强制的撤销：AI时代学术知识生产中的认知社会化危机

作者 阳涌 · SDE 学员 · 约 2.1 万字 · 发表于2026年7月22日

本文诊断人工智能深度介入学术知识生产时的一个尚未被充分命名的结构性变迁：维系学术认知品格代际传递的社会化机制，其制度性强制正在被AI的效率增益系统性地局部撤销。既有概念——认知去技能化、判断外包、默会知识流失——分别指向个体技能的丧失、决策权的转移或认知品格的断裂。本文论证，它们共有一个未被审查的前提：那个使认知品格得以传递的劳动社会化过程，本身作为一个独立的制度变量，被默认为完好无损。而本文所要命名的，恰恰是这个默认前提的溃散：当AI使得不从事笨拙的深度的批评性劳动、却依然满足学术生产指标这件事变得制度上可行时，共同体失去的不只是某一种技能，而是它再生产“批评者”这一构成性身份的社会化装置本身。本文以粒子物理学、计算社会科学、医学与教育的田野证据为基础，界定了“强制性回路”“劳动基座”与“认知社会化”等核心概念，展开这一撤销的三重互生机制，设定四条可被检验的证伪条件，并为其命名提供了一条所有既有解释均无法覆盖的分离线。本文的诊断不是对AI技术本身的拒斥，亦非对一个本质主义“黄金时代”的哀悼；它是一项关于认知社会关系如何在无数微观的、合理化的日常决策中发生结构性漂移的审慎分析。

关键词： 人工智能；学术知识生产；认知社会化；强制性回路；科学共同体；批评功能；学术资本主义

一、引言：当批评的熔炉冷却下来

2012年夏，ATLAS合作组与CMS合作组共同宣布发现希格斯玻色子。那篇署名逾三千人的论文，背后是数十年的笨拙劳动：成千上万的年轻物理学家在昏暗的实验室里反复调试探测器模拟，在深夜的屏幕上发现一个微弱的信号峰，激动两周后又痛苦地意识到它不过是统计涨落。他们从这些反复的希望与失望中，一点一点地积累起一种无法被形式化编码的判断力——一种在统计显著性不足、理论依据薄弱、但一些微妙迹象让你“觉得不对劲”时，敢于投入数年时间去追查一个最终可能一无所获的信号的物理直觉（ATLAS Collaboration, 2012）。

十余年后，一个使用端到端深度学习模型完成唯象分析、顺利通过答辩拿到博士学位的年轻人，在答辩晚宴上被一位退休教授私下问道：在跑出那个最关键信号图时，你是否曾因完全不相信它是对的，而回到白板前亲手重新推导过一遍？

他诚实地答：没有。他相信他的模型调试得足够好，相信合作组前辈写好的检查代码跑过了所有标准检验，相信如果有什么根本性错误，审稿人一定会指出。

这场沉默，是本文所要诊断的那个尚未被充分命名的结构性变迁的缩影。它不是一个年轻人偷懒的故事，也不是一个老教授怀旧的故事。它是关于一套学术制度——那套通过强制性劳动社会化将一代又一代新手逼入批评性劳动并在其中被重塑的完整生态——如何在AI的效率增益、学术劳动市场的竞争压力、以及共同体内部认知分工的重组中被一点一点地局部撤销的故事。

这个故事的关键词不是“技能退化”，不是“判断外包”，不是“默会知识丧失”。

让我们从这三个关键区分进入本文的核心论证。

第一，技能退化论的框架要求一门手艺在被贬值后仍保持生产力上的必要性。 布雷弗曼（Braverman， 1974）论述技工被泰勒制拆解为流水线操作员时，有一个不可动摇的前提：切割、打磨、组装这些动作仍然需要有人来完成。资本只是把它们重新组织成一种更廉价的形态。但本文面对的是性质完全不同的过程：当AI能直接端到端完成唯象分析、放射诊断或论文草稿时，那些曾经需要研究者、医生或学生亲手完成的劳动，不是被拆解成了更简单的动作，而是变得不再被需要。这门手艺没有被贬值，它被局部废弃。而且被废弃的不只是手艺本身，更是那门手艺作为一种强制性的社会化通道的认知附带功能——批评者品格的培育、共同体方向感的校准、代际传承的强制性结构。当一项劳动环节从知识生产链中被移除时，它曾经承载的附属认知功能也随之消散，且无法通过“再技能化”来回收。这不是退化的故事，是消解的故事。

