认知视域下 DeepSeek 技术构境的三重逻辑

马俊峰　温兆伦
(西北师范大学 马克思主义学院,甘肃 兰州 730070)

[摘　要] DeepSeek 已经成为数字时代具代表性的大型语言模型,这要求学界不断深化对 DeepSeek 的研究。从认知视域展开 DeepSeek 的多层构境,首先要重视 DeepSeek 作为“第三持存”对人类认知能力的建构作用,关注其与人类认知劳动的交互关系。通过辩证地把握 DeepSeek 在突破“算法黑箱”、塑造分布式认知网络的基础上超越资本逻辑进而解构“信息茧房”的技术潜能,与经由其架构机制、训练方式以及推理特性再构“AI 幻觉”的双重属性,探寻依据基础软硬件可信、数据可信、模型可信与应用可信的“可信可控”原则降低幻觉风险。从而以动态框架揭示大型语言模型的认知演进逻辑,明确人机协同中人类的价值判断与主体地位,为人工智能时代的认知治理确立价值坐标。

[关键词] DeepSeek;第三持存;信息茧房;AI 幻觉
[中图分类号] G206　[文献标识码] A　[文章编号] 1008-7699(2026)01-0001-10

2025 年 1 月杭州深度求索人工智能基础技术研究有限公司发布 DeepSeek-R1 推理模型,随之引起全球科技圈震动。① DeepSeek 在群体相对策略优化算法、知识蒸馏与长链思维推理等技术的使用中展现出突破以往生成式人工智能性能边界的技术实力。② 在此前提下,当前学界已从自动化技术③、新闻传播④、信息经济⑤、行政法制⑥等多维度对其展开系统性探讨。然而,DeepSeek 的问世同样“标志着一次深具理论与实践意义的认知架构跃迁”⑦,而这要求我们回归哲学话语体系来解析现代技术所蕴含的抽象性特征。⑧ 有鉴于此,依循“建构—解构—再构”的逻辑进路对 DeepSeek 在认知维度的构境机制展开分析,实质上是以动态分析框架揭示人工智能对人类认知的深刻影响,并为人工智能时代的认知治理确立价值坐标。

[收稿日期] 2025-07-31
[基金项目] 国家社会科学基金项目(21VSZ127)
[作者简介] 马俊峰(1969—　),男,甘肃张家川人,西北师范大学马克思主义学院教授、博士生导师;温兆伦(1999—　),男,内蒙古包头市人,西北师范大学马克思主义学院博士研究生。

① 张慧敏:《DeepSeek-R1 是怎样炼成的?》,《深圳大学学报(理工版)》2025 年第 42 卷第 2 期,第 226-232 页。

② 邓建鹏、赵治松:《DeepSeek 的破局与变局:论生成式人工智能的监管方向》,《新疆师范大学学报(哲学社会科学版)》2025 年第 4 期,第 99-108 页。

③ 蔡天琪、蔡恒进:《DeepSeek 的技术创新与生成式 AI 的能力上限》,《新疆师范大学学报(哲学社会科学版)》2025 年第 4 期,第 136-143 页。

④ 刘建明:《DeepSeek 震惊世界新闻业的精准效能》,《新闻爱好者》2025 年第 4 期,第 4-10 页。

⑤ 陆岷峰、高伦:《DeepSeek 赋能商业银行创新转型:技术应用场景分析与未来发展路线》,《农村金融研究》2025 年第 2 期,第 19-34 页。

⑥ 邓建鹏、赵治松:《DeepSeek 的破局与变局:论生成式人工智能的监管方向》,《新疆师范大学学报(哲学社会科学版)》2025 年第 4 期,第 99-108 页。

⑦ 令小雄:《DeepSeek 开启后 ChatGPT 时代——基于数字范式革新及其运演哲思》,《西北工业大学学报(社会科学版)》2025 年第 2 期,第 59-67 页。

⑧ 韩庆祥:《以哲学把握经济的基本方式》,《哲学研究》2020 年第 11 期,第 18-27 页。

一、认知建构层:DeepSeek 作为“第三持存”的技术本质

DeepSeek 的出现意味着“人—技术”关系进入崭新阶段,其在重构技术装置与人类认知能力的基础上,进一步加速了技术装置对主体生命的影响。依斯蒂格勒所言,“人—技术”关系的演进逻辑源自于人类通过技术补全主体“原始性缺陷”的需求。“人类没有过失,有的只是起源的原初缺陷,它使缺陷的共同体成为共同体的缺陷”。① 即人类的生物进化过程具有先天匮乏,人类既缺乏锋利的爪牙、厚实的皮毛等生物本能装备,又面临大脑记忆容量有限、神经传递速度缓慢等认知局限。由此人类必须借助外化技术装置实现自我补完。以石质工具的发明强化肢体力量,语言符号系统的形成突破生物记忆的时空限制,数字媒介的迭代实现意识活动的体外延展,当生物基因无法通过自然选择及时应对环境剧变时,体外技术装置的持续进化便成为确保物种存续的关键策略。正如斯蒂格勒所言,“如果一个外在的东西构成它所面对的存在本身,那么这个存在就是存在于自身之外。人类的存在就是在自身之外的存在”。② 这意味着,在人类的进化谱系中,技术始终扮演着人类“后种系生成”的补偿机制角色。随着人类演化过程中认知能力的发展与语言符号系统的形成,“后种系生成”的技术性补偿机制逐渐由生理层面扩展至精神层面,在记忆机制中表现为超越个体神经系统的外部化记忆形态,从而在技术的支持下形成具有持续性与积累性的记忆系统,也就是依靠“技术和语言的记忆”③机制。斯蒂格勒认为,如果说“第一持存”是神经系统的即时感知留存,“第二持存”是大脑皮层的内化记忆,那么“第三持存”则通过符号刻写、媒介储存等技术实践将认知活动对象化为可脱离生物体存在的技术装置。这样一来,“第三持存”作为人类记忆的技术补偿构成人类认知活动的关键环节。④ 由此,“人类不断通过技术装置实现记忆的外在化持存,实现历史踪迹的保留与传承”。⑤

