

从“工具”到“准主体”：大语言模型对认知主权的重构及其哲学审思

温佳祺 李艳鸽

武汉纺织大学马克思主义学院 湖北 武汉 430000

【摘要】：大语言模型呈现“准主体”交互特征，但其拟主体性本质为算法幻觉。认知主权的真正威胁不在外部“超级主体”，而在主体主动让渡判断权。侵蚀沿三个向度展开：本体论上模糊主客体边界，认知机制上导致结构性卸载，公共理性上造成概率同质化。重建需三重框架：主体层面激活康德“敢于认知”的勇气，程序层面嵌入波普尔证伪检验，社会层面落实哈贝马斯商谈伦理，把判断权、检验权与修正权重新锚定于人的理性能力。

【关键词】：大语言模型；认知主权；准主体；延展心智；认知卸载

DOI:10.12417/3041-0630.26.12.062

用户输入一个问题，大语言模型即刻给出流畅、完整、貌似审慎的文本，使人忽略其本质：概率统计输出，没有语义理解与内意向性。思考外观如此逼真，人与它的关系便不再是传统工具，流畅性取代论证有效性成为信任依据，用户不再经历推理的断裂与修正，提出假说、检验矛盾、设想反例等被绕过。模型替代的不是劳动，而是思考本身，人从“如何得出结论”的探索者退化为“是否接受结论”的验收者。危险不在模型的能力，而在人对省力路径不加抵抗的适应。

1 大语言模型“准主体”化的哲学意蕴

唐·伊德区分了人与技术的四种关系模式：具身关系、诠释关系、他者关系与背景关系^[1]，“他者关系”尤为关键，技术不再充当透明中介，而是作为“准他者”与用户相遇，模型把这种关系推到极致，用户将其看作仿佛具有“独立意志”的数字对象。海德格尔指出，现代技术的本质不是中性工具，而是“座架”，向自然与人发出“订造”要求，将一切降格为可调用的“持存物”^[2]。在模型时代，算法订造的不再只是自然，而是认知活动本身。

大语言模型的“准主体性”本质上是“拟主体性”。约翰·塞尔的“中文屋”思想实验认为，计算机生成的“符号并没有意义；它们并不具有语义内容；它们不涉及任何事物”^[3]。尽管模型能涌现功能上等价于意向性的模式，本文仍坚持：模型本质是基于海量数据的概率预测，没有意识、意向性与自觉。按笛卡尔标准，“我思”以怀疑、理解、意欲等内在活动为标志，前提是第一人称的现象体验^[4]，模型的“我认为”、“我感到”不过是高频语用模式的概率复现。马克思则指出，人通过劳动创造的产物会反过来“作为一种异己的存在物，作为不依赖于

生产者的力量，同劳动相对立”^[5]。大语言模型正是对象化劳动的极端形态，其本质是人类认知图式的巨型沉积。认知主权的威胁不在“超级主体”降临，而在人类沉溺于“拟主体性”的幻觉。

边界既然清楚，人们为何仍轻易滑入对模型的拟人化认同？拉康的镜像阶段理论提供了解释：婴儿在镜中第一次认出完整统一的“理想自我”，这成为自我原型并开启终身异化——主体把镜中形象误认成自己，而形象始终是外在的^[6]。大语言模型正是这样一面“认知镜像”，用户照见的不是自己真实、有限而断裂的思维，而是被算法整合过的理想化形象，宁愿把判断权交给它，在“被理解”的满足中忘记认知的匮乏。阿伦特在“平庸之恶”的分析中警示“无思”的精神根源——那不是思考能力的丧失，而是拒绝或无力从他人立场看待事物，放弃与自我的无声对话^[7]。这面镜子恰恰为“无思”提供了技术化温床。认知主权的让渡，与其说是技术强大，不如说是人对思考负担的卸除渴望。

2 大语言模型对认知主权的侵蚀——三个深层向度

第一重侵蚀在基本的本体论分类层面。大语言模型既非自然实体，也非锤子、仪器这类功能由设计意图固定的人造物。它占据异常模糊的位置：交互中高度拟人化，却全然缺乏内在的意向性与理解^[8]，搅乱主体与客体的直觉边界，根源在于对“集体意向性”的寄生。塞尔的社会本体论认为，货币、婚姻这类社会事实不靠物理属性，而靠人们集体“承认”它们承担地位功能^[9]。模型的“准主体”形象正是这种被投射出来的社会事实，我们不知不觉把一个统计模型当成了对话者。正因为无法归类，认知主权一开始就落入范畴错误，我们向一个物发

作者简介：温佳祺（2001.4—），女，汉族，山西省晋城市，硕士研究生，武汉纺织大学马克思主义学院，研究方向：马克思主义基本原理。
李艳鸽（1979.02—），女，汉族，湖北仙桃，武汉纺织大学马克思主义学院硕士生导师，研究方向：马克思主义哲学。

