学奠定

起点, 主体遂由世界中的承载者转变为认识的出发点; 康德则进一步指出: “我思必须能够伴随着我的一切表象; 因为否则的话, 某种完全不可能被思考的东西就会在我里面被表象出来。” [2]至此, 主体不再只是经验事实中的心理存在, 而成为经验之所以可能的条件。认识主体问题, 也由此被推到哲学讨论的中心。

不过, 主体概念自近代确立以来, 并未沿着单一方向持续推进。十九世纪以后, 历史主义、实践哲学与社会理论不断冲击孤立主体的理论图景, 主体逐渐不再被理解为一个脱离现实关系、先验自足的理性中心, 而是被置于社会实践、制度结构与历史进程之中重新把握。马克思曾明确指出: “全部社会生活在本质上是实践的”; [3]同时又强调: “人们自己创造自己的历史, 但是他们并不是随心所欲地创造。” [4]前一句揭示了主体必须在实践中理解, 后一句则表明主体的能动性并非毫无限制, 而总是在既有条件中展开。主体由此不再只是形而上学意义上的自我, 而成为一种在历史与现实中生成的行动位置。

二十世纪的主体理论进一步强化了这一转向。现象学并没有简单消解主体, 而是从意向性结构和生活世界经验出发, 重新阐释主体与世界的关系; 社会理论和建构论则把主体放入制度、结构和交往网络之中, 使主体问题越来越多地被表述为如何理解主体, 而不是主体究竟是什么。正是在这一理论背景下, 生成式人工智能的广泛兴起, 使主体问题重新获得了现实迫切性。以大语言模型为代表的生成式人工智能已经深度介入科研写作、课堂教学、知识问答、政策咨询与舆情分析等活动, 不仅能够整合信息、生成文本, 而且能够参与问题重述、结构搭建和论证推进。它在事实层面上已不再只是外在工具, 而开始进入传统上主要由人承担的认知活动内部。

也正因此, 生成式人工智能对主体问题的冲击不应仅被理解为一项单纯的技术挑战, 而应被视为知识生产方式变化所引发的哲学与规范问题。现实中的争论已经出现明显分化: 有的研究聚焦于生成式人工智能是否具备真正的理解能力, 是否拥有某种意义上的意向性; 有的研究则强调在具体实践中, 人和机器如何形成新的协同关系以及谁应承担的责任; 还有的研究把视线推进到平台、制度与资本结构层面, 认为生成式人工智能引发的不是局部工具更新, 而是知识生产网络的整体重组。若缺乏相对稳定的分析坐标, 这些讨论便容易停留在各说各话的状态, 难以形成真正有效的比较与对话。正因为如此, 将当下关于生成式人工智能的主体争论加以类型化梳理, 不仅必要, 而且构成进一步讨论的前提。

2. 从经典主体哲学到当代主体重估

生成式人工智能时代的主体问题, 并不是脱离思想史脉络而凭空出现的。今天围绕算法、协同与平台所展开的争论, 实际上都可在近代以来主体哲学的演进中找到理论背景。若不回到这一背景, 便难以把握当代讨论为何会以意向性交互性分布性等不同方向展开。

二十世纪主体理论的多向展开, 为理解今天的生成式人工智能问题提供了更丰富的资源。一方面, 现象学强调主体的意向性结构, 指出主体并非孤立地拥有知识, 而是在关于某物的关系中展开世界经验。另一方面, 社会建构论与结构化理论则强调主体并不是先在于社会结构之外的实体, 而是在行动、制度与规范的循环中形成。吉登斯指出, 社会系统的结构性特征既是后者的中介, 又是它的结果[5]; 伯格和卢克曼则概括说: “社会是人类的产物……人是社会的产物” [6]。前者表明行动与结构之间存在双向生成关系, 后者则揭示了主体与社会世界之间的辩证循环。由此, 主体不再被看作纯粹内部性的意识中心, 而被理解为在结构和实践中被组织出来的位置。

当代主体理论并没有简单宣布主体死亡, 而是使主体问题从本体论和认识论意义上的固定主体转向历史的、生成的、交互的和结构化的主体。这一转变, 正是生成式人工智能主体讨论得以展开的理论前提。今天的关键问题已不再只是主体是否存在, 而是主体如何在技术 - 社会条件中被重新组织起来。正因为理论重心已经发生这样的转移, 当代关于生成式人工智能的争论才会自然分化为几种不同的认识主