DeepSeek 作为数字化的“第三持存”,所使用的知识蒸馏技术是通过师生模型的知识迁移实现认知经验的体外化留存,也就是将历时性积累的认知成果压缩为可复制的数据参数,由此能够将人类的经验知识集中展现⑥,这意味着技术装置具有了超越个体经验的演化路径。可以发现,知识蒸馏技术在实现认知成果传递的基础上能够通过持续的数据分析突破个体认知界限,从而使得技术装置具有自主优化的能力。在此基础上,技术装置的演化特征可以从西蒙的“机器思维”思想进一步考察,西蒙在对人类思维与机器思维的比较中认为,人类“复杂的行为智能是通过连续多个步骤而非一大步跳跃还原成神经过程”。⑦ 也就是说,人类的思维并非源于整体性、不可约化的意识本体,而是由一系列可被解析的符号操作序列构成,符号操作又可形式化为算法指令并在有限的时间与资源条件下得以进行精确地模拟与计算,从而使人类的行为具备可分析性与可复制性。这意味着,我们可以将“人类所有的思维过程都看成是物理符号的事件,让机器模拟人类智能有了可行

① [法]贝尔纳·斯蒂格勒:《技术与时间(修订合卷本)》,裴程等译,译林出版社 2023 年版,第 232 页。

② [法]贝尔纳·斯蒂格勒:《技术与时间(修订合卷本)》,裴程等译,译林出版社 2023 年版,第 221 页。

③ [法]贝尔纳·斯蒂格勒:《技术与时间(修订合卷本)》,裴程等译,译林出版社 2023 年版,第 203 页。

④ 张一兵:《斯蒂格勒〈技术与时间〉构境论解读》,上海人民出版社 2018 年版,第 95 页。

⑤ 马俊峰,赵海蕴:《“第三持存”在生活世界的运演逻辑及价值旨趣》,《南开学报(哲学社会科学版)》2025 年第 1 期,第 1-10 页。

⑥ Geoffrey Hinton, Oriol Vinyals & Jeff Dean, *Distilling the Knowledge in a Neural Network*, at https://arxiv.org/pdf/1503.02531(Last visited on February 27, 2025).

⑦ [美]赫伯特·西蒙:《科学迷宫里的顽童与大师:赫伯特·西蒙自传》,陈丽芳译,中译出版社 2018 年,第 248 页。

性”。[①] 这样一来，技术装置就能够将基于物理符号系统的认知过程映射为可执行的计算模型，并通过模块化算法架构与实时反馈机制实现对环境变化的动态感知与自主优化。因此模拟人类思维的 DeepSeek 既包含目标导向的设计内核，又具备环境适应性调节机制。DeepSeek 依据知识蒸馏构建的符号化知识体系与数据驱动型的反馈机制形成互动关系，使技术系统的认知方式从人类的认知映射转向人机协同的认知活动。这意味着 DeepSeek 能够与人类之间形成认知互动的关系，人类将感知、推理与创造等认知能力外化为数据与算法，而算法系统的运算结果又增补并塑造着人类的认知，由此形成动态的认知循环回路。换言之，人类主体与技术装置在持续的交互中互为因果，DeepSeek 作为人类认知能力的体外延伸载体也在塑造人类的思维方式与认知内容，从而推动着人类的认知模式向“人机协同”转型。

从笛卡尔的“我思故我在”，可以窥见主体的认知能力如何被数字技术所建构。笛卡尔认为，“要想追求真理，我们必须在一生中尽可能地把所有的事物都来怀疑一次，凡可怀疑的，我们也都应当认为是虚妄的。思想本身就是思想的活动，当思想在怀疑时，思想可以怀疑外在的对象，也可以怀疑思想之内的对象，思想可以怀疑思想的一切对象和内容，而‘我’就是怀疑活动的主体”。[②] 从普遍怀疑的方法论出发，笛卡尔认为应该将“我思故我在”作为第一原理，“所谓第一原理就是一个绝对不怀疑的东西，找到这个绝对不可怀疑的东西之后，才能以这个东西为基础进行推理”。[③] 在此意义上，身体与技术便是需要悬置的“可疑之物”。然而大型语言模型通过对身体经验与心智活动进行形式化的分解与符号化处理，将其转化为可编程、可操作的算法模型，使技术本身在笛卡尔式的普遍怀疑中不再只是被悬置的外在对象，而成为能动的认知参与者。质言之，DeepSeek 是能够将人类的认知成果转化为可操作可复制的数据参数，这样一来技术便演化为具有认知能力的“准主体”，这与笛卡尔将技术限定为广延实体的物质性存在形成对立。在此基础上，可以进一步借助笛卡尔的理论体系分析 DeepSeek 在认识论层面引发的变革，笛卡尔将知识的确定性建立在“领会得清楚明白的观念都是真实的”[④]基础之上，我们之所以可以“领会”是因为普遍理性的存在，并且理性人皆有之“良知，是人间分配得最均匀的东西……也就是我们称为良知或理性的那种东西，本来就是人人均等的”。[⑤] 而大型语言模型通过数据驱动所形成的认知形式，主要是以概率分布而非逻辑演绎来建构知识体系。这种认知方式的转变使得笛卡尔强调的“基础主义”认识论[⑥]让位于基于技术装置的“融贯论”真理观[⑦]，也就是说，对于 DeepSeek 而言，知识的有效性并不取决于与先验理性的契合度，而是维系于算法模型内部参数的系统自洽性。

