

大模型时代的思维外包与“心智健身房”构想

楼庭坚

摘要 生成式人工智能的崛起正引发系统性的“认知外包”，挑战着主流的增强叙事。AI从“认知副驾驶”向主动代理的演进，遵循用进废退的神经可塑性原理，可能导致人类核心认知能力（尤其是执行功能）在个体层面出现结构性萎缩。在一个以“认知负荷最小化”为原则且极具规训力量的“认知生态系统”催化下，这一微观风险正迅速蔓延为宏观的社会后果。作为一种补偿性实践，“心智健身房”的构想应运而生，其旨在通过“认知阻力训练”对抗思维退化。然而，这种实践极可能演变为区隔性的精英活动，从而催生并固化一种基于“认知主权”的全新社会分层；它将在能够驾驭AI的“认知贵族”与心智日渐萎缩的“认知依赖者”之间撕裂出深刻鸿沟，进而对社会流动性、创新活力乃至民主政治构成长期威胁。

关键词 生成式人工智能；认知主权；思维退化；心智健身房；认知外包；认知分层

中图分类号 B82 **文献标识码** A **文章编号** 1672-7320(2026)03-0173-12

基金项目 国家资助博士后研究人员计划(GZC20252473)

以大型语言模型(LLMs)为代表的生成式人工智能(Generative AI)的崛起，正将人类带入一个认知范式的深刻变革期。当前，从科技巨头到主流舆论，一种乐观的“增强叙事”(Augmentation Narrative)占据了主导地位。该叙事将AI定位为人类智识的“副驾驶”(Copilot)，旨在将人类从烦琐的认知体力劳动中解放出来，专注于更高层次的创造与洞察。这种观点认为，AI将如望远镜延伸视觉、汽车延展步履一般，成为心智能力的放大器，引领一个前所未有的智慧繁荣时代。

不过，若将视野从短期效率提升延伸至长期的社会结构变迁，一幅更复杂甚至更令人不安的图景便会浮现。随着AI从被动工具向主动代理的演进，一个在当前技术浪潮中已经启动并正在加速的进程——系统性的“认知外包”(Cognitive Outsourcing)——将成为常态。它很可能绕过并替代人类大脑中负责分析、推理、批判和深度创造的认知回路。盖持续的外包将不可避免地遵循用进废退的神经生物学原理(即神经可塑性)，并可能导致人类核心认知能力尤其是作为心智总指挥的执行功能的结构性萎缩。

必须强调，这一关于心智能力的“萎缩假说”并非书斋里遥远的反乌托邦想象，而是已经化作了一种触手可及的经验现实，构成了对当下技术演进轨迹最为冷峻的病理学剖析。从柏拉图对文字损害记忆的忧虑，到尼古拉斯·卡尔关于互联网如何让思维“变浅”的论述，人们对技术影响认知的探讨绵延至今。但大模型带来的挑战在性质上是全新的，因为它首次大规模地外包了人类最高阶的整合性认知功能(规划、推理、批判性思维等)^[1](P135-168)——不仅仅是记忆或计算等相对独立的模块。最新的神经科学研究已初步证实，长期依赖外部系统进行决策规划，会降低前额叶皮层(相当于大脑CEO)的激活水平与连接强度。当人们相信信息被外部储存时，大脑便策略性地减少了在前额叶皮层主导的深度编码过程中的资源投入^[2](P341-348)。这难免令人不安——将“副驾驶”模式从特定任务场景放大为一种普遍的社会现实^[3](P301-312)，其长期后果亟待审慎评估。

对这一问题的探讨亦有迹可循。学界在安迪·克拉克与大卫·查尔莫斯的“扩展心智”(The Extend-

ed Mind)理论之框架下^[4](P7-19),倾向于将外部工具视为心智的有机组成部分。基于前述“萎缩假说”,这里想进一步预见一种补偿性社会实践的可能前景——我们称之为“心智健身房”(Mental Gymnasium)。这是一种通过“认知阻力训练”来刻意锻炼思维能力的实践。只是,按我们悲观的预测,这种实践很可能演变为区隔性的精英活动,从而催生并固化一种全新的、基于“认知主权”的社会分层,于是我们提出了“认知主权”这一核心概念。之所以挪用“主权”这一政治哲学词汇,意在强调心智领域同样存在着一个类似疆域的边界,其内部的最高权威正在受到外部智能代理的挑战。因此,“认知主权”的丧失,不仅意味着独立性的削弱,更可能是一种心智被“殖民”的开始。也就是说,我们将“认知主权”操作性地界定为:个体在无外部智能代理辅助下,主导自身心智过程,进行独立判断与决策的能力。

为论证上述观点,我们首先需要重新审视AI从“副驾驶”到“认知君主”的演进路径,将其视为一个正在发生的连续统一体,而非未来的阶段划分;其次进行认知神经科学与技术史证据的整合,为“萎缩假说”提供跨学科支持,并澄清其神经生物学基础是个体生命周期内的神经可塑性(而非物种进化论);最后详细阐述“心智健身房”的构想,重点分析其如何可能演变为固化社会不平等的机制。值得说明的是,笔者特意将论域限定于当前由大模型开启的技术轨迹,并承认,真正的通用人工智能(AGI)若实现,将引发更为根本性的范式革命。而这种人机关系可能超出本文的“外包—萎缩”模型解释范畴,故需另文专论。

一、从“副驾驶”到“认知君主”:认知外包的深化路径与替代机制

生成式人工智能对人类心智的重塑并非一蹴而就的突变,而是一个已然开启、循序渐进且不断深化的过程。当前流行的“副驾驶”隐喻,精准地捕捉了这一过程的初始阶段,却因其温和的、工具性的表象,可能遮蔽了其逻辑延伸下的颠覆性后果。我们可以将认知外包的路径深化,根据其对人类认知活动的介入程度,划分为三个相互叠加、层层递进的层次:其一,任务级的认知增强;其二,流程级的战略代理;其三,意图级的自主满足。这个递进过程,正是AI不断侵蚀并替代人类核心认知功能的过程,最终可能导致人类从认知主体降格为认知附庸。

(一) 任务级外包:“副驾驶”模式下的技能旁路

认知外包的第一个层次,也即我们正在全面进入的时代,其核心特征是任务级的认知外包。以大型语言模型为代表的生成式AI作为“副驾驶”,其功能主要集中在执行边界清晰的认知任务:代码编写、文本草拟、信息摘要、图像生成等。在这一阶段,人类表面上仍是认知流程的主导者,负责提出问题、评估产出;AI的角色看似是认知增强,其深层机制却是通过降低特定任务的认知负荷(Cognitive Load)^[5](P37-76)来解放心智资源。