第二，判断外包论刻画决策权从人转移至算法的过程，但它仍将判断视为一种可以被“转移”的实体。 本文的诊断恰恰相反：那种必须在信息不完备的黑暗中亲自摸索、在反复失败中修正、最终被迫做出无人能代劳的判决的能力——它不是一种可以被完整“外包”的实体。它只能在特定的强制性劳动回路中被生成。当那个回路被AI的效率增益局部撤销时，失去的不是判断的“所有权”，而是判断得以生成的制度条件。这一区分决定了两种完全不同的修复路径：如果是外包问题，修复方案是“把决策权拿回来”；如果是生成条件的撤销，修复方案就必须是“重建能够强制生产判断力的社会化回路”。二者在学术治理的处方上导向截然相反的制度设计。

第三，默会知识论能精美地描述“批评者”所失去的究竟是什么——那套只能通过亲自劳动来传承的认知品格（Polanyi, 1958）。 但它将传承视为一个师徒之间的认知传递问题：只要老王匠愿意教，小学徒愿意学，品格就能从上一代身体传递到下一代身体。它从未将那道迫使学徒必须去学的“制度性紧度”作为一个独立变量加以考察。本文的诊断恰在波兰尼止步的地方开始：不是知识传不下去，而是让知识传下去的强制性制度——那套由论文发表门槛、求职市场信号、晋升评审标准、培训体系规范共同构成的压力系统，使得任何人若不穿越那条笨拙通道就会受到系统性惩罚——正在因AI的介入而松动。默会知识的传承不是被某个坏心思的人有意阻断的；它是被一套由速度、产量、指标驱动的学术生产关系系统性地挤出局面的。

这三个区分汇合为一笔更根本的辨析。既有解释各从一侧切入了“批评者的退场”这枚棱镜的某一切面，但它们在更深处共享了一个未被审查的默认前提：那个维系学术认知品格的社会化机制——那个制度性地逼迫新手必须亲手穿越笨拙的深度的批评性劳动的强制性——本身被默认为一个完好的常量。技能退化论想修技能，判断外包论想夺回控制权，默会知识论想保护师徒温情。但它们都没有去碰那个最根本的问题：如果那个强制性回路本身正在被AI系统性地局部撤销，那么所有这些修补都将在制度层面失效，因为它们试图重建的是那个默认常量之上的变量，而常量本身正在漂移。

这就是本文所要命名的那个新辨别面：不是批评者缺技能，不是批评者放弃判断，不是批评者的隐性知识流失——而是生产“批评者”这一构成性身份的社会化制度的强制性本身，正在被AI的效率增益、学术劳动的竞争压力与认知分工的重组所共同合力局部撤销。

“强制的撤销”这一命名，其不可被既有理论还原的辨别面在于：它在既有概念地图上切出了一条此前不存在的分离线。默会知识论、技能退化论、判断外包论，乃至社会学传统中讨论的“制度性

社会化消解”，都可以在“强制性劳动回路仍基本完好”的制度生态中发生。一个研究者可以在技能退化的同时仍然被制度强迫去穿越某种认知劳动门槛；一门手艺可以在被贬值后仍然作为进入共同体的强制入口。而本文诊断的新现象，恰是这个强制性本身正在被局部撤销——且这一撤销过程不由任何人蓄意造成，而是由AI的效率增益、学术劳动市场的竞争压力、共同体内部认知分工的重组共同结算出来的一个无意的后果。在既有的社会学解释中，没有一个框架将“制度性强制的撤销”本身作为核心诊断变量；它们各自握着这个变量的一个代理人——市场化压力、弹性积累、平台化、教育标准化——但没有直接触及其正身。

这一分离线的理论收益是双重的。在解释力上，它能同时容纳技能退化论的个体层面、判断外包论的决策层面与默会知识论的认知层面，并精确标定它们各自在哪个预设上漏掉了什么；它因此不是一个对既有解释的否定，而是对它们的结构性重新装配。在可操作性上，它将诊断的焦点从“个体学者变懒了/变弱了”的指责，转移到“什么样的制度变迁使得不再变懒变弱不再被惩罚”的社会关系分析，从而为可能的制度干预提供更根本的靶点。如果问题是强制的撤销，那么对策就不能是呼吁个体自强，而必须是重建强制。