就此而言，现象学视角的引入能够进一步凸显这一认知形式的转变，胡塞尔以现象学对笛卡尔理论进行了改造，他认为，唯有意识自身才能够以绝对自给、自证与明示的方式显现，因此应将意识确立为哲学研究的根本起点，以意识作为起点便使得“一切非意识的存在物都变成了意识的意向相

① 李宁宁、宋荣：《人-机-思维模型：对赫伯特·西蒙机器发现思想的审思》，《科学技术哲学研究》2022 年第 5 期，第 87-93 页。

② [法]笛卡尔：《哲学原理》，关文运译，商务印书馆 1958 年版，第 1 页。

③ 谢宜麟：《笛卡尔“我思故我在”的含义及意义》，《吉首大学学报(社会科学版)》2017 年第 S2 期，第 94-96 页。

④ 康萍：《论笛卡尔的知识“确定性”问题》，《河南理工大学学报(社会科学版)》2020 年第 5 期，第 14-19 页。

⑤ [法]笛卡尔：《谈谈方法》，王太庆译，商务印书馆 2000 年版，第 3 页。

⑥ 杜以芬：《浅谈近代基础主义认识论的特点》，《济南大学学报(社会科学版)》2008 年第 4 期，第 62-66 页。

⑦ 曾志：《西方知识论哲学中的真理融贯论》，《社会科学辑刊》2005 年第 1 期，第 4-9 页。

关物”①。在 DeepSeek 中,意识与存在物的关系体现为算法与数据的交互关系。换言之,知识蒸馏能够将人类经验转化为数字空间中的数据参数,在此意义上现象学所描述的“意识流”便被技术化为可量化的信息流,也就是超越生物神经可塑性的“技术意向性”②。具体而言,DeepSeek 是凭借算法系统建立起数据特征与任务目标之间的耦合关系,模拟了知觉与行动的具身循环,使得原本依赖生物神经网络时空延续性的意向弧不再局限于人类意识的投射,也就是通过将动机生成、行为执行与反馈评估等意向性环节转化为自主运行的算法模块,并以迭代优化为驱动,在数字化的闭环回路中实现了对意向弧的重构与延展。在此意义上,一方面,DeepSeek 承继并延展了人类意识的意向弧,使得感知与行动的循环可以在数字空间中自主运作。另一方面,DeepSeek 又能主动生成新的关联结构,对环境进行选择性呈现与意义建构。由此可见,DeepSeek 作为人类智能的技术性延伸,在“技术意向性”意义上承载着人类经验的原有指向,同时也以自身的运算逻辑开拓出超越个体意识的新型认知活动场域。

如果把 DeepSeek 对主体认知能力的建构置于生产劳动的语境下,可以看到 DeepSeek 对人类认知劳动的替代。大卫·哈维认为,自二十世纪七十年代起,知识经济与技术革新协同发展,并持续推动着现代劳动形式的转型升级。“价值生产不再仅仅是建立在物质生产基础上,而是日益建立在难以衡量和量化的非物质要素之上,如信息、情感、智力”。③ 在此过程中,“技术正取得一些我们认为属于活着的生物的特性。随着技术变得可以感知周遭环境并作出反应,随着技术变得能够自我组合、自我设定、自我疗愈和有‘认知’能力,它们越来越像活着的生物。技术越是精密和‘高科技’,便越像生物”。④ 因此可以这样认为,技术生物化的趋势实质上是人类认知能力的外化与客体化,当人工智能装置通过深度学习获得类似生物体的自主认知能力时,原本专属于人类的认知劳动便转化为可被机器系统接管的技术过程。正如马克思在《资本论》中所言,“劳动资料取得机器这种物质存在方式,要求以自然力来代替人力,以自觉应用自然科学来代替从经验中得出的成规”。⑤ 这就是说,人类的认知劳动随着机器的使用抛弃了工业时代机械性重复劳动与服务业时代程式化知识劳动,进入了数字时代以人工智能为主体的认知劳动。而当我们对认知劳动进行考察时,就意味着我们要对人的认知能力作出一些限定性说明。就此而言,西蒙在认知科学与计算理论的基础上指出人类认知活动的多层特征。在他看来,人类认知并非生物刺激与反应机制的简单延伸,而是在实践活动中形成的复杂信息处理系统。西蒙认为,人类认知结构的层次化特征具体展现为基础层的感知处理、中间层的语义与知识整合、元认知层的逻辑建构与策略规划功能,而这三层认知能力构成了从物理感知到知识建构的完整过程。⑥