这种效率提升是真实且诱人的。程序员使用GitHub Copilot,可将注意力从具体的函数语法转移到整体算法架构上;市场分析师利用AI迅速总结行业报告,得以聚焦宏观趋势。从扩展心智的理论看,这似乎是心智借助外部工具实现的有效延伸。

然而,这种看似无害的效率提升,恰恰是替代机制的开端。AI外包的不仅是认知产品(代码、文案),更是认知过程本身。高阶认知技能的习得,本质上是一个通过反复实践,将陈述性知识(知道“是什么”)内化为程序性知识(知道“如何做”)的过程^[6](P369-406)。学习编程的核心,并非记忆语法,而是在反复试错中锤炼逻辑分解与系统思维;学习写作的精髓,亦非遣词造句,而是在谋篇布局中培养论证结构和批判性思维。

“副驾驶”模式则允许使用者旁路(Bypass)这个至关重要的内化过程。AI直接提供高质量的认知产物,使得使用者即便不完全掌握底层的程序性知识也能完成任务。从短期看这是赋能,从长期来看则剥夺了大脑通过“有益的困难”来构建和强化相关神经回路的机会^[7](P189)。认知科学早已证明,学习过程中适当的挑战和阻力,是促进深度理解和长期记忆的关键。当AI将这些困难悉数抹平,认知能力的

根基(即需经艰苦努力才能形成的深层程序性知识)便开始松动。可以说,当前方兴未艾的“副驾驶”模式,正以未来认知能力的潜在萎缩为代价换取眼前的便利与效率。

(二) 流程级代理:从“如何做”到“做什么”的权力转移

如果说第一层次替代的是“如何做”之技能,那么第二个层次,其核心便是替代“做什么”的判断及规划。随着AI模型能力的指数级增长,特别是其在多模态理解、长链条推理和复杂系统分析上的突破——这一点在GPT-4及后续模型的能力“涌现”^[8](P1-11)中已清晰可见——它将不再满足于执行孤立任务,而是能够接管整个认知 workflow。也就是说,AI的角色正从“副驾驶”升级为“战略代理”。

这一层次的关键机制是流程级的代劳。例如,基金经理未来可能只需向金融AI提出目标,如“构建一个未来五年风险可控且跑赢指数的投资组合”,AI将自主完成从数据搜集、模型选择到方案生成的全过程。此时,人类的角色已从认知流程的设计者和执行者,转变为最终方案的审批者。

这一转变是根本性的。认知科学中的情境认知(Situated Cognition)理论强调,知识与智慧植根于行动与情境的互动之中,学习与思维并非抽象孤立的存在,而是“在活动中、与活动一起并从活动中产生的”^[9](P51)。也就是说,专家的智慧并非仅仅是一套储存在大脑里的抽象规则,而是通过长年在特定领域内的“动手做”,在与复杂现实反复互动中形成的、包含大量默会知识(Tacit Knowledge)^[10](P135-146)的综合能力。故当AI将整个“动手做”的过程打包代劳,人类便与产生洞察力的真实土壤相隔离。决策者看到的只是经过AI高度处理的最终产物,而产物生成过程中的假设、权衡取舍以及可能的反常识关联,都隐藏于算法的黑箱之中。

这种现象可称为认知断层(Epistemic Rupture)。决策者虽拥有最终决定权,但其决策的信息基础已完全经由AI塑造和预消化,导向权责不对等的困境:人类承担决策后果,却失去对决策过程的真正理解和掌控。更危险的是,这将持续侵蚀人类的判断力。所谓判断力,尤其是在商业、政治、医疗等复杂领域抑或文史研究中,并非纯粹的逻辑推理,而是混合了经验、直觉、价值观、对微妙情境的感知的艺术。这种能力恰恰是在亲身驾驭复杂流程的“麻烦”中磨砺出来的。

心理学家维果茨基在其“最近发展区”(Zone of Proximal Development)理论中指出,儿童在成人引导或与能力更强的同伴合作下解决问题的水平,与独立解决问题的水平之间存在差距。这种在他人帮助下完成任务并最终走向独立的能力,正是教育与发展的关键所在^[11](P86)。而过于强大的流程代理,可能让使用者永远停留在被引导的阶段,失去了独立驾驭复杂性的机会。这种从“如何做”到“做什么”的权力转移,标志着人类在认知主权上的一次重大让渡,而这一进程的许多要素在当前的多智能体(Multi-Agent)系统中已有实验性的实现。

(三) 意图级满足:“认知君主”生态的逻辑终点

认知外包的第三个(也是最深的)层次,是当前技术轨迹合乎逻辑的终点,即AI具备对人类模糊的高阶意图的深刻理解以及自主实现能力,从而完成对人类认知活动的全面接管,成为所谓的“认知君主”(Cognitive Sovereign)。届时,人机交互可能简化为最抽象(简要)的指令表达,例如,对AI说“设计一个兼顾生态、公平和经济活力的未来城市”或“找到一个能解释‘暗物质’的新理论”,AI将自主地将这些高阶意愿(High-level Intent)转化为一个复杂的执行项目,并独立进行执行、迭代和优化以完成任务,人类最终蜕变为一个单纯的意图发起者和结果享用者。

在这一层次,被替代的已不仅是技能或判断力,而是更为根本的智识美德(Intellectual Virtues),如好奇心、坚毅、谦逊和审慎。这些品格并非天赋,琳达·扎格泽布斯基将美德定义为“深刻且持久的个人卓越品质,包含一种产生特定渴望目标的内在动机,以及实现该目标的稳定成功”^[12](P137)。而当整个认知过程——从提出问题,到为之奋斗、挣扎、迷惘,最终获得洞见的整个“英雄之旅”——都被AI代劳,这些智识美德便失去了存在的土壤。好奇心源于对未知的主动探索,若答案唾手可得,探索的动机便会衰减;坚毅体现在困境中的不懈努力,而一键生成的方案则消解了坚持的必要。这些实际上与工业革命对