体观。

3. 生成式人工智能对认识主体的三重重塑

生成式人工智能对主体问题的冲击,根本上不是一个简单的机器是否比以前更聪明的问题,而是认识活动本身的组织逻辑正在发生变化。若只笼统地把这类系统理解为“会生成文本的机器”,便难以说明它为何会在事实上逼近传统上由人承担的部分认知位置。以大语言模型为代表的生成式人工智能,其运行至少包含几个值得注意的技术环节:其一,预训练使模型能够在大规模语料中学习词语、句式与知识片段之间的统计关联;其二,RLHF(基于人类反馈的强化学习)通过人工排序、偏好反馈与奖励建模,将“什么样的回答更像合适回答”嵌入模型输出之中;其三,提示工程则通过角色设定、任务分解、少样本示例、上下文约束等方式,显著改变模型的生成路径与回答质量。正是这些技术机制的叠加,使生成式人工智能在表面上不仅“会说”,而且越来越像“会配合”“会解释”乃至“会思考”。

这一点对于主体判断具有直接影响。RLHF所塑造的,并不是严格意义上的原生意向性,而是一种经由人类偏好训练出来的“对齐性表达”。模型之所以常常显得连贯、礼貌、合作,甚至似乎能够把握提问者意图,并不必然意味着它已经获得了独立的理解能力,而更可能意味着它被优化为更擅长呈现符合人类期待的回答样态。提示工程则进一步表明,生成结果并不是模型单方面“想出来”的,而是在提问者、任务设计、上下文材料与模型参数共同作用下被生产出来的。换言之,生成式人工智能越是表现出近似主体的语言能力,就越提示我们:主体问题已不能仅凭表面输出加以判断,而必须同时考察其背后的技术机制、互动条件与结构环境。

若从更细的层面看,这种变化至少体现在三个彼此勾连的方面:主体构成的演进、主体形态的重组与主体功能的重构。

在主体构成上,认识主体出现了由单一行动者向复合行动者演进的趋势。传统意义上的认识主体,大多默认是某个具体的人或人的共同体。即便在社会理论中,主体也主要被理解为人类行动者在历史与制度中的表现形态。然而,生成式人工智能的广泛应用已经使这一前提发生动摇。今天的知识生产过程往往包含多个彼此关联的行动节点:研究者提出问题并设计提示,模型依据预训练与对齐机制给出若干可能路径,平台系统再通过界面、排序与调用规则塑造信息流向和表达形式。主体不再由某一个点单独承担,而是呈现出由人、模型、平台和制度共同参与的复合结构。所谓主体构成的变化,并不是说人已经被排除出知识生产,而是说原本主要由人内部完成的认知流程,现在越来越需要在多重节点之间被组织起来。

在主体形态上,生成式人工智能推动主体由中心化图景走向网络化和分布化图景。原先的主体理论往往设定一个稳定的认知中心:无论是笛卡尔的“我思”,还是康德的先验统觉,抑或后来许多政治哲学和道德哲学中的责任主体,主体大多被视为统一活动的源点。然而,随着知识活动越来越依赖平台、接口、模型和数据库,这种单中心想象不断被削弱。尤其是提示工程所揭示的“上下文决定生成路径”,使主体越来越表现为一个处在多重关系中的动态位置。人的判断能力并未消失,但它已经不再独占知识形成的全部环节;模型并非完整主体,却在事实层面上承担部分认知功能;平台和制度则通过界面设计、评价逻辑和数据分配重塑认知活动的外部条件。主体形态因此由中心化向网络化转变。

在主体功能上,生成式人工智能促使主体由直接的奠基者转向协同组织者和责任承担者。近代主体哲学强调主体的功能主要在于奠定知识的可能性,为判断和经验提供统一根据。生成式人工智能进入知识生产后,主体的功能并非消失,而是被重新分配。今天的主体越来越少以独立生成全部知识内容的方式出现,更多体现为对生成过程的审查、筛选、组织、修正和负责。人类主体不再只是内容的唯一生产者,更是问题的设定者、提示的设计者、规范的把关者和结果的承担者。模型系统则在信息整合、语义