依据西蒙的理论考察 DeepSeek 对于人的认知能力的影响可以发现,在基础感知层面,DeepSeek 作为文本型大语言模型,通过文本数据的分布式表征技术实现了对不同类型文本信息的统一处理,其表征网络能够将结构化文本与非结构化文本编码为可计算的向量数据,在此意义上突破了传统单通道文本处理中海量信息筛选效率低、多源文本关联难的局限。例如,由于 DeepSeek

① 吴增定:《“我思”及其主体性——简析胡塞尔在〈第一哲学〉中对于笛卡尔的解释》,《哲学动态》2017 年第 3 期,第 69-76 页。

② 韩连庆:《技术意向性的含义与功能》,《哲学研究》2012 年第 10 期,第 97-103 页。

③ 施灿业:《认知资本主义的劳动价值理论及其批判》,《江淮论坛》2023 年第 1 期,第 66-72 页。

④ [英]大卫·哈维:《资本社会的 17 个矛盾》,许瑞宋译,中信出版社 2016 年版,第 102 页。

⑤ 《马克思恩格斯全集》(第 23 卷),人民出版社 1972 年版,第 423 页。

⑥ [美]赫伯特·西蒙:《认知:人行为背后的思维与智能》,荆其诚等译,中国人民大学出版社 2020 年版,第 21 页。

能够快速整合医学知识与海量病例数据，以此帮助医生缩小疾病判断范围并在处理复杂病例时系统可交叉对比多种信息从而有效降低误诊率，DeepSeek 已在临床诊断领域展现出显著价值并被全国十几所医院用于辅助诊疗。① 此外，DeepSeek 能够建立动态的知识框架，因此具有了人类中间层的认知能力。可以发现，过去的内容存储设备表现为静态的知识框架，输入与输出均是已被设定的内容。而 DeepSeek 具备实时演化的特征，能够凭借即时的网络搜索持续丰富数据库，并将新获取的信息与已有的内容进行语义关联。在这样的优势下，DeepSeek 能够基于原有数据库与实时信息，开展投资行为或金融风险评估。② 进一步而言，DeepSeek 具备对使用者提出的目标进行分解和生成策略的高级认知能力，能够对已经给出的内容进行自我质疑与修正，在作答时会自动提出一些假设性的问题，这种自我优化机制能够不断突破已有的逻辑结构，从而对给出的回答进行持续地调整和改进。

以 DeepSeek 作为代表的大型语言模型印证了斯蒂格勒所说的“第三持存”对于人类的“后种系”生成作用。数字时代之前的“第三持存”是以书籍、雕刻等为代表的物质性“第三持存”，在对于人类认知活动的辅助方面表现为被动性与外在性的特征，而生成式人工智能实现的是对人类认知过程本身的持存与再造。也就是说数字时代的第三持存是在人类思维成果外化存储的基础上，将人类的认知能力本身转化为可计算、可优化的技术装置，从而使得第三持存不再是单纯的被动容器，而是具备自我学习与自主决策能力的人类“合作者”。

二、认知解构层：DeepSeek 对“信息茧房”的技术解域

要理解 DeepSeek 对“信息茧房”的突破，就需要对“信息茧房”的概念及其内蕴作出一些澄清。美国学者桑斯坦认为，在数字时代的传播环境下，算法技术构建的精准信息筛选机制正在重塑人们的认知模式。他发现当代的信息获取方式显著受到算法推荐的影响，普遍表现出“选择性接触”倾向，也就是说个体会非自觉地过滤与既有观念相悖的信息，如同桑斯坦所言“只听我们选择的东西和愉悦我们的东西”③，正如蚕在桑叶间织茧自缚，用户也在看似开放的数字空间中塑造了自我封闭的信息获取方式，桑斯坦将这类现象形象地称为“信息茧房”。具体而言，“信息茧房”的特征表现为基于用户画像的定制化信息服务、群体内部的认知趋同化与跨圈层的信息流通障碍。④ 从“信息茧房”的分类形式来看，“信息茧房”分为用户主动选择信息源的“自主型茧房”——个体基于既有立场和兴趣偏好在选择性接触、屏蔽或强化特定信息中逐渐形成封闭的内容获取路径以及由算法推荐主导的“程序型茧房”——平台通过信息过滤和语义关联等技术将用户锁定在与其历史行为高度匹配的信息闭环中，从而导致外部观点难以突破算法设定的边界。值得注意的是，“自主型茧房”与“程序型茧房”相互交织共同作用于数字时代的信息传播，也就是说用户的主动筛选为算法提供了数据分析的基础，而算法的精准推送又进一步固化了用户的信息获取，这样一来便形成了双向强化的“信息茧房”现象。

当把“信息茧房”置于海德格尔的语境之中思考，就变成对此在“在世界之中存在”方式的追问。在海德格尔看来，现代技术的本质绝非纯粹工具性中介，而是作为“集置”。“集置”是那种摆置的聚