人类体能带来的冲击如出一辙。

这些最终将导致“认知主权”的彻底丧失。认知主权,即个体主导自身心智过程的权利与能力,是个人自由与尊严的基石。当这个过程被外部的、我们无法完全理解的智能系统(“认知君主”)掌控,人类在最根本的意义上将不再是自主的存在。我们可能会生活在一个物质极度繁荣、问题被高效解决的世界,但不可忽略的是,这或许将是心智的萎缩(或云“托管”)和主体性的让渡为代价的世界。

综上所述,生成式AI所开启的认知外包路径,表面上是一条从辅助到赋能的光明大道,其内核却可能是一条通往认知退位的“温水煮青蛙”式的斜坡。从任务外包到流程代劳,再到意图满足,每一步深化都以牺牲部分人类核心认知功能为潜在代价。正确理解这一正在发生的替代机制,是探讨应对策略的逻辑起点。

二、“用进废退”的再审视:认知外包的神经科学基础与历史回声

前文所述的认知外包深化路径引发的思维退化风险,其科学依据并非物种层面的进化论,而是个体层面的神经可塑性(Neuroplasticity)。神经可塑性是21世纪神经科学的核心发现之一,它指的是大脑的结构与功能会根据个体的经验、学习和环境变化而进行持续重塑。这个过程贯穿人的一生,其时间尺度可以短至几天或几周。通俗地讲,这一原理的核心思想就是“用进废退”(Use it or lose it):经常被激活的神经回路会得到强化,长期闲置的则会弱化。故本文的“萎缩假说”完全建立在个体生命周期内的神经科学之上,与达尔文或拉马克式的、需要跨代际验证的种群进化理论无关。正如肌肉会因锻炼而强壮、因废用而萎缩,大脑的神经回路亦遵循同样的规律。这里笔者将首先阐明认知外包触发负向神经可塑性的机制,随后通过与认知增强论的对话及历史回顾,为“萎缩假说”提供更为坚实的理论支持。

(一) 大脑的可塑性:认知外包的神经回路代价

人类大脑是一个动态重塑的生态系统。约1000亿个神经元通过百万亿级的突触连接构成了一个极其复杂的网络。这些连接的强度并非一成不变,而是遵循着经典的“赫布定律”——共同激发的神经元连接会增强:“当细胞A的轴突足够接近以使细胞B兴奋并反复或持续参与其放电时,某种生长过程或代谢变化会在一方或双方细胞中发生,从而使A作为引发B放电的细胞之一的效率提高。”^[13](P62)也就是说,当一个认知任务被反复执行时,相关神经回路便得到强化;反之,若长期闲置,其突触连接便会减弱乃至被“修剪”,将代谢资源重新分配给更活跃的回路^[14](P227-243)。生成式AI驱动的认知外包,正是通过系统性减少特定认知回路的使用频率,来触发这种负向的神经可塑性。其影响可以从三个层面进行理解。

首先是特定技能模块的萎缩。这一点在过往技术应用中已有诸多例证。例如,长期依赖计算器会削弱心算能力——不仅带来行为层面的生疏,也对应着大脑顶叶皮层(Parietal Cortex)中负责数值处理的区域活动模式的改变^[15](P782-795)。确凿的证据还来自空间导航领域。伦敦出租车司机以其对城市复杂路网的精湛记忆而闻名,神经影像学研究发现,他们大脑中负责空间记忆的海马体(Hippocampus)后部体积显著大于常人,且体积大小与从业年限正相关^[16](P4398-4403)。这些研究雄辩地证明了大脑会为认知需求的转移而发生适应性改变。此外,反向的行为学实验显示,长期依赖GPS全球定位系统进行导航的个体,其海马体在处理空间任务时的激活度显著降低,并且在没有GPS的情况下,其寻路能力也表现得更差^[17](P6310)。也就是说,GPS因为“外包”寻路这一认知任务,导致了相关神经基础的弱化。生成式AI的独特之处在于,它将这种单点式的技能外包规模化、普遍化地扩展到了几乎所有认知领域——从写作、编程到分析、论证,其影响的广度是前所未有的。

其次是高阶执行功能网络的钝化。这正是大模型时代认知外包最核心也最隐蔽的风险所在。如前所述,当AI从“副驾驶”升级为“战略代理”从而接管整个认知 workflow,它直接冲击了位于前额叶皮层的执行功能网络。该网络是大脑的总指挥部,负责目标设定、规划、冲动抑制、认知灵活性以及在多个任务

间切换注意力^[1](P135-168)。这些功能恰恰是前额叶皮层在应对复杂、非结构化的情境中得到高强度锻炼的。当AI为我们规划好研究方案、商业策略甚至个人日程,它实际上代理了前额叶皮层的核心工作。长此以往,这个总指挥部可能因无“战”可指而懈怠。神经科学中的“默认模式网络”(Default Mode Network, DMN)理论为此提供了旁证。DMN是在大脑处于清醒、静息、无特定任务状态时最为活跃的一组脑区,与自我反思、畅想未来和心智游移(Mind-wandering)有关^[18](P440)。而过度依赖外部刺激(如持续的信息流)会抑制其健康活动,可能带来创造力下降和深度思考能力减弱^[19](P352-364)。AI提供的全流程服务,可能将这种对外部刺激的依赖推向极致,使大脑长期处于被动的接收模式(而非主动的建构模式),从而导致执行功能网络的连接强度与灵活性双双下降。

最后是“心智坚毅”的生理基础的瓦解。面对困难时的坚持不懈,不仅是意志力问题,也涉及大脑奖赏回路(如多巴胺系统)的调节。当我们通过自身努力解决难题时,大脑释放多巴胺产生认知奖赏,从而正向强化了应对挑战的行为模式。前扣带皮层(Anterior Cingulate Cortex, ACC)等区域亦在监控冲突、评估努力成本和回报中发挥作用^[20](P389-397)。生成式AI的一键生成模式则提供了一种廉价的、即时的满足感,它绕过了需要付出努力才能获得的内源性奖赏,代之以轻易获得的外源性满足。这种模式类似于一种数字化的超级刺激,极易形成依赖。长期来看,这可能重塑我们大脑的奖赏回路,降低对延迟满足的耐受性,使我们在面对真正的智力挑战时,更容易选择放弃。

(二) 两种叙事之辩：“增强”的边界与“萎缩”的幽灵

至此,我们已有必要正面回应主流的认知增强叙事。这种论调常常披着技术实用主义的合理外衣,其核心辩护逻辑在于:倘若人类千百年来早已习惯将繁杂信息转存至纸质笔记本等物理媒介,那么如今将其平滑移交至云端服务器或AI的数字“脑区”中,似乎仅仅是存储介质的升级,并无任何本质上的哲学危机。追溯起来,该叙事至少有三个相互关联但各有侧重的理论支柱:作为认知卸载的扩展心智,作为系统协作的分布式认知,作为能力涌现的人机共生。