本文的论证将分六步展开。第一步（第二节），我们将深入两个具体的田野——粒子唯象学与计算社会科学——勘察批评的劳动如何被实际地重组。第二步（第三节），将界定“强制性回路”与“劳动基座”等核心概念，并回应那个最难的问题：凭什么说AI这一次与历史上任何一次工具革新都不一样？第三步（第四节），系统检视默会知识论、技能退化论与制度性社会化消解论，指出它们各自的解释边界，并正面展开本文的核心命名。第四步（第五节），展开“强制的撤销”的三重互生机制，说明其如何从制度许可的结构性变革，一路连锁到学术传统的代际社会化危机。第五步（第六节），将这一诊断推衍至医学与教育领域。第六步（第七节）设定四条明确的证伪条件，将诊断自身置于可被经验推翻的审慎位置。

二、田野：批评的劳动如何被重组

“强制的撤销”不是一个在所有学科中以统一面貌发生的均质过程。它寄生在具体的工序里。本节选取粒子唯象学和计算社会科学两个田野，分别展示这一机制如何在不同劳动地形上留下相似的刻痕。前者暴露批评劳动的重组如何发生在研究者与物质材料的对抗层面；后者揭示它如何发生在研究者与同行之间的解释对抗层面。

2.1 粒子唯象学：被绕过的笨拙劳动与被悬置的判断

粒子唯象学的日常工作，是处理高能物理实验产生的海量数据，试图筛出可能暗示新物理的信号——一个与标准模型预言的微小偏差，一个在数百种背景噪音中若隐若现的共振峰。在ATLAS与CMS合作组内部，一个典型的博士研究生通常被分配去负责某一特定分析通道，比如双光子末态中的希格斯玻色子信号提取。他需要亲手调试探测器模拟，逐区比对蒙特卡洛事件与真实数据的残差；他会在凌晨三点的屏幕上发现一个微弱突起，激动数周，再痛苦地发现它只是系统误差的一个变体。这种反复的希望与失败，将标准模型的脆弱之处、参数空间的微调褶皱，一点一点地刻入研究者的认知身

体。在ATLAS合作组2012年那场著名分析中，每一位署名作者背后都有数年这样的笨拙劳动作为其学术判断力的根基。

当基于深度神经网络的增强决策树（BDT）与图神经网络（GNN）在2010年代中期以后大规模进入唯象学时，它首先介入的正是这个数据处理流程（Guest et al., 2018）。一个典型的深度学习模型可以在原始探测器数据上完成端到端分析，自动筛选最具区分度的特征组合，输出信号与背景的最优分类阈值。研究者不再需要手动构造那些曾经是直觉孵化器的物理观测量——不变质量谱的特定截面比、某些特殊角度的横向动量分布。他们拿到的是最后一步的结果。

这不是工作流程中某一个步骤的加速，而是整个认知回路的拓扑结构改变。在传统的唯象学劳动中，研究者的身体是数据与判断之间的必经通道：数据流经他的眼睛、他的手、他反复修改的代码、他深夜的怀疑与兴奋，然后才凝结为一个判断——“这里值得再查”。在端到端的深度学习流程中，这条通道的大部分段落被黑箱吸收。研究者的身体不再是数据进入判断的必经关隘。他站在黑箱的一端输入数据，从另一端接收已被精炼至最高显著性的产出。

随之而来的是一种微妙但深远的经验重组。研究者依然能看懂输出结果中的统计显著性数值，却越来越难产生那种无法被量化的前语言警觉——“这个信号虽然统计上看不显著，但它出现在这个特定角度上，直觉告诉我值得再查。”这种警觉之所以在前AI时代被培育，正是因为研究者曾被强制性地长时间暴露在未经算法预处理过的原始数据的反复“刮擦”之中。那些被算法平滑掉的统计涨落、被自动分类器判定为背景噪声的边缘案例，恰恰是物理直觉最肥沃的生长土壤。