① 顾泳：《DeepSeek 本地化部署与医院系统深度对接》，《解放日报》2025 年 2 月 24 日，第六版。

② 任国省：《金融机构争相拥抱 DeepSeek》，《河北日报》2025 年 2 月 24 日，第十二版。

③ [美]凯斯·R. 桑斯坦：《信息乌托邦：众人如何生产知识》，毕竞悦译，法律出版社 2008 年版，第 8 页。

④ 李貌、韩璞庚：《数字时代“信息茧房”束缚下主体性的解构与重建》，《江苏社会科学》2024 年第 3 期，第 47-54 页。

集,“这种集置摆置着人,也即促逼着人”①,它使技术将人解蔽为可计算、可预测的持存物。随着数字技术的发展,海德格尔所论及的“集置”不再仅限于对物理世界的技术化支配,而是由互联网平台与算法技术将个体的每一次线上行为和决策纳入可测量、可预测的操作轨道之中。通过持续地数据采集与行为建模,算法技术不断将人的存在压缩为可量化的数据集合,进而构建出个体专属的“用户画像”,原本敞开的可能性领域逐渐被技术理性预先划定的轨道所促逼,数字平台中的用户此时表现为海德格尔在《存在与时间》中揭示的“常人”沉沦状态。社交媒体的点赞机制、短视频平台的成瘾性推送、搜索引擎的过滤气泡,都表明数字技术以更为隐蔽的方式强化着“常人”的统治,不仅将人的理解范式局限在既定框架内,更通过神经认知层面的奖赏机制重塑着主体的欲望结构。当用户深陷由算法编织的认知闭环时,他们实际上已成为海德格尔所说的“技术座架”中的持存物,其存在方式被简化为可预测的数据流。更为重要的是,“信息茧房”使此在陷于“自由选择”的幻象之中。海德格尔在《技术的追问》中指出,技术“对人类的威胁不只来自可能有致命作用的技术机械和装置。真正的威胁已经在人类的本质处触动了人类”。② 当我们以效率和便利为名拥抱算法推荐时,实际上已默许数字技术接管了自身“在世界之中存在”的方式。此时主体不再是主动建构意义世界,而是算法系统通过数据挖掘反向定义主体的认知边界。如同海德格尔笔下的农民面对机械耕作时丧失了对土地的原始经验,数字原住民在算法喂养下也逐渐丧失了遭遇“他异性”的能力。

在此,可以进一步追问,为何具有知识开放性特征的ChatGPT、Claude以及Grok等模型未能突破“信息茧房”现象?ChatGPT头顶上的“乌云”可能始终同资本座架下失控的竞赛以及技术开发者也无法理解、预测和控制其发展过程有关。③ 当人工智能被嵌入资本主义生产体系,其技术演进轨迹将必然受到资本增殖逻辑的规训。也就是说,ChatGPT、Claude等模型的局限性源于其资本基因中固有的封闭性。当科技巨头通过资本注入主导技术发展时,训练数据的筛选标准便无可避免地被商业利益所裹挟。资本通过建立数据准入机制将符合其增殖需求的语料赋予优先权重,使得知识开放性在算法层面遭到削弱。这意味着,商业机构提供的标准化内容将占据主导性,而具有公共价值但缺乏商业变现能力的信息则被边缘化。进一步从信息的输出环节来讲,付费用户与普通用户之间、企业级应用程序接口与公共免费版之间形成的响应速度、内容深度与知识完整性的级差,实质上是以技术参数为标尺的认知资源再分配。马克思在《1857—1858年经济学手稿》中指出,“资本是资产阶级社会的支配一切的经济权力”④,而今这种支配已从物质生产领域延伸至认知领域。在福柯的语境下,“知识为权力规定范围,权力为知识确定形式,两者相互支撑,知识是无处不在,权力也是无处不在的”。⑤ 这意味着,资本驱动的算法技术已经转化为新的权力形式,算法权力对知识生产与分配实施选择性过滤,并通过技术性规训深度嵌入个体的认知过程,将用户的信息获取路径预设在资本增殖的逻辑轨道之上。由此,“信息茧房”不再是偶发的算法偏差,而是资本与技术联手设置的“剩余数据”捕获牢笼,其封闭性和同质性已被固化为数字平台的默认运行模式,从而使得多元开放的公共话语场域在现有机制下难以自发重塑。总体来看,ChatGPT、Claude以及Grok等生成式人工智能所形塑的“信息茧房”现象,本质上是科技巨头通过控制数据资源的获取渠

① 马丁·海德格尔:《演讲与论文集》,孙周兴译,商务印书馆2018年版,第22页。
② 马丁·海德格尔:《演讲与论文集》,孙周兴译,商务印书馆2018年版,第28页
③ 张亮:《关于“ChatGPT”的历史唯物主义三重审思》,《理论探讨》2024年第3期,第150-159页。
④ 《马克思恩格斯全集》(第12卷),人民出版社1962年版,第758页。
⑤ 张国清:《他者的权利问题——知识—权力论的哲学批判》,《南京社会科学》2001年第10期,第14-18页。

道，迫使用户支付准入费用，从而在数字时代开展着类似于封建社会时期的新型“圈地运动”。正如齐泽克所言，“我们今天正在目睹新封建主义、封建资本主义的崛起。通过控制我们的公共资源，新领主（比尔·盖茨、埃隆·马斯克）的行为类似于封建主。……资本家的利润来源于雇佣工人通过生产商品所创造的剩余价值，而领主则不同，他通过垄断、胁迫和地租攫取价值。……数字平台是新的水磨坊，亿万富翁是新的领主，成千上万的工人和数十亿用户是新的农奴”。①