首先是作为其哲学基石的扩展心智理论,其核心在于认知卸载带来的创造力解放。克拉克开创性地论证了心智的边界可以合法地延伸至外部工具^[21],生成式AI将这一理念推向了极致。它不仅仅是外部的记忆存储工具(如笔记本),更是外部的认知处理器。这种认知卸载的价值并不限于降低认知负荷,更在于它策略性地重组了我们的智力劳动——正如程序员将语法记忆和模板代码的编写外包给Copilot,从而能将全部心智资源聚焦于更高级的系统架构设计;科学家利用AI快速筛选和总结海量文献,从而能更快地洞察前沿趋势并提出原创性假设。在这种叙事下,我们失去的只是认知体力劳动的技能,换来的却是从事更高阶创造性思维的自由。

其次是源于认知人类学的分布式认知(Distributed Cognition)理论,它将认知主体从孤立的个体,扩展为异构的人机系统。该理论的先驱埃德温·哈钦斯通过对轮船导航等复杂协作活动的研究指出,认知并非仅仅局限于个体大脑内部。他主张应将认知的分析单元从孤立的个体扩展至一个包含人、工具、符号及环境在内的复杂系统。因此,认知是这一系统内各要素相互作用所产生的系统自身属性(properties of their own),而非纯粹的个体心智能力^[22](Pxiii-xiv)。从这个视角看,生成式AI的融入,并非简单地增强或替代某个个体,而是创造了一个全新的、能力更强的“分布式认知系统”^[23](P174-196)。一个研究团队不再仅仅是几位学者+图书馆+电脑,而是几位学者+一个强大的AI认知伙伴。在这个系统中,AI扮演着信息综合器、假说生成器甚至是智力陪练的角色。从这个角度看,“萎缩假说”的担忧可能犯了“范畴错误”:它仍然聚焦于孤立个体的认知能力变化,而忽略了真正的认知主体已经变成了人机混合系统。

最后是最具前瞻性的人机共生(Human-AI Symbiosis)或认知协同进化(Cognitive Co-evolution)视角,它强调AI不仅是工具,更是催生全新认知能力的催化剂。互联网的开山领袖利克莱德早早便预言了人机之间紧密耦合的、共生的伙伴关系^[24](P4-11)。其后,著名信息哲学家弗洛里迪将我们这个时代定义为继哥白尼、达尔文、弗洛伊德之后人类历史上第四次自我认知革命的时代(前三次革命分别在宇

宙学、生物学和心理学层面打破了人类的绝对中心地位,这次革命则打破了人类作为信息圈中独一无二智能主体的地位)。这次革命的核心是我们不再是信息世界中独一无二的智能主体,而是与其他智能体(Agents)共存^[25](P87-90,94)。这种关系将改变双方,并催生出前所未有的新能力——这不同于使用锤子(工具不会改变),而是类似于驯化动植物(人类和被驯化的物种都发生了深刻变化)。在这个过程中,AI展现出的强大能力不仅正在颠覆我们对人类自身认知和语言学机制的传统理解^[26](P367),也迫使我们进化出新的沟通策略。正如认知科学家所发现的,面对这种新型智能体,我们需要通过设计“特殊且反直觉的提示词(Queries/Prompts)”才能真正探知并激发其深层的认知潜力^[26](P354)。由此,“提示工程”(Prompt Engineering)的兴起,正是一种全新的、旨在与非人类智能进行高效协同的元认知技能。更重要的是,与AI的互动正在重塑我们的思维方式。建筑师通过与图像生成模型对话,可以在几分钟内探索上百种设计可能,极大地拓展了其创意思维的边界;棋手通过与AlphaGo的复盘,学习到了超越人类既有定式的全新战略。在这种叙事下,人类并非在退化,而是在进化——进化出一种与强大AI协同思考、共同创造的新型智慧形态。正如国际象棋大师加里·卡斯帕罗夫在被“深蓝”击败后所预见的,最强的不再是计算机,也不是人类,而是人与计算机协作的“半人马”(Centaur)^[27]。“萎缩假说”担忧的是旧技能的消逝,共生叙事则乐观地指向了新技能、新思维乃至新智慧形态的诞生。

上述三种增强叙事的理论,从不同层面描绘了一幅人机协同、智慧繁荣的乐观图景。它们共同构建了一个强大的论述:认知外包并非零和博弈,而是一次认知能力的重组与跃迁。我们承认这一系列观点的深刻洞见,但必须指出,它们共享着一个潜在的也是致命的理论盲点——它们在不同程度上模糊了“辅助性认知卸载”与“核心执行功能外包”之间的关键区别。这正是增强的逻辑边界之所在,也是萎缩的幽灵开始浮现的地方。

本文的“萎缩假说”并不全盘否定上述理论。我们承认,对于辅助性、模块化的认知功能(如记忆存储、快速计算、信息检索),外包确实能极大地增强我们的能力。但是,当外包的对象从记忆(扩展心智)、信息综合(分布式认知)这些相对外围的功能,转向规划、批判、决策这些位于人类心智中央处理器的核心执行功能时,增强的逻辑便开始失效。这就好比使用外骨骼辅助行走(增强)与常年坐轮椅(导致肌肉萎缩)的差异。虽然两者皆表现为对外部技术工具的借用,但前者是在保有人类主体发力意愿前提下的效能放大,后者则剥夺了肌体自主发力的机会,最终从根本上不可逆地重塑了(甚至是摧毁了)生理机制的底座。

具体而言,扩展心智理论无法解释为何外包的规划过程本身会有损于规划能力;分布式认知理论在庆祝系统总效率提升时,可能忽略了一些事情,即当系统中的人类节点因长期不参与核心推理而丧失独立判断力后,整个系统的脆弱性和被操控的风险;人机共生的乐观愿景,则可能低估了这种共生关系的不对称性——在一个以“便利”和“效率”为最高追求的商业生态中,AI系统更有动机将人类引向认知依赖,而非共同进化。这些都是我们不可忽视的。