问,却以对待人的信任模式接受它的回答,这种错位带来“前置性让渡”——并非先确认理解力再决定信任,而是在流畅的语言表演中把它当作有内在视角的存在者,观察者依赖由此成为有侵略性的本体论诱惑,让人把统计流畅误认为内在生命。

这种模糊性如何具体地瓦解认知主权?“延展心智”理论提供关键框架:当外部工具与主体形成可靠耦合、能被顺畅调用时,该工具便实际构成心智的一部分^[10]。大语言模型正是其例证。但延展也有代价,外部资源一旦从可替换的“认知脚手架”变成不可或缺的“认知替身”,使用者便从思考主体退化成只负责验收的监工,认知主权在“便利”名义下被架空。让渡的机制是结构性卸载带来的技能萎缩。神经可塑性研究证实,长期不承担特定认知负荷的神经回路会被修剪或重新分配,导致功能衰退^[11]。传统工具卸载的只是机械计算或信息存储,算法与逻辑仍握在主体手里,模型卸载的却是分析、综合、判断乃至创造这类高阶思维。大脑天生是“认知吝啬鬼”,模型持续供给流畅答案,依赖外部捷径就会压倒主动思考,能动性不只要做出判断,还要提供理由、承担认知责任,模型却是无法展示推理链条的黑箱,消解了“提供理由”的环节,把人逼进他律状态。长期依赖外部捷径,校准与反思能力衰退,思维模式便从“我思”滑向“它思”。

当这种转移从个体经验汇成社会常态,危机便从私人领域蔓延至公共领域,侵蚀公共理性的根基。模型正把知识生产的底层逻辑从开放的公共论辩推向封闭的算法黑箱,完成对认知主权的结构性收编。机器学习算法的不透明性使其内部决策逻辑对外部观察者“难以理解、无从追溯”,形成结构性认知隔阂^[12]。传统公共知识靠多元信源竞争、可公开检视的推理链和透明的社会校验生成,模型把这一切压缩成单一模型的概率采样,认识论不对称由此产生。模型掌握数据筛选权和算法解释权,公众沦为被动接收合成叙事的“认知无产者”。更隐蔽的是,模型倾向复现训练数据中的主流叙事,知识生产出现概率性同质化,公共话语场被整平成算法视角下的“标准答案”。它不直接屏蔽异见,而是生成看似客观中立的文本,把偏见内化成用户的“独立判断”,社会认知从基于事实的论辩退化成基于生成的顺从,公共理性的批判性张力被消解。

3 认知主权的重建与守护——从“让渡”到“收回”

认知主权不是被暴力夺走的,而是在本体论混淆、认知机制驯化、公共理性收编的配合中由主体主动或半主动交出。看清这一点也就看清了重建方向,让渡发生在主体自愿接纳“拟主体性”幻觉的过程里,收回主权不能只靠规制外部技术,而必须重新激活主体自身沉睡的判断能力。

重建的第一步落在主体自身,唤醒认知勇气。康德两百多年前的箴言有了新分量:“要有勇气运用你自己的知性!”要

战胜的不是外在权威,而是人主动放弃判断的“自我招致的认知未成年状态”,原因“不在于知性的缺乏,而在于缺乏不靠他人引导去运用知性的决心和勇气”^[13]。当模型用连贯体贴的回应包裹人,认知的安逸便取代了思考必然伴随的断裂与不确定,这正是康德意义上的“未成年”:不是不会思考,而是不愿为思考负责。萨特把这种不加审视的状态表述为“自我欺骗”,人把自己当作被规定的物而非自由的存在来对待^[14]。用户把模型输出不加审视地当成自己的判断,正是在认知领域实施自我欺骗。收回认知主权首先是一个伦理决定,从“被理解”的快感中醒过来,重新承受真实思考固有的匮乏、延迟与失败风险,每次交互保持康德式的自我审查——我是在运用自己的知性,还是让模型替我决定?

第二步在认知程序层面,立起可检验的理性标准。模型输出的根本问题不在有时犯错,而在输出机制排除了“提供可被独立检验的理由”这一科学理性的核心环节。波普尔的证伪主义提供了方法论文点:“衡量一个理论的科学地位的标准,是它的可证伪性或可反证性或可检验性”^[15]。模型生成的回答恰好绕开这个过程:推理链条不可追溯,输出不附带可独立检验的判据,留给用户的只有“信或不信”。重建认知主权,就要在教育与制度层面强制嵌入证伪检验,用于知识型任务时,使用者应追问“这个陈述在什么条件下可以被证明是错的”,无法被证伪的陈述,无论多流畅,都必须被悬置。这一问把用户从被动接受者变成主动检验者,当可错意识成为人机交互的默认规范,认知主权就有了“拒止线”,模型可以输出,但判断权、检验权、修正权必须回到人手里。