生成和结构组织中发挥辅助乃至准核心作用。主体功能因此从单向生产转向协同共创, 从自我奠基转向对复杂流程的反思性把控。

也正是在这三重变化基础上, 当代关于生成式人工智能的主体争论才会形成不同类型。有的讨论沿着经典主体哲学延伸, 首先追问机器是否真正具备意向性和理解力; 有的讨论从现实实践出发, 强调在协同场景中主体如何被重新配置; 还有的讨论则把目光放到平台、制度和分布式认知网络层面, 认为主体已成为一种结构性效果。这些并非互不相关的观点, 而是同一现实在不同分析层面上的理论表达。

4. 意向性认识主体观：从是否理解界定主体

意向性认识主体观是最容易与经典主体哲学接续的一条路径。它所关心的是一个根本性问题：一个行动者之所以能够被称为认识主体, 关键不在于它能否表现出复杂行为, 而在于它是否真正具有关于对象世界的关于性, 即是否能够以某种方式理解、指向并把握其所面对的对象。由此, 生成式人工智能能否成为主体, 便首先被转化为它是否真正理解的问题。

这一立场的理论优势非常明显。面对生成式人工智能日益逼真的语言表现, 意向性路径提醒我们, 不能把文本流畅、逻辑连贯或表达拟人化, 直接等同于主体资格的成立。即便模型能够在形式上模仿论证、生成复杂叙述、回应抽象问题, 它仍然可能只是通过概率分布和符号关联运作, 而并不拥有与人类相同的世界经验、生活世界和意义奠基方式。换言之, 意向性主体观通过把主体资格系于理解和语义指向, 保持了主体概念的哲学严肃性。它可以有效抵御一种现实中十分普遍的倾向, 即把像主体误当作是主体, 把行为上的拟人化误判为存在论和认识论上的主体性成立。

然而, 意向性路径的限制也很突出。它常常将主体问题过于集中于内部状态的成立与否, 因而容易把一个实践性极强的问题, 压缩成一个关于系统内部是否拥有真正理解的概念争论。问题在于, 现实中的生成式人工智能即便不具备严格意义上的原生意向性, 也已经在教育、科研、媒体与治理场景中承担了事实性的认知功能。它未必真正懂得自己生成的内容, 但它确实在改变人类问题提出的方式、信息处理的节奏和论证展开的路径。若只围绕它究竟懂不懂来界定主体, 便无法充分说明为什么今天的知识生产会因它的加入而发生结构性变化。

更重要的是, 意向性路径往往容易引出二元化判断: 要么承认人工智能拥有主体地位, 要么将其重新归回工具。然而, 生成式人工智能恰恰处在一个介于两者之间的复杂地带。它不是传统意义上被动的工具, 因为它能够主动生成内容并影响认知流程; 但它也不是完整意义上的主体, 因为它缺乏独立责任担当与价值反思能力。意向性认识主体观在划清边界方面十分重要, 但若试图独自承担对现实技术实践的解释, 便难免显得过于抽象和单薄。

5. 交互性认识主体观：从协同实践理解主体

与意向性路径不同, 交互性认识主体观不首先追问机器内部是否存在真正理解, 而是把视野转向具体实践中的人机协同过程。在这一视角下, 主体并不是一个先于互动而完整存在的实体, 而是在互动中不断被组织、分配和生成的关系结构。生成式人工智能最值得重视的变化, 并不只是它是否像人思考, 而是它已经在越来越多场景中成为与人一起完成认知任务的行动参与者。

若进一步与计算机支持的协同工作(CSCW)和人机交互(HCI)领域对话, 这一路径的理论内涵可以得到更清楚的展开。HCI 研究长期强调, 技术系统不是被动执行命令的中性容器, 而是通过界面设计、交互流程与可供性结构塑造用户的行动方式。对于生成式人工智能而言, 提示框、追问机制、多轮对话、历史上下文调用与即时反馈, 并不是纯粹外在的技术细节, 而是直接影响认知活动如何展开的交互装置。提问者之所以能够把模糊的问题逐步推进为清晰的问题, 往往不是因为其一开始就拥有完整想法, 而是

在与模型的往复互动中不断修正表达、筛选路径和重组判断。由此可见,交互性主体观所说的“主体生成”,并不是抽象意义上的关系论表述,而是具体发生在界面、反馈与修正机制中的认识过程。