在此基础上，为何认为 DeepSeek 改变了“信息茧房”现象？又为何能够改变？一方面，DeepSeek 的技术解域体现为对于“算法黑箱”的突破。简单来说，“‘算法黑箱’是指这样一种系统或机制，它的输入和输出是明确的，但是输入到输出是如何转化却是部分或全部隐藏的”。② 以往推荐算法通过封闭的技术架构使用户的行为数据被转化为商业平台的价值生产资料，信息过滤机制便成为资本权力规训用户的隐性工具。而 DeepSeek 的开源特性为算法决策过程的透明化提供了可能，DeepSeek-R1 在发布时便已公开了模型的权重，并于近期再次公布《模型原理与训练方法说明》，披露了大模型 V3/R1 的训练流程。在预训练环节，模型依托大规模数据掌握语言规律与各类知识，而进入优化训练阶段时，会借助特定任务对应的数据集调整模型参数，以适配实际应用场景。为保障数据质量与使用安全，DeepSeek 声称采取了多项针对性举措，具体包括数据清洗、算法偏见检测以及匿名化处理等，同时明确强调模型的训练过程并未依赖用户个人信息。③ 值得注意的是，虽然目前 DeepSeek 仍有部分技术信息尚未完全公开，但其公开的模型权重、参数以及推理工具代码等，也已打破了传统封闭算法的技术垄断壁垒，使信息筛选机制从完全不可见的操控向部分可追溯、可验证的协作转变。另一方面，在知识生产维度，DeepSeek 构建了分布式的认知网络。DeepSeek 使用宽松的 MIT 协议供使用者自由下载并部署使用，也已发布各模型的完整技术报告供社区和研究人员参考④，允许个体或组织基于基础模型开发垂直应用，将知识生产从平台方转移至用户共同体。教育机构可构建学科知识图谱增强型模型，科研组织能开发专业文献分析工具，公民群体可创建社区信息验证系统，这些分布式节点通过应用程序接口形成互联的知识网络，使信息传播从封闭算法分割的“条纹空间”转向德勒兹所描述的“平滑空间”。平滑空间强调向量、方向、流动，这就类似于游牧民在沙漠中寻找水源和植被的空间轨迹。⑤ 在此意义上，DeepSeek 消除了中心化的知识边界，允许知识在网络中自由流动，也就是能够依据任务需求自主定义信息交互路径。值得注意的是，DeepSeek 的基础模型能够确保基本的信息质量基准，而垂直应用则根据具体场景进行适应性调整，可有效避免完全分散导致的结构失序并防止过度集中引发的“创新抑制”。

三、认知再构层：DeepSeek 引发“AI 幻觉”的技术桎梏

马克思指出，“人类始终只提出自己能够解决的任务，因为只要仔细考察就可以发现，任务本身，只有在解决它的物质条件已经存在或者至少是在形成过程中的时候，才会产生”。⑥ 也就是说，

① 蓝江：《技术封建主义假说及其悖谬——如何解读塞德里克·杜朗的〈技术封建主义〉？》，《广西师范大学学报（哲学社会科学版）》2025 年第 3 期，第 1-11 页。

② 张珺皓：《算法黑箱研究：基于认知科学的视角》，《科学学研究》2025 年第 9 期，第 1872-1880 页。

③ 《模型原理与训练方法说明》，https://cdn.deepseek.com/policies/zh-CN/model-algorithm-disclosure.html，2025 年 9 月 4 日访问。

④ 《模型原理与训练方法说明》，https://cdn.deepseek.com/policies/zh-CN/model-algorithm-disclosure.html，2025 年 9 月 4 日访问。

⑤ Gilles Deleuze & Felex Guatri，*A Thousand Plateaus：Capitalism and Schizophrenia*，The University of Minnesota Press，1987，p. 7.

⑥ 《马克思恩格斯文集》（第 2 卷），人民出版社 2009 年版，第 592 页。

在审视 DeepSeek 所具备的优势之后,应进一步将理论焦点转向 DeepSeek 在具体应用场景中引发的实践问题。正如埃隆·马斯克等千名全球科技人士在 ChatGPT 横空出世之际联名呼吁“高级 AI 可能意味着地球生命史上的深刻变革,我们应当投入相称的关注和资源对其进行规划和管理”。① 而 DeepSeek 的突破性进展尤其凸显出对大语言模型“幻觉”现象展开反思的学理需求,我们亟须在把握技术演进规律的基础上结合中国本土化实践经验,探索具有中国特色的数字技术治理路径。