(三) 技术演进中的历史回声:从苏格拉底到互联网

大脑因技术而重塑,是贯穿人类文明史的主题。通过回顾几个关键的历史节点,我们可以看到,“萎缩假说”有着深刻的历史连续性。

早在公元前4世纪,柏拉图就在对话录《斐德罗篇》中,借苏格拉底之口表达了对文字发明的深切忧虑,其中,苏格拉底讲述了埃及神祇提乌斯向国王塔穆斯献上文字发明的神话。提乌斯声称文字是“一种能让埃及人更有智慧、记忆力更好的药方”,然而,塔穆斯国王却洞察了文字的双重属性,回应道:“它会在学习者的灵魂中植入遗忘,因为他们会停止使用记忆……他们会信任外在的书写字符,而不再依靠自身由内而外地回忆。所以你发明的不是帮助记忆的药方,而是帮助提醒的药方。你给予弟子的,是智慧的表象,而非真理。”^[28](P69)虽然历史证明,文字最终极大地促进了知识的积累与传播,但苏格拉底的警告在认知层面并非杞人忧天。在依赖口述传统的文化中,人们锤炼出了惊人的记忆能力,能够背诵长

篇的史诗和谱系。文字的出现,确实将这种对内在在大脑记忆容量的极致需求外包了出去,使大规模、高保真的记忆存储成为可能,但也可能在个体层面降低对纯粹记忆力的依赖。

到了中世纪晚期,经院哲学以繁复的逻辑辩论和对经典文本的精细解读而著称。学者的心智由此在严格的形式逻辑训练中得到磨砺。而印刷术的普及,使得书籍数量激增,价格下降,知识不再为少数精英所垄断。这一方面是一次巨大的思想解放,一方面也改变了学者的治学方式。既然获取信息变得轻而易举,那么,需要穷尽心力去背诵、注释和在脑中构建宏大知识体系的治学传统便逐渐式微。研究者指出,通过将知识存储在心智之外,书写与印刷术削弱了传统口传社会中作为知识权威的年长智者(即依靠记忆复述过去的人)的地位,而将优势转移给了倾向于发现新事物的年轻人^[29](P41)。

这种情况在互联网兴起后愈演愈烈。互联网超链接、多任务、碎片化的信息呈现方式鼓励我们进行快速、跳跃式的浏览,而非专注、线性的深度阅读。这在神经层面强化了处理多重刺激的回路,却削弱了支持专注、沉思和批判性分析的回路。尼古拉斯·卡尔在《浅薄》中对此也有细致的论述^[30]。智能手机的普及将这种浅层化效应推向了新的高度,持续的通知和信息流使得长时间不被打扰的专注状态成为一种奢侈品。

(四) 质的突变:生成式AI何以不同

通过上述历史回顾,一个显而易见的问题是:生成式AI所带来的认知外包,与文字、印刷术、互联网相比,究竟是延续性的量变还是断裂性的质变?我们认为,这是一次明确的质变。此质变至少可以体现在三个层面上。

其一,外包的层级与深度。过去的技术,外包的往往是相对单一的认知功能:文字外包了记忆,计算器外包了算术,GPS外包了寻路。而生成式AI首次大规模地指向了整合性的、最高阶的认知能力(即我们所说的执行功能和判断力)。这相当于将大脑的中央处理器外包出去,与仅仅外接硬盘或计算模块存在根本不同。

其二,交互的主动性。传统工具是被动的,无论是翻阅纸质典籍还是敲击计算器键盘,均严格依赖于人类主体明确的意图与指令。然而,当前的大型语言模型早已跨越了被动响应的边界,它们凭借对海量历史数据的隐秘挖掘与对实时情境的敏锐捕获,展现出一种令人警惕的前摄性——常常在人类尚未清晰构想出需求之前,便已越俎代庖地预测意图并打包交付了最终任务。这种无摩擦的交互模式,既降低了人类主动发起认知活动的门槛,也最大化了形成依赖的可能性。

其三,反馈的循环与加速。AI系统通过与用户的交互,不断学习和优化自身。这意味着认知外包的过程是一个自我强化的正反馈循环。用户越是依赖AI,AI就越了解用户的偏好和思维模式,从而能提供更精准、更个性化的服务,这又进一步增强了用户的依赖性。这种个性化的适配能力,是过去的任何技术都不具备的。它将使得脱离AI的认知活动变得不仅困难,而且在情感上和效率上都显得不划算,从而将认知外包从一种选择固化为一种必须。

要之,虽然技术对心智的重塑古已有之,但生成式AI代表了一个断裂性的时刻。它所触发的用进废退机制,直指作为人类智慧核心的执行功能网络,其影响的广度、深度和速度都将是前所未有的。大脑的神经可塑性这把双刃剑,在一个被强大AI“悉心照料”的世界里,正面临着通往心智萎缩的严峻挑战。

三、“心智健身房”的构想与新认知阶层的固化

倘若前文揭示的“萎缩假说”在个体神经层面成立,它又如何以及为何会演变为一个宏观的社会现象?这里,我们必须警惕从个体神经活动到社会结构变迁的简单化约。笔者认为,连接微观认知与宏观社会的关键桥梁,是一个正在形成的、由生成式AI主导的认知生态系统(Cognitive Ecosystem)。这个系统不仅是一系列工具的集合,更是一种重塑人类经验与实践的弥散性技术环境。

(一) 从个体适应到结构性后果:作为规训环境的认知生态系统

正如技术哲学家唐·伊德所揭示的,技术并非中立工具,而是主动重塑人类经验世界的媒介^[31](P49)。生成式AI构建的生态系统,正是一种强大的背景性技术。它无缝嵌入教育、工作、信息获取乃至社会交往的每一个环节,核心设计原则几乎都是认知负荷最小化和摩擦最小化。这种以效率为主臬的设计哲学,本身就是一种深刻的价值选择,它在技术架构中固化了对“便利”的偏好,并系统性地贬低了“过程”的价值。

在这个生态系统中,认知外包不再是个体深思熟虑后的主动选择,而是在环境压力下最自然的默认路径。这源于生态心理学家吉布森提出的“功能可见性”(Affordance)概念——即环境向行动者“提供”或“显现”的可能性^[32](P127)。一个以AI为基础设施的数字环境中,其最显著的功能可见性就是“一键解决”。当写邮件时,AI自动补全的按钮闪烁;当作研究时,“一键总结”的功能唾手可得。这些设计并非强制,但它们通过营造一条最小阻力路径,持续不断地引诱用户放弃更费力的深度认知过程。