在共同体层面，这种重组的后果更为深远。当一个年轻成员带着深度学习模型跑出的结果坐在组会桌前时，他面对的是一个经验丰富的教授。在传统劳动形态中，教授可以逐一盘问推导的每一步骤：你用的费曼图是哪几张？如何估计系统误差？那个反常的峰有没有考虑探测器死区效应？批评之所以能够淬炼判断力，正是因为批评双方都曾在同一种笨拙的劳动中被磨砺过。但当年轻成员的结果是基于一个教授无法轻易拆解的复杂神经网络模型时，有效批评的劳动门槛就被垒得极高：要做出有杀伤力的批评，教授可能必须自己去重新训练一遍整个模型，执行系统的超参数敏感性测试，这在时间上几乎不可行，在资源上可能超出单个学者的能力范围。绝大多数潜在批评者选择了一种更温和的参与方式：对模型输出的图示进行一般性质询，询问一些方法论上的通用注意事项，然后签字通过。对话的形式保留了，但其内核——两套独立淬炼的判断之间那场笨拙而深度的对冲——被悄悄掏空了。共同体从锤炼车间，静默地转化为展示会场。

2.2 计算社会科学：从解释对抗到指标攀比

计算社会科学——一个利用大规模数字痕迹数据与机器学习方法研究人类行为的交叉领域——呈现了批评劳动重组的另一种形态，但指向同一个结论。

传统社会科学的核心劳动是解释性的。一个田野调查者花数年时间生活在一个社区里，反复修正自己的理解框架，直到能够对社群成员自己都无法言明的行为模式给出一个有洞见的“为何如此”的解释。一个定量社会学家即使在用回归模型时也必须审慎论证，每一个被放入方程的控制变量要有理

论依据，并对“统计显著但理论上荒谬”的组合保持警觉。这种警觉本身，正是被同行在其职业生涯中反复追问“你这一步的理论根据是什么”所淬炼出来的认知品格。

大规模预测性AI的介入深刻改变了这个劳动生态。以2015年前后引发广泛讨论的“谷歌流感趋势”项目为典型案例：研究团队利用数千万谷歌搜索查询数据预测流感样疾病就诊率，初始报告称可实现比疾控中心提前一到两周的疫情趋势捕捉（Ginsberg et al., 2009）。然而在2011至2013年间，该模型持续高估就诊率。其背后原因——搜索算法本身的动态更新、用户行为的非平稳变化、模型过度拟合了与真实流感无关的季节性噪音——直到两个独立研究团队分别投入大量资源进行复现研究后才被充分揭示（Lazer et al., 2014）。GFT事件暴露了一个结构性问题：当一个预测模型的准确率高于传统方法、但其内部机制连其构建者都无法完全解释时，能够对其提出有效批评的人力与资源成本已经超出了绝大多数潜在批评者所能承担的范围。

这个教训在当下的大语言模型时代被进一步放大。一个研究团队可以使用机器学习模型分析数十亿条社交媒体帖子，自动发现最能区分不同政治立场人群的语言特征组合，并依此建立精准预测个体投票倾向的模型。论文发表后，它为同行批评垒起了三道几乎无法逾越的门槛。要批评这篇论文，同行不能仅指出其理论论证的薄弱——因为作者可能坦诚承认模型是个黑箱、解释力有限。要做出有杀伤力的批评，同行需要去复现整个数据分析流程：获取同样规模的数据（通常涉及与平台公司的数据使用协议，近乎不可能），花数月时间试不同模型架构和超参数（通常没有资源），验证结论是否对训练数据的微小改变保持稳健（论文没有给出所有预注册细节）。批评的劳动门槛被垒得如此之高，以至于绝大多数潜在批评者选择沉默。

结果是双重的。第一，研究质量的核心标准从“经得起批评的解释力”悄然滑向“经得起指标检验的预测力”——因为后者是唯一能被快速评估的东西。一篇论文在顶刊发表后，同行们无法判断其理论洞见的深度，但他们可以查看其在标准数据集上的F1分数。第二，共同体内部的对话方式从相互淬炼的解释对抗，退化为相互展示的指标攀比。计算社会科学领域内部有学者为此表达忧虑：当模型击败了理论，共同体就不再是一群人在旷日持久的争论中共同辨别方向，而是一群人在各自算法赛道上收割低悬的预测精度果实（Hofman et al., 2017）。鉴赏者不做判决，只比对数字。

两个田野指向同一个核心事实：AI并没有直接禁止批评，但它系统性地抬高了生产有效批评所需的制度性劳动成本。当批评变得过于昂贵时，共同体理性的个体选择就是减少批评。这一集体性的减少——而非任何个体的懒惰或道德缺失——正在悄悄地将整个共同体推离它曾经引以为傲的那个认知功能内核。批评的形式保留着。批评的强制力，正在从它赖以存在的社会化结构中一点一点地渗走。