在现阶段的技术演进中,“幻觉”问题已成为大语言模型的关键研究维度。新南威尔士大学的学者奥利弗·鲍恩认为,“当人们说 GPT 出现幻觉时,他们指的就是 GPT 对事实的这种篡改”。② 这意味着,大语言模型所生成的内容并非是简单的错误输出,而是以表面合理的话语包装事实偏差。此类偏差往往并非明显易见的低级错误,而是深度嵌入在文本逻辑与表达形式中的伪装性错误。正因如此,大语言模型生成的内容表现出较强的隐蔽性与验证困难性,容易形成严肃化虚构的现象。大语言模型的隐蔽性,体现在输出的内容在语言风格、论证逻辑与结构形式上与人类所撰写的内容高度相似,也就是说虚构的信息可以通过看似权威且连贯的方式嵌入语境之中,从而降低了用户对输出内容真实性的顾虑。从输出内容的验证困难性来看,大语言模型生成的内容会融合大量缺乏出处标注与追溯路径的信息,使得使用者在短时间内难以辨识生成内容的正确性,从而进一步加剧了错误内容的扩散与误信风险。总之,大语言模型“幻觉”现象的产生是由于它所生成的内容具有事实上的偏差、逻辑上的漏洞以及信息空洞的缺陷③,从而削弱了使用者的信息甄别能力与逻辑思辨能力,并在长期的使用中造成认知能力的退化与认知内容的污染。有鉴于此,目前中央网信办已发布 2025 年“清朗”系列专项行动的整治重点,在 AI 技术滥用乱象方面,明确指出应加强“AI 技术管理和信息内容管理,强化生成合成内容标识,打击借 AI 技术生成发布虚假信息、实施网络水军行为等问题,规范 AI 类应用网络生态”。④

基于 DeepSeek 思考“AI 幻觉”问题可以进一步从技术视角剖析其生成机理。DeepSeek 作为混合专家系统的典型代表,在能够理解语义的同时,也受限于数据参数与真实世界之间存在的解释性不足,也就是说 DeepSeek 所“理解”的世界并非来自于直接经验或具身感知,而是从大规模的文本数据中提取出来的语言关联方式。这意味着,当用户的提问超出大型语言模型的数据库时,模型会生成看似合理但缺乏事实依据的内容,即出现“幻觉”现象。进一步从模型的训练方式来看,DeepSeek“幻觉”现象的产生是由于对模型路径的依赖。在自回归生成过程中,“DeepSeek-V3 模型一次性可预测多个 Token,无需逐个预测,在推理过程中不使用 MTP 模块,只在训练过程中利用该模块约束模型的优化,MTP 的训练目标函数能够同时考虑多个 Token 的估计准确性,可以捕捉 Token 间的依赖关系”。⑤ 也就是说 DeepSeek 的预测会受到前序序列的强约束,这就导致 DeepSeek 在处理复杂逻辑推理任务时,局部最优解的选择可能偏离全局语义的一致性,此时幻觉

① 《马斯克等千名全球科技人士联名表布公开信 呼吁暂停推出更强大的 AI 系统》,中国日报网,http://cn.chinadaily.com.cn/a/202303/29/WS64240fba3102adab235e93.html,2025 年 2 月 27 日访问。

② 冯岩、[美]奥利弗·鲍恩:《当 AI 出现幻觉时》,《世界科学》2024 年第 3 期,第 28-30 页。

③ 陈万球、罗一人:《生成式人工智能的“知识幻觉”及其风险治理探论》,《上海市社会主义学院学报》2024 年第 4 期,第 38-51 页。

④ 2025 年“清朗”行动重点开展 8 项整治任务,中华人民共和国中央人民政府,https://www.gov.cn/yaowen/liebiao/202502/content_7004896.htm. 2025 年 2 月 27 日访问。

⑤ 蔡天琪、蔡恒进:《DeepSeek 的技术创新与生成式 AI 的能力上限》,《新疆师范大学学报(哲学社会科学版)》2025 年第 4 期,第 136-143 页。

现象的发生概率便会显著提升。从认知神经科学领域来看，“当前依托于大模型的生成式人工智能并未达到人类‘心智’层面的智能程度，其无法真正理解人类与它进行语言交互的深层含义”。[①] 也就是说，大语言模型并不具备人类心智中的意图识别、语境感知与意义建构的能力，它所生成的回应是对语言模式的统计拟合，而非建立在真正语义理解基础上的交互行为。这样一来，为维持回复的稳定性模型，便会主动生成符合其数据库的结果，而回复闭合的需求在信息不完备情境下极易产生虚构内容。总之，可以将 DeepSeek 所呈现的“幻觉”现象理解为，是由于其数据驱动的生成机制、训练过程中的路径依赖以及缺乏真正的语义理解能力这三重结构性约束所导致的问题。