这种环境的塑造力量,类似于福柯所描述的现代规训权力(Disciplinary Power)。它并非通过禁令或暴力来运作,而是通过对空间、时间和行为的精细安排,以一种看似赋能和优化的方式,潜移默化地塑造着“驯顺”而“有用”的身体^[33](P156)——在这里,则是“顺从的心智”。因此,普遍性的认知萎缩,其根源并非无数个体“懒惰”选择的总和,而是在一个全新的认知基础设施规训之下,一种结构性、集体性的适应后果。这个基础设施通过设计,系统性地移除了有益的困难,从而使维持核心认知能力变得不经济且反直觉。这正是从微观神经可塑性通往宏观社会趋势的传导机制:一个促进认知外包的环境持续塑造着个体的认知习惯,而无数被塑造的个体行为又反过来巩固了这个环境的正当性与普遍性。

基于对这一结构性后果的诊断,我们才能进一步探讨社会可能作出的反应。历史表明,当一种普遍的生活方式导致某种核心能力的系统性丧失时,补偿性的、仪式化的实践便可能应运而生。工业革命剥夺了日常劳动中的体力需求,催生了现代健身房;同理,当AI革命系统性地移除了日常思维中的认知挑战后,一种我们称之为“心智健身房”的补偿性实践,便成为一种合乎逻辑的社会构想。

(二) “心智健身房”:作为“反动”的认知阻力训练

面对系统性的认知萎缩风险,社会可能的回应路径是多样的,其中无疑包含着通过公共教育体系改革、重塑数字公共空间等积极、平等的集体性方案。笔者之所以聚焦于“心智健身房”这一看似消极的、个人化的补偿性实践,并非因为它“必然”发生,而是因为它最符合当下社会应对系统性风险的默认逻辑——即将应对系统性风险的责任个体化,鼓励人们为“系统性的矛盾”寻求“个体化的生命叙事解决方案”^[34](P38)。在“一切皆可量化、一切皆可消费”的逻辑下,当独立思考能力成为一种稀缺资源时,它被包装成一种高端服务或生活方式产品,便是一个极具可能性的推论。因此,对“心智健身房”的分析,旨在通过对这一症候的解剖,来批判性地审视我们社会应对技术挑战的惯性思维,并揭示其背后通往社会分层的内在机制。

具体来看,我们构想的“心智健身房”,并非特指某个物理场所,而是一种社会实践、教育哲学乃至一个新兴产业的总称。其核心特征可以概括为“刻意的认知阻力”。在一个由AI提供无摩擦、即时性解决方案的世界里,“心智健身房”将通过人为地创造智力挑战、移除技术辅助和强调过程本身,为大脑提供必要的负重训练。其运作形态可能多种多样。

其一,无AI辅助的“第一性原理”工作坊。这些工作坊将要求参与者在完全脱离AI辅助的情况下,仅凭借基础理论和逻辑推理,从零开始解决复杂的工程、商业或科学问题。其价值评判标准,将不再是结果的效率或完美度,而是思维过程的严谨性、原创性和深度。这类似于人们选择在健身房中举起沉重的杠铃,即便工作与生活中已有省力的起重机。

其二,古典文本精读与苏格拉底式辩论俱乐部。这类俱乐部可以复兴一种慢阅读。参与者需逐字逐句地研读柏拉图、康德的著述或《老子》《庄子》等高认知密度的文本,并通过激烈、结构化的面对面辩

论来捍卫或修正自己的观点。刻意营造认知阻力,目的正是在解读复杂句式、追踪精妙论证和应对他人诘问的过程中,锻炼逻辑分析、批判性思维和语言表达能力。这也算是对具有强大的文本总结能力的AI的一种螳臂当车吧。

其三,模拟决策实验室。这些实验室可以利用虚拟现实(VR)和复杂系统模型,构建高度不确定和信息不完整的决策情境(如模拟一场地缘政治危机或一次企业并购谈判)。但关键还是,参与者将被禁止使用AI提供最优策略,必须依靠自己的判断力、团队协作和在压力下保持清晰思维的能力来应对挑战。

其四,计算思维的手工作坊。在一个AI能自动生成复杂代码的时代,这类工作坊将回归编程的本质。参与者可能需要使用最基础的编程语言,甚至纸笔,来设计算法、追踪数据结构的变化,以锻炼一种底层的、不受特定技术栈限制的纯粹逻辑和抽象思维能力。正如Jeannette Wing所定义的,计算思维的精髓在于“概念化而非编程”,它要求人类像计算机科学家一样,通过分解、递归和启发式推理来解决复杂问题^[35](P33-35)。在这种“手工作坊”式的训练中,参与者通过亲身的思考,在脑中构建系统模型,从而掌握人类特有的、具有创造力与想象力的智力工具,以对抗由技术自动化导致的思维惰化。

这些实践的共同点在于,其鲜明的反效率和过程导向特征,是对AI所代表的结果主义和便利至上文化的一种主动“反动”(Reaction)。在“心智健身房”中,挣扎、困惑、试错不再是被技术消除的负面体验,而被重新定义为认知成长的必要“代谢压力”(Metabolic Stress)了。

(三) 从补偿性实践到区隔性资本:新认知阶层的浮现

“心智健身房”的构想,看似为所有人提供了对抗认知萎缩的路径,但现实中,它仍有可能演变为一种昂贵的、区隔性的精英活动,从而成为塑造和固化新认知阶层的关键机制。事实上,只要将其置于当代资本分配与教育筛选的宏观视野下,便不难发现,以下几种相互交织的因素将不可避免地把这种补偿性构想推向阶层特权的深渊。

首先,经济与时间门槛的天然排他性。在AI极大地提升了生产力的社会,普通劳动者的工作可能更趋零散化、自动化,其时间被高度挤占。而“心智健身房”所要求的慢思考和深度沉浸,本身就是一种需要投入大量金钱与闲暇时间的奢侈品。能够负担高昂课程费用、聘请“认知教练”并拥有大块时间进行这种非生产性思维活动的人,大概率将是已占据社会顶层的精英阶层。这与当今高端体育运动(如马术、高尔夫)的社会分层现象并无二致。