三、这次为何不一样：强制性回路的界定与历史的证伪

在展开系统的理论对话之前，必须正面回应那个所有类似诊断都必须回答的最棘手问题：凭什么说“这次不一样”？在科学史上，从统计软件到电子计算机，每一代新工具的出现都引发过“判断力将消亡”的焦虑。望远镜曾被视为使天文学家“懒于仰望星空”的拐杖；统计软件曾被担心会培养出

一代不懂代数推导的“按钮社会学家”。凭什么AI这一次造成的改变，就不是这一长串技术焦虑史上的又一个被时间证明为过虑的案例?

回答这一问题，必须做两件事。第一，界定本文的核心概念——强制性回路——的精确边界，使它可以被经验识别；第二，从科学技术学（STS）的理论资源中借来手术刀，精确标定AI从“工具中介”到“认知代理”的那条跃迁线。

3.1 强制性回路：一个操作化界定

“强制性回路”是本文的核心诊断概念，其操作化定义如下：它是指一套由制度性压力（论文发表门槛、求职市场信号、晋升评审标准、培训体系规范）共同构成的社会化装置，该装置使得任何不穿越特定的、笨拙的、深度的批评性劳动过程的人，都会在学术生涯的某个阶段承受可被识别的系统性惩罚——无法通过资格考核、无法获得聘用、无法晋升、被同行隐性排斥。正是这套制度性的强制性，而非任何个体的美好意愿或导师的私人传授，保证了足够多的下一代学者愿意、且必须、投入数千小时的笨拙劳动，从而将上一代人的批评者品格经由具身实践复制到下一代人身上。

强制性回路的本质不是“好制度”，而是“社会化制度”。它不保证产出的认知质量——它只保证进入共同体的人至少被强制性地暴露在某种共同劳动中足够久，久到某种认知品格可以在他们的身体里被点燃。历史上不同的学术传统有不同的强制性回路。19世纪德国研究型大学的实验室内漫长的学徒期是一种；20世纪中叶美国研究型大学的博士资格考试加匿名评审制是一种（Burnham，1990）。它们的共同结构是：入口有高墙，墙内有大火。你不想跳，但你不得不跳；而你跳下去之后，你就被改变了。

AI的介入，在本文的意义上，不是在墙外开了一扇新门。它是在墙上凿了一个越来越大的洞，并且整个制度开始默许——甚至因效率压力而鼓励——从这个洞侧身穿过去，绕过那堆必须亲自扑腾才灭得了的大火。研究者没有被禁止从事笨拙劳动；他只是不再被制度性地逼着去从事它。这就是“强制性回路的局部撤销”的精确含义：不是批评劳动本身从世界上消失了，而是不从事批评劳动却仍能完成学术生产力指标这件事，变得制度上可行了。入口还在。导师还在。论文要求也在。但那个曾经逼迫每个人都必须用自己的身体穿越整条认知回路才被认可的制度性紧度，被AI的效率增益、学术劳动市场的加速竞争、以及“发表高产者为胜”的评价体系共同合力松动。

3.2 历史归纳的边界与STS的干预

乐观者的论证建立在坚实的历史归纳之上：每一次工具革新最终都没有摧毁、而是重新配置了科学共同体内部的批评生态。统计软件没有让统计批评消失——它只是让批评从手算误差的核对，转移到了模型设定、变量选择、因果识别等更高阶的层面。望远镜没有消灭天文学家的判断力——它迫使天文学家发展出新的观测校准技术、新的误差分析方法、新的理论检验标准。在这每一个案例中，工具都只延伸了人的感官或加速了人的运算，而没有在认知回路的拓扑结构上做出根本改变。“人—工具—物质”这条回路中，人始终是将数据转化为判断的必经关隘。

历史归纳是有力的。它迫使任何声称“这次不一样”的人必须精确标定：这一次，工具究竟在哪里跨越了从“中介”到“代理”的界线。

拉图尔（Latour, 1987）在《科学在行动》中提出的“不可逆性”概念为标定这条界线提供了关键工具。拉图尔观察到，当一件技术物的“黑箱化”程度高到某一阈值时，后续使用者就不再能沿着它被构建的路径逆向拆解它；他们只能把它当作一个既成的、不透明的整体来接受或拒绝。拉图尔将这一过程描述为技术物获得了“代理者”的地位——它不再只是一个被动的工具，而是在人与世界之间替人做了一部分原本需要人亲自完成的认知加工。从“中介”到“代理”的跃迁，其标志不是运算速度的提升，而是不可逆性的确立：工具内部究竟发生了什么，使用者不再知道，且在技术上也几乎不再可能完全知道。