CSCW 的讨论则进一步提醒我们,协同工作的关键并不只是任务被多人共同完成,更在于工作如何被分解、协调、可视化和问责。将这一思路引入生成式人工智能场景,可以看到,模型之所以呈现出“共思考者”的特征,并不是因为它已经成为与人等等的主体,而是因为它在事实上参与了任务分配、信息中介、草稿生成和方案比较等环节。在论文写作中,研究者可能负责问题意识、论证方向与结论判断,模型则负责备选框架、材料整理与表达试探;在教育场景中,教师负责目标设定、规范边界与价值判断,模型则承担即时反馈、差异化解释与练习生成;在治理和行政场景中,管理者负责标准设定、风险把控与最终决策,模型则参与初步分类、摘要和文本起草。主体在这里并不是单数,而是一种在协作过程中被不断重新配置的任务结构。

从这一角度看,提示工程对主体问题的意义尤其突出。提示工程并不只是提高回答质量的技巧,它事实上揭示了人机协同中的认识分工:问题由谁设定,角色由谁指定,标准由谁嵌入,输出由谁筛选。模型的回答之所以常常显得“聪明”,很大程度上正是因为人类提问者已经通过提示把任务边界、表达风格与评价准则预先写入了互动过程。也就是说,生成式人工智能的交互性并不是机器主体性的直接证明,反而提示我们:所谓“智能回答”常常是由人类任务设计能力与模型生成能力共同完成的结果。

交互性认识主体观的突出优点,在于它能够贴近现实应用场景,看到主体性在实践中如何被具体配置。它把讨论重心从“机器是否本体上是主体”转移到“在何种协同机制中,谁承担何种认知职能”这一更具解释力的问题上。尤其在人机协同场景中,主体之所以重要,不只是因为它能认识,更因为它要承担认识带来的后果。CSCW 中关于可见性、协调与责任链条的讨论,在这里就转化为一个更直接的问题:谁对生成内容负责,谁拥有审查与否决权,谁承担事实错误、偏见扩散和规范失当的后果。正是在这一意义上,交互性认识主体观能够把主体问题从抽象的存在论争论,引回到实践中的责任分配与制度安排上。

不过,交互性路径也并非没有缺陷。无论是 HCI 对界面的重视,还是 CSCW 对协同工作的分析,主要都着眼于中观层面的互动场景。它们有助于解释人机关系如何在具体任务中被组织,却未必足以说明这些关系为什么会以当前方式被组织起来。换言之,人机互动并不是在真空中发生的。界面设计由平台决定,信息可得性受数据结构限制,输出风格受到 RLHF 等对齐机制影响,使用场景又受到组织制度与评价体系的塑造。如果忽略这些更大的结构条件,交互性分析就容易把看似自然的人机协同理解得过于中性,从而低估平台与制度对主体配置的预先塑形作用。

6. 分布性认识主体观:从网络与制度层面把握主体

分布性认识主体观是在前两种路径基础上的进一步推进。它认为,生成式人工智能带来的主体问题,不仅表现为某个系统是否有理解能力,也不只是局部互动中谁在承担任务,而是整个知识生产网络的组织方式发生了变化。主体在这里不再被理解为某个单独点上的属性,而更接近一种分布于多重节点之间的系统性效果。

若与行动者网络理论(ANT)展开对话,这一路径的理论内涵可以得到明显深化。ANT 的重要启发在于,它不预设只有人类才构成有效行动者,而是主张从异质要素的连接、转译和委托中理解行动如何发生。若从这一视角考察生成式人工智能,那么提示词、训练数据、模型参数、接口平台、内容审核规则、云端算力、用户习惯与组织制度,都不是外在背景,而是共同参与知识生产的网络节点。模型之所以在现实中表现出近似主体的效应,并不是因为它单独拥有完整主观性,而是因为它在网络中承接并转译了原本分散于不同节点的功能:数据集为它提供可学习的表达资源,RLHF 为它写入人类偏好的排序机制,

平台界面为它组织交互入口, 用户提示为它限定任务方向, 组织制度则为其输出规定使用边界。所谓“模型在思考”, 在很多情境下毋宁说是一个异质网络经由模型节点所呈现出来的集中效果。

但分布性主体观又不能简单等同于 ANT 的完全对称立场。ANT 的长处在于揭示人类与非人要素如何共同构成行动网络, 但若把这种方法直接移用于主体规范问题, 便容易滑向责任扁平化。因为主体问题并不只是说明“谁产生了效果”, 还要回答“谁应当负责”。在这一点上, 分布性认识主体观更适合作为一种有选择的借鉴: 它吸收 ANT 关于异质网络、功能委托和行动转译的洞见, 用来说明生成式人工智能的主体效应何以形成; 同时又必须保留人的规范优先性, 以防止主体责任在网络叙事中被完全稀释。