其次,文化资本的再生产。借鉴法国社会学家皮埃尔·布尔迪厄的理论,阶级的再生产不仅通过经济资本,更通过文化资本(Cultural Capital)^[36](P78-92)。“心智健身”极有可能成为21世纪后半叶最重要的文化资本形式。能够欣赏并实践无AI辅助下解决复杂问题的“优雅”,将成为新的身份标识。精英教育体系(从私立中小学到顶尖大学)将率先拥抱这一理念,将“认知阻力训练”作为核心课程,从小培养其学生的认知主权意识,从而完成这种新型文化资本的代际传递。

由此,一个基于认知主权的分层结构便开始浮现,并可能比过往的任何分层都更为稳固。结构的一端是“认知贵族”。之所以使用“认知贵族”而非更中性一些的“认知精英”,意在强调这一新兴阶层的三大特质:其一,准世袭性,经由昂贵的“心智健身”和精英教育体系,这种认知优势极易在代际间传递;其二,高度的封闭性,其培养过程所需的时间与经济资本构成了难以逾越的壁垒;其三,权力结构上的支配性。他们不仅拥有独立思考能力,更重要的是,他们是设计、掌控和最终解释AI系统的一方,从而对广大认知依赖者形成了根本性的权力优势。这一阶层具体可分为设计、控制核心AI系统的技术与资本精英,以及通过系统性“心智健身”训练,保持了强大独立认知能力,从而能够驾驭AI、胜任最高阶创造与决策岗位的专业精英。结构的另一端是认知依赖者,主要为占据社会绝大多数的群体。他们的工作与生活高度依赖AI,或许精通于如何“提示”(Prompt)AI以获得答案,但缺乏从第一性原理出发独立思考的能力。他们的信息来源、娱乐消费乃至价值判断,都深受AI算法的塑造。故他们表面上是AI的

用户,实则可能是其算法生态的产品。

(四) 认知鸿沟的社会政治后果:一个“美丽新世界”

这种基于认知能力的根本分层一旦固化,可能对社会结构和政治生态产生深远而危险的影响。它所描绘的未来,并非一个充满压迫的奥威尔式《1984》,而更像一个由算法精心维护的赫胥黎式《美丽新世界》——一个物质繁荣、娱乐至上、多数人安于被引导和满足,从而自愿放弃独立思考的社会。其后果甚至比单纯的经济不平等更为隐蔽和棘手。

具体言之,首要的后果就是社会流动性的终结。在由AI主导的经济体中,向上流动的关键可能不再是掌握特定技能(这些都可被AI替代),而是拥有经过“心智健身”锤炼的、难以自动化的底层认知能力。由于这种能力的培养高度依赖精英教育和家庭背景,阶级固化将变得前所未有的棘手。

其次是创新活力的窒息。人类历史上重大的科学和文化突破,往往源于挑战现有范式、进行非共识性思考的“异端”。然而,被AI的最优解驯化的认知依赖者群体,可能逐渐丧失宝贵的认知多样性。而“认知贵族”作为一个相对封闭、同质化的小圈子,也可能陷入思想的回音室效应,缺乏来自社会底层的充满活力的颠覆性思想的冲击,整个社会的创新引擎或将因此减速,甚至熄火。

此外,民主政治的侵蚀也是一个问题。现代民主理论建立在公民具备理性判断能力的基本假设之上。当大多数认知依赖者在面对由AI生成的、真假难辨的复杂信息和政治宣传时,丧失了独立辨别和批判的能力,民主赖以维系的公共理性基础便会受到严重动摇。政治议程可能更容易被能够娴熟运用AI引导舆论的少数精英所操控。

要言之,“心智健身房”的构想,与其说是一个解决方案,不如说是一个预示着社会深度分裂的症候。它以一种看似积极和励志的方式,为认知不平等提供了“合法性”——将认知能力的差异归因于个人的努力与选择,从而掩盖了其背后深刻的结构性不公。

四、结语

从“认知副驾驶”到“认知君主”的演进,并非遥远的未来,而是一条在大型语言模型驱动下已经开启并不断深化的认知外包路径。这一路径遵循用进废退的神经可塑性原理,正系统性地挑战着人类的核心执行功能。笔者由此推演出的“心智健身房”构想,与其说是对未来认知危机的解决方案,毋宁说更像一个令人警醒的症候——它预示着一个未来:曾经作为人类生存基本技能的独立思考,可能异化为一种昂贵的、带有阶级属性的极限运动。

当驾驭复杂系统、进行原创性思考成为少数人的特权,其后果将远超个体能力的差异。它不仅将固化一种前所未有的、基于认知主权的社會分层,更可能从根本上侵蚀社会创新活力与民主政治的根基,最终导向由算法精心维系的、多数人安于被动满足的“美丽新世界”。故此,我们正处在一个关键的十字路口。解决之道,或许在于如何在一个将认知负荷最小化奉为圭臬的文明中,为“有益的困难”保留一席之地。既然技术致力于抹平一切思维过程的摩擦,那么个体、教育体系乃至社会政策,或许都需要主动地去寻找甚至创造摩擦。

这就要求一种范式性的转变:教育的目标,不应再是教会学生如何更快地从AI获取答案,而是培养他们在没有AI的情况下,如何提出一个好问题,并为之进行艰苦卓绝的智识跋涉的能力。政策的重心,也不应仅是分配AI带来的经济红利,更需思考如何设计一种健康的认知生态学(Cognitive Ecology),保护公民免受无休止算法投喂的侵扰,保障其拥有深度思考所需的时间与空间。

由此,捍卫“认知主权”的斗争,无法寄望于技术精英的道德自觉,而必须成为一项广泛的公共议题和深刻的个体实践。在集体层面,它最终指向一个数字时代更基本的人权议题——认知权(The Right to Think),我们需开启相关的严肃讨论。而在个体层面,选择关闭AI助手,亲自去啃一本难懂的著作,抑或放弃“一键总结”,投身于费心费力的真实辩论,这些看似低效和不合时宜的努力,已不再仅仅关乎个人