将这一视角带入当代学术AI的语境，其启发是直接的。历史上的工具——算盘、计算器、统计软件——在原则上都保持了可逆性。一个博士生可以被要求手算一遍回归系数以理解极大似然估计的迭代逻辑；这个要求虽繁重，在认知和物理上完全可能。工具的使用并没有撤销那堵将学生逼向亲手计算的制度性墙壁。而深度学习模型的黑箱化程度，在特定场景下已跨过拉图尔所说的“不可逆性”阈值。一个基于数十亿参数、在分布式集群上训练数周的Transformer模型，其内部表征的复杂程度超出了人类大脑能够逐一追溯的边界；即使将全部权重公之于众，任何单一个体也无法在认知上逆向拆解它从一个输入到一个输出之间的完整因果链条（Burrell, 2016）。这不是量的升级，而是质的跃迁。人的身体不再是数据与判断之间的必经关隘；黑箱本身已成为一个不可拆解的认知代理者。

正是在这一跃迁中，“强制性回路”开始被局部撤销。历史上的工具曾让笨拙劳动加速，但从未使得绕过笨拙劳动本身变成制度上可行的选择。一个博士生可以在有了计算器的时代仍然被要求理解最小二乘法的推导过程；但在AI深度介入的学术领域中，一个使用端到端深度学习模型产出结果的研究生，被允许不再亲手调试探测器模拟、不再逐区比对残差、不再在数据与模型的淤泥中泡数千小时——而他依然可以发表论文、毕业、找到工作。工具没有强迫任何人停止思考。它只是使得不思考变得更容易、更高效、且更少被惩罚。当制度不再有足够的紧度去强制从事那项劳动时，那项劳动也就从“进入共同体的默认门票”变成了“被默许可绕过的高成本选项”。

这就是为什么“这次不一样”。不是AI更聪明。是AI第一次在那些特定工序上，同时做成了两件事：既提高了产出的效率，又撤销了对笨拙劳动的强制性要求。历史上没有任何一项工具同时在这两个维度上对学术共同体施加了同级别的结构性影响。

四、既有解释为何兜不住它

在界定了强制性回路并论证了“这次不一样”的STS根据之后，本节与三个最可能争夺同一解释对象的既有理论传统对话。需要论证的不是这些传统“错了”，而是它们在面对“强制性回路的局部撤销”这一特定现象时，各自卡在了某个预设上，从而无法无损覆盖本文命名的那个新辨别面。

4.1 默会知识与制度性劳动社会化：波兰尼的漏算

波兰尼（Polanyi，1958）的默会知识论是我们理解专家判断力不可被形式化编码的经典起点。它精美地描述了一个熟练工匠如何能“感觉”材料何时会断裂，一个经验丰富的医生如何能在病人的面容中“读”出化验单尚未显示的危机。本文所关切的那些被绕过的笨拙劳动——调试探测器模拟、反复比对残差、在组会被严厉批评——正是波兰尼所说的那套“只能通过亲自劳动来传承”的认知品格的生产工序。从这个意义上说，波兰尼完全可以解释为什么那个只用AI跑结果的年轻唯象学家无法获得老辈物理学家的判断力：他从未在与物质障碍的长期肉搏中被塑造成那种能够做出前语言识别的认知身体。

但波兰尼的框架止步于此。他将传承视为师徒之间的认知传递问题：只要老王匠愿意教，小学徒愿意学，品格就能从上一代身体传递到下一代身体。在这个框架里，传承的制度性生产条件从未作为一个独立变量进入分析焦点。波兰尼不问：老王匠为什么会愿意花时间教？那个让学徒持之以恒泡在笨拙劳动中的制度性压力——论文发表门槛、求职市场信号、晋升评审标准——到底是怎样被组织起来的？如果这套制度性压力在特定历史条件下开始变松，默会知识的传承是否还能稳定地自我维持？