若进一步与媒介理论对话, 则可以更清楚地看到, 生成式人工智能不仅是网络中的一个行动节点, 而且正在成为一种新的认知媒介。媒介理论的基本启发在于, 媒介从来不是中性的通道, 它会主动塑造信息的可见性、表达的形式和理解的节奏。对于生成式人工智能而言, 平台化的对话界面、模板化的回答结构、即时生成的时间机制以及与搜索、办公、教学、传播系统的深度整合, 使其不只是帮助人“表达已有思想”, 而是在很大程度上预先组织了“什么更容易被表达”“什么更容易被相信”以及“什么更容易被继续扩散”。也就是说, 生成式人工智能并非单纯输出内容, 它还在重塑认知活动的媒介条件本身。

由此, 分布性认识主体观的解释力首先体现在, 它能够把认识问题与权力问题、媒介问题联系起来。知识从来不是在真空中生产的, 它总是受到平台结构、资本逻辑、媒介形式和制度规则的影响。生成式人工智能进入知识生产后, 并不只是帮助生成文本, 还会影响哪些问题更容易被提出, 哪些材料更容易被看见, 哪些表达更容易被放大, 哪些立场可能被系统性边缘化。主体问题在这里已经不是单纯的谁有能力认识, 而是谁在控制认知条件, 谁在塑造知识流向, 谁在组织可见性与可说性。正是在这个意义上, 分布性主体观获得了比单纯“人机协同”更强的结构批判维度。

然而, 分布性认识主体观也有自身问题。它在揭示宏观结构、平台权力与媒介塑形方面极具洞察力, 却容易在规范层面导致主体责任的抽象化。一旦主体被理解为网络的整体效果, 具体责任该如何落实就会变得困难。若一切都被归于系统, 那么谁对错误输出负责, 谁对偏见传播负责, 谁对事实失真承担义务, 反而可能变得更加模糊。因此, 分布性路径虽能把握结构条件, 却不能完全取代对具体行动者和具体场景的分析。更准确的做法, 应该是在 ANT 与媒介理论所揭示的网络-媒介条件之上, 重新追问人的审查义务、平台的治理责任与模型部署者的制度责任。

7. 从非此即彼到互补制衡

比较以上三种认识主体观可以发现, 它们之间并不是简单并列的替代关系, 而是构成一个彼此竞争、又彼此补充的理论场域。意向性认识主体观强调主体资格的概念边界, 交互性认识主体观强调主体在协同实践中的生成, 分布性认识主体观则强调主体在平台和制度网络中的结构化组织。三者所面对的问题层次并不相同, 因此不能用单一标准来裁决其全部优劣。

也正是在这一点上, 互补制衡思路具有重要意义。所谓互补, 是指三种主体观分别揭示了生成式人工智能时代主体问题的不同面向。意向性路径提醒我们, 不应因为算法具备复杂表达能力, 就轻率取消人与机器在主体资格上的差异; 交互性路径提醒我们, 主体不是一成不变的实体, 而是在具体实践中通过分工、协作和反馈被不断重组; 分布性路径则提醒我们, 任何局部互动都不是孤立发生的, 而是深深嵌入平台、制度和资本逻辑之中。只有把这三个层面放在一起, 主体问题才不会被过度简化。

所谓制衡, 则意味着三种路径可以相互纠偏。意向性路径能够防止交互性和分布性分析滑向泛主体化, 避免因为机器参与了认知过程便直接赋予其完整主体地位; 交互性路径能够防止意向性分析停留在抽象概念辨析中, 而忽视现实场景中的责任配置与协同机制; 分布性路径则能够防止前两者只盯住个体

和局部互动, 而忽略平台权力、制度安排和知识可见性的整体条件。互补保证了分析的完整性, 制衡则避免任何单一立场走向偏狭。

从这一意义上说, 生成式人工智能时代真正值得追问的问题, 已经不再是简单的人工智能是不是主体。这一问题过于抽象, 而且极易陷入赞成或反对的二元对立。更重要的问题应当是: 在怎样的知识-权力-资本结构中, 哪些行动者、装置和制度被组织成某种类型的主体? 在这种组织方式中, 人类作为认识和实践主体的判断权、解释权与责任权究竟发生了怎样的变化? 这样的提问方式, 既避免了技术神秘化, 也避免了简单工具论, 而是把主体问题重新放回知识生产的现实机制之中。