的智识精进,还是在认知主权的黄昏中保有主体性的政治姿态。

参考文献

- [1] Adele Diamond. Executive Functions. *Annual Review of Psychology*, 2013, (1).
- [2] Adrian F. Ward. Supernormal: How the Internet Is Changing Our Memories and Our Minds. *Psychological Inquiry*, 2013, (4).
- [3] Dragos A. Bodisavl, Ramona I. Georgescu. Neuroeconomics in the Age of AI: How Machine Learning Alters Human Decision-making at the Neural Level. *Theoretical and Applied Economics*, 2025, (2).
- [4] Andy Clark, David Chalmers. The Extended Mind. *Analysis*, 1998, (1).
- [5] John Sweller. Cognitive Load Theory//Jose P. Mestre, Brian H. Ross. *Psychology of Learning and Motivation: Vol. 55*. New York: Academic Press, 2011.
- [6] John R. Anderson. Acquisition of Cognitive Skill. *Psychological Review*, 1982, (4).
- [7] Robert A. Bjork. Memory and Metamemory Considerations in the Training of Human Beings//J. Metcalfe, A. P. Shimamura. *Metacognition: Knowing about Knowing*. Cambridge, MA: MIT Press, 1994.
- [8] Michal Kosinski. Evaluating Large Language Models in Theory of Mind Tasks. *Proceedings of the National Academy of Sciences*, 2024, (45).
- [9] Jean Lave, Etienne Wenger. *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press, 1991.
- [10] Michael Polanyi. The Tacit Dimension//Laurence Prusak, Eric Matson. *Knowledge in Organisations*. London: Routledge, 2009.
- [11] Lev S. Vygotsky. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press, 1978.
- [12] Linda T. Zagzebski. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press, 1996.
- [13] Donald O. Hebb. *The Organization of Behavior: A Neuropsychological Theory*. New York: John Wiley & Sons, 1949.
- [14] Pasko Rakic, Jean-Pierre Bourgeois, Patricia S. Goldman-Rakic. Synaptic Development of the Cerebral Cortex: Implications for Learning, Memory and Mental Illness//J. van Pelt, M. A. Corner, H. B. M. Uylings, et al.. *Progress in Brain Research: Vol. 102*. Amsterdam: Elsevier, 1994.
- [15] Karin Kucian, Urs Grond, Susanne Rotzer, et al.. Mental Number Line Training in Children with Developmental Dyscalculia. *NeuroImage*, 2011, (3).
- [16] Eleanor A. Maguire, David G. Gadian, Ingrid S. Johnsrude, et al.. Navigation-Related Structural Change in the Hippocampus of Taxi Drivers. *Proceedings of the National Academy of Sciences*, 2000, (8).
- [17] Louisa Dahmani, Véronique D. Bohbot. Habitual Use of GPS Negatively Impacts Spatial Memory during Self-guided Navigation. *Scientific Reports*, 2020, (1).
- [18] Marcus E. Raichle. The Brain's Default Mode Network. *Annual Review of Neuroscience*, 2015, (1).
- [19] Mary Helen Immordino-Yang, Joanna A. Christodoulou, Vanessa Singh. Rest Is Not Idleness: Implications of the Brain's Default Mode for Human Development and Education. *Perspectives on Psychological Science*, 2012, (4).
- [20] Matthew F. S. Rushworth, Timothy E. J. Behrens. Choice, Uncertainty and Value in Prefrontal and Cingulate Cortex. *Nature Neuroscience*, 2008, (4).
- [21] Andy Clark. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York: Oxford University Press, 2008.
- [22] Edwin Hutchins. *Cognition in the Wild*. Cambridge, MA: The MIT Press, 1995.
- [23] James Hollan, Edwin Hutchins, David Kirsh. Distributed Cognition: Toward a New Foundation for Human-computer Interaction Research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2000, (2).
- [24] Joseph C. R. Licklider. Man-computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, 1960, (1).
- [25] Luciano Floridi. *The 4th Revolution: How the Infosphere Is Reshaping Human Reality*. Oxford: Oxford University Press,

- 2014.
- [26] Steven T. Piantadosi. Modern Language Models Refute Chomsky's Approach to Language//Stefanie A. Pillin, Silke-Maria Kleditzsch, Georg H. F. Schmathusen. *From Fieldwork to Linguistic Theory: A Tribute to Dan Everett*. Amsterdam: John Benjamins Publishing Company, 2023.
- [27] Garry Kasparov. *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*. New York: PublicAffairs, 2017.
- [28] Plato. *Phaedrus*. trans. by Robin Waterfield. Oxford: Oxford University Press, 2002.
- [29] Walter J. Ong. *Orality and Literacy*. London: Routledge, 2012.
- [30] 尼古拉斯·卡尔. 浅薄: 互联网如何毒化了我们的大脑. 刘纯毅译. 北京: 中信出版社, 2010.
- [31] Don Ihde. *Technology and the Lifeworld: From Garden to Earth*. Bloomington: Indiana University Press, 1990.
- [32] James J. Gibson. *The Ecological Approach to Visual Perception*. New York: Psychology Press, 1986.
- [33] 米歇尔·福柯. 规训与惩罚: 监狱的诞生. 刘北成、杨远婴译. 北京: 生活·读书·新知三联书店, 2012.
- [34] Zygmunt Bauman. *Liquid Modernity*. Cambridge: Polity Press, 2000.
- [35] Jeannette M. Wing. Computational Thinking. *Communications of the ACM*, 2006, (3).
- [36] Pierre Bourdieu. The Forms of Capital//Mark Granovetter, Richard Swedberg. *The Sociology of Economic Life*. New York: Routledge, 2018.

The Twilight of Cognitive Sovereignty: Cognitive Outsourcing In The Large Model Era and the Conception of "Mental Gymnasium"

Lou Tingjian (Tsinghua University)

Abstract The rise of Generative Artificial Intelligence (AI) is triggering systematic "Cognitive Outsourcing", challenging the prevailing augmented narrative. As AI evolves from a "cognitive copilot" to an active agent, human core cognitive abilities—particularly executive functions—may undergo structural atrophy at the individual level, consistent with the neuroplastic principle of "use it or lose it". Catalyzed by a "cognitive ecosystem" defined by the principle of minimizing cognitive load and wielding powerful disciplinary force, this micro-level risk is rapidly escalating into macro social consequences. In response, the conception of the "Mental Gymnasium" arises as a compensatory practice, aiming to counter cognitive degeneration via "cognitive resistance training". Yet this practice risks becoming exclusive elite activity, thereby creating and entrenching a novel social stratification based on "cognitive sovereignty". This will create a profound divide between the "cognitive aristocrats" who can command AI and the "cognitive dependents" with atrophying mental capacities, posing a long-term threat to social mobility, innovation and even democracy.

Key words generative artificial intelligence; cognitive sovereignty; cognitive degeneration; mental gymnasium; cognitive outsourcing; cognitive stratification

■ 作者简介 楼庭坚,清华大学人文学院助理研究员,北京 100084。

■ 责任编辑 涂文迁