鉴于目前 DeepSeek 产生的“幻觉”现象，与其说是技术演进的结果，不如说是技术实质上需要人类引导去实现其真正价值与可控应用。为了“构筑全民畅享的数字生活”[②]，2025 年 1 月全国数据工作会议强调应“努力完成‘十四五’规划和《数字中国建设整体布局规划》目标任务”[③]，由此从顶层设计层面为合理应对“AI 幻觉”提供了发展指导与根本遵循。这意味着应在持续推动技术发展的基础上严控大语言模型的幻觉根源。为此，我国在《新一代人工智能伦理规范》中明确提出“可信可控”的技术发展原则。[④] 从可控层面来看，要求人工智能技术的研发与部署必须保障人类对其目标的设定与行为边界的控制，防止人工智能在使用过程中出现不可预测的行为或违背伦理规范的内容。同时，需要进一步构建完整的风险评估机制包括提升算法透明度的措施与设计应急联动机制，确保在人工智能出现偏差或歧视时能够及时地介入和纠偏。此外，强化对关键技术的审查与测试并推动可解释性与可追溯性技术标准的实践也是实现“可控”目标的重要路径。在可信层面，中国信通院人工智能研究所在《大规模预训练模型技术和应用评估方法》系列标准中明确提出，基础软硬件可信、数据可信、模型可信与应用可信的四点要求[⑤]，由此为构建人工智能信任体系的建设提供了方向遵循。在此意义上，基础软硬件的可信强调算法技术和运行平台的安全可靠，使其能够确保在部署和执行过程中不受系统漏洞、恶意植入代码或其他潜在安全隐患的干扰。数据可信则指向了数据采集、标注与处理过程中的合法性与完整性，也就是要防止训练过程中因为数据偏差而引发模型的错误输出。模型可信主要涵盖模型结构的合理性与训练过程的可解释性，确保模型在不同应用场景中具备稳定且可预测的行为表现。而应用可信则涉及大语言模型在具体业务流程中要做到合规、安全与符合伦理规范，确保使用者的知情权与隐私权。总体而言，遵循“可控可信”的技术发展原则能够有效降低“AI 幻觉”的出现概率与潜在风险。

四、结语

在数字文明建设的宏观视野下，对 DeepSeek 在认知维度的审视，既需正视其作为“第三持存”对人类认知能力的辅助性建构价值，也不能忽视其技术特性所潜藏的多重风险，避免陷入技术乐观主义的认知偏差。从实践层面来讲，DeepSeek 虽为突破“算法黑箱”提供了可行路径，但其文本处

① 陈万球、罗一人：《生成式人工智能的“知识幻觉”及其风险治理探论》，《上海市社会主义学院学报》2024 年第 4 期，第 38-51 页。

② 《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》，中国人大网，http://www.npc.gov.cn/npc/c2/c30834/202312/t20231227_433830.html，2025 年 2 月 27 日访问。

③ 《全国数据工作会议在京召开》，中华人民共和国国家发展和改革委员会，https://www.ndrc.gov.cn/fzggw/wld/llh/zyhd/202501/t20250115_1395691.html，2025 年 2 月 27 日访问。

④ 《新一代人工智能伦理规范发布》，中华人民共和国科学技术部，https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html?ref=salesforce-research，2025 年 2 月 27 日访问。

⑤ 《中国通信标准化协会关于发布〈面向行业的大规模预训练模型通用要求 第 1 部分：金融〉等 51 项团体标准的公告》，https://www.ccsa.org.cn/detail/id=54926&title，2025 年 2 月 27 日访问。

理的技术边界既使其在专业领域的应用依赖人类对输出内容的事实性校验,也因“AI 幻觉”在事实性场景中的误导性输出引发决策偏差和风险误判等连锁问题。同时,DeepSeek 对认知劳动的部分替代虽提升了生产效率,但也可能导致人类在基础认知环节的能力退化,削弱主体的批判性思维进而陷入“技术依赖”的被动局面。这种对风险的审慎认知并非否定技术进步的价值,而是确保 DeepSeek 正向赋能的必要前提。也就是说,对 DeepSeek 的治理需以“风险防控”与“价值实现”的辩证统一为导向,一方面,需强化“可信可控”原则的落地实践,从基础软硬件安全、训练数据合规性、模型行为可解释性等维度,构建全流程风险评估机制;另一方面,需通过全民数字素养教育,提升公众对 AI 生成内容的甄别能力,避免因过度依赖技术而丧失自主认知判断。随着人工智能技术的加速演进,DeepSeek 这类大型语言模型对社会认知生态的影响将愈发深远。唯有始终坚守以人为本、技术为辅的核心原则,在多方协同中完善伦理规范与安全防护体系,既发挥其在认知建构、信息解域中的技术优势,又通过制度设计与能力建设化解潜在风险,真正推动数字技术深度融入教育、医疗、环保等民生领域,最终实现人工智能服务于人的全面发展与社会进步的根本目标。

Triple Logic of DeepSeek Technological Construction: A Cognitive Perspective

MA Junfeng, WEN Zhaolun

(*School of Marxism, Northwest Normal University, Lanzhou* 730070, *China*)

Abstract: DeepSeek has become one of the most representative large language models in the digital age, demanding continuous deepening of research on DeepSeek in academia. To unfold the multilayered construction of DeepSeek from a cognitive perspective, it is essential to emphasize DeepSeek's role as a "tertiary retention" in constructing human cognitive abilities and to focus on its interactive relationship with human cognitive labor. This paper dialectically analyzes DeepSeek's technical potential to transcend the logic of capital and deconstruct the "information cocoons" based on its breaking through the "black box algorithm" and shaping distributed cognitive networks. Meanwhile, the paper also explores its dual properties of reconstructing "AI illusions" through its structural mechanisms, training methods, and reasoning characteristics. Therefore, we can seek to reduce illusion risks based on the "trustworthy and controllable" principles of reliable foundational hardware and software, trustworthy data, reliable models, and trustworthy applications. Thus, with a dynamic framework, we can reveal the cognitive evolution logic of large language models, clarify human value judgments and subject status in human-machine collaboration, thereby establishing value coordinates for cognitive governance in the era of artificial intelligence.

Key words: DeepSeek; tertiary retention; information cocoons; AI illusions

(责任编辑:傅　游)