事实上, 若从马克思“全部社会生活在本质上是实践的”这一命题出发, 再结合吉登斯关于结构既是后者的中介, 又是它的结果的判断, 就可以看到: 生成式人工智能所带来的主体变化, 并不是主体被机器简单替代, 而是主体活动的条件、方式和边界被重新组织。今天的主体问题必须同时在实践层面和结构层面被理解。主体既不是完全内在的意识事实, 也不是完全外在的制度效果, 而是在行动、技术和结构的循环中不断被生成、被限制、被赋权和被重写的关系性位置。

由此, 互补制衡主体观所坚持的, 并不是对人的主体地位的放弃, 而恰恰是在新的技术条件下重新确认人的主体地位。人仍然是价值判断、规范设定和责任承担的最终主体, 但这并不意味着知识生产仍保持着旧有的纯人类中心结构。相反, 人的主体性今天必须通过对算法系统、平台结构和制度安排的反思性控制来重新实现。坚持人的主体地位, 不是回到技术尚未深度介入的旧主体状态, 而是在承认技术已深度改变认知活动的现实前提下, 重新确立人的审查能力、解释能力和责任能力。

8. 结语

总体来看, 生成式人工智能的兴起并没有让主体问题失去意义, 反而使这一问题以更复杂的方式重新浮现出来。与传统技术相比, 生成式人工智能不仅提高了信息处理和文本生成的效率, 更深刻改变了知识生产的组织方式。主体不再只是单一的人类个体, 而是在构成上趋于复合, 在形态上趋于网络化, 在功能上趋于协同化与共创化。也正是在这一变化背景下, 当代关于认识主体的讨论才会分化为意向性、交互性和分布性三种具有代表性的主体观。

意向性认识主体观强调主体资格的概念边界, 提醒我们不要把复杂工具轻率误判为主体; 生成式人工智能的“主体样态”并不是孤立内部状态的自然流露, 而是在界面设计、提示过程、任务分工和责任链条中被具体组织出来的; 分布性认识主体观则通过与 ANT 和媒介理论比较, 揭示出知识生产背后的网络结构、平台权力与媒介塑形条件。与此同时, RLHF、提示工程等关键技术又进一步说明, 生成式人工智能之所以会给人以“仿佛具有主体性”的印象, 并不只是因为它能生成内容, 更因为它在技术上被训练为更符合人类偏好, 并在互动中被不断嵌入特定任务与规范之中。三种路径各有解释力, 也各有局限。任何单一视角都不足以独立把握生成式人工智能时代主体问题的全部复杂性。

因此, 更合理的理论选择不是在三者之间作简单裁决, 而是建立一种互补制衡的综合框架。这样的框架既能够保持主体概念的哲学严肃性, 又能回应现实实践中的责任配置问题, 还能对平台化和制度化条件下的结构性权力保持敏感。它把主体问题从人工智能是否具有主体身份的抽象争论, 转化为当代知识生产如何组织主体、重配权力与界定义务的现实追问。也只有在这种意义上, 生成式人工智能时代的主体研究, 才能真正兼顾理论深度与现实关怀。

参考文献

- [1] 海德格尔·尼采[M]. 孙周兴, 译. 北京: 商务印书馆, 2004: 773.
- [2] 康德. 纯粹理性批判[M]. 邓晓芒, 译. 北京: 人民出版社, 2004: 89.

- [3] 马克思. 关于费尔巴哈的提纲[C]//马克思恩格斯选集: 第1卷. 北京: 人民出版社, 1995: 56.
- [4] 马克思. 路易·波拿巴的雾月十八日[C]//马克思恩格斯选集: 第1卷. 北京: 人民出版社, 1995: 585.
- [5] 吉登斯. 社会的构成: 结构化理论纲要[M]. 李康, 李猛, 译. 北京: 中国人民大学出版社, 2016: 89.
- [6] 伯格, 卢克曼. 现实的社会建构: 知识社会学论纲[M]. 汪涌, 译. 北京: 北京大学出版社, 2009: 61.