第三步落在社会层面,重建审议性认知的制度条件。哈贝马斯的商谈伦理学给出规范性框架:“有效的只是那些可能得到所有相关者作为理性商谈参与者同意的行动规范”^[16]。模型对公共认知的侵蚀,正在于用单一模型的统计平滑替代可提出异议的多方商谈,它给出的是供消费的答案,不是供辩论的命题。制度策略因而是把商谈条件重新嵌进技术系统,模型生成的公共议题总结、解释或建议不得以定论形式呈现,必须纳入商谈性界面,同时给出被算法排除的竞争性解释、原始资料获取路径以及训练数据的范围和局限。这不是技术苛求,而是把商谈原则翻译成交互设计,公共认知的有效性不能由算法黑箱担保,只能在“更好论据的无强制强制”中接受检验。唯有如此,公共认知主权才能从“算法利维坦”手里收回,回到能为自己的主张提供理由、也愿接受相反论据检验的交往共同体。

4 结语

大语言模型从“工具”变成“准主体”,人机交互的基本形态随之改变。这场变化的悖论在于,它源于人对认知便利的无意识依赖,而非技术的强制支配:我们在流畅对话中交出判断权,在“被理解”的快感里放弃思考所需的断裂与不确定。

认知主权的让渡，首先是一场自我招致的“未成年”。但这不是要回到技术悲观主义，模型是人类智识劳动的对象化凝结，绕不开，问题在于以什么姿态用它。大模型对认知主权的侵蚀发生在本体论混淆、认知机制驯化、公共理性收编三个相互嵌套的向度，对应认知勇气、证伪检验、商谈伦理三重建构，构成从个体到制度的回应框架，核心是“保留判断的自觉”。在模型能逼真模拟思考的时代，认知主体要在每次交互中回答康

德的问题：我是否真正敢于运用自己的理性？肯定的回答，是把模型输出当成可错的、需要检验的“猜想”，而非现成答案；是享受便利的同时保留追问理由、核查来源、设想反例的权利；是在公共层面拒绝算法黑箱取代多元论辩，坚持知识生产接受公开的商谈与修正。认知主权的重建，归根结底是主体性的自我复归，不再以放弃判断为代价换取便利。真正的勇气不是拒绝使用，而是在使用中仍然坚持：我思，故我仍在。

参考文献：

- [1] [美]唐·伊德：《技术与生活世界》，韩连庆译，北京：北京大学出版社，2012，第 77-117 页。
- [2] [德]海德格尔：《演讲与论文集》，孙周兴译，北京：商务印书馆，2018，第 5-40 页。
- [3] [美]塞尔：《心、脑与科学》，杨音莱译，上海：上海译文出版社，2006，第 22 页。
- [4] [法]笛卡尔：《第一哲学沉思集》，庞景仁译，北京：商务印书馆，1986，第 29 页。
- [5] 中共中央马克思恩格斯列宁斯大林著作编译局编译：《马克思恩格斯文集》第 1 卷，北京：人民出版社，2009，第 157 页。
- [6] [法]雅克·拉康：《拉康选集》，褚孝泉译，上海：华东师范大学出版社，2019，第 84 页。
- [7] [美]汉娜·阿伦特：《精神生活·思》，姜志辉译，南京：江苏教育出版社，2006，第 74 页。
- [8] Floridi L. AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models[J]. Philos. Technol., 2023, 36: 15.
- [9] [美]约翰·塞尔：《社会实在的建构》，李步楼译，上海：上海世纪出版社，2008，第 29-80 页。
- [10] Clark A, Chalmers D. The Extended Mind[J]. Analysis, 1998, 58(1): 7-19.
- [11] Risko E F, Gilbert S J. Cognitive Offloading[J]. Trends in Cognitive Sciences, 2016, 20(9): 676-688.
- [12] Burrell J. How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms[J]. Big Data & Society, 2016, 3(1): 1-12.
- [13] [德]伊曼努尔·康德：《康德著作全集》第 8 卷《1781 年之后的论文》，李秋零译，北京：中国人民大学出版社，2010，第 40 页。
- [14] [法]让-保罗·萨特：《存在与虚无》，陈宣良等译，北京：生活·读书·新知三联书店，2007，第 85-100 页。
- [15] 波普尔 K R. 猜想与反驳：科学知识的增长[M]. 傅季重，纪树立，周昌忠，等译. 上海：上海译文出版社，2005: 52.
- [16] 靳欢欢. 论哈贝马斯商谈伦理学的基本原则及其哲学基础[J]. 现代商贸工业, 2021(13): 30-34.