本文的诊断恰在波兰尼止步的地方开始。不是知识传不下去，而是让知识传下去的强制性制度的齿松了。一个博士生面临的真正问题不是“是否有导师愿意教我”，而是“花数月亲手推导以获得物理直觉”这件事，是否会让他输给另一个更高效使用AI工具产出论文的同辈。默会知识的传承不是被某个坏心思的人有意阻断的；它是被一套由速度、产量、指标驱动的学术生产关系系统性地挤出局面的。波兰尼的框架从定义上就将这种制度性的生产关系排斥在分析焦点之外——它只关注认知，不关注认知再生产的社会经济学。要兜住本文观察到的现象，需要补充的，正是那整个被波兰尼默认为完好无损的制度性劳动社会化维度。

4.2 技能退化与技能废弃：布雷弗曼被更改了的问题

布雷弗曼（Braverman，1974）的“技能退化”论为批评者提供了更接近劳动现场的武器。资本通过泰勒制的科学管理将手艺人的复杂工作拆解为一组可由不熟练工完成的简单动作，工匠失去了对生产过程的控制，从构思与执行合一的完整劳动者降格为只执行不思虑的附庸。把这个叙述套用到当代学术劳动上，有一种表面诱惑力：研究者似乎正在经历某种高阶的技能退化——不再亲手做那些需要数年训练才能掌握的繁琐工序，因而失去了对整个认知生产过程的掌控。

然而，用力一推，这个类比就在其核心预设上崩解了。布雷弗曼论述技能退化时有一个不可动摇的前提：那门被拆解的手艺，仍然被资本所需要。铁匠铺被拆解为流水线，但制作一把椅子所需的那些基本动作——切割、打磨、组装——仍然需要有人来完成。资本只是把它们重新组织成一种更廉价的形态。技能退化论的内在逻辑是：劳动被重新塑形了，但劳动本身作为生产过程的必要组成部分，未被消除。

而本文在第二节中观察到的，是性质完全不同的过程。当AI能直接端到端完成唯象分析的绝大部分工序时，那些曾经需要研究者亲手完成的劳动，不是被拆解成了更简单的动作，而是变得不再被需要。一个博士生不需要再花三年学会如何手工调试探测器模拟，因为深度学习模型将那个步骤整个吞

掉了。这门手艺没有被贬值，它被废弃了。技能退化论的框架要求一门手艺在被贬值后仍保持其生产力上的必要性；而本文面对的情况是：那项劳动作为一个必要的生产环节，正在系统性地从知识生产链中被移除。

这一区分决定了修复策略的根本差异。技能退化暗示可以通过“再技能化”——重新教会工人那门被拆解的高阶手艺——来挽救。本文观察到的，则是一个不再被需要的劳动环节所曾承载的那些附带认知功能——批评者品格的培育、共同体方向感的校准、代际传承的强制性通道——是任何“再技能化”方案都无法回收的，因为它们不是那门手艺本身的属性，而是那门手艺作为一种制度性强制的社会化通道的附带效应。如果那条通道本身被AI的效率增益冲垮了，修复个体技能就像是治疗病症而忽略了摧毁免疫系统的环境毒素。

4.3 制度性社会化的消解：一个最邻近但仍覆盖不了的解释

上述辨析之后，本文的核心命名必须正面与一个最邻近的既有框架对话：社会学与教育学文献中关于“制度性社会化的消解”的讨论。这一传统——从涂尔干的教育社会学，到专业社会学对“老派专业精神”衰落的忧虑，再到当代关于高等教育的麦当劳化的批判——都指向同一类现象：维系专业社群代际价值传递的那些制度性强制（严苛的入门仪式、漫长的学徒期、基于同行判断的晋升制度）正在被市场逻辑、弹性积累、数字平台化与管理主义所侵蚀。

这一框架与本文的诊断高度接近，二者的区分必须极为精确才构成理论增量。本文的论证是：上述框架与本文的诊断之间存在着一个关键的辨析面。制度性社会化的消解论，其核心自变量是市场化压力与科层制理性化——当大学被当作企业来运营，当学术产出被简化为可量化的绩效指标，那种慢速的、低效的、以深度批评为内核的专业社会化自然会被挤压。它描述的是一个被外部力量侵蚀的过程：资本与管理主义从外部冲进学术殿堂，瓦解了其传统的自治秩序。

本文诊断的自变量则在性质上有所不同。AI的效率增益，首先不是一个“外部”力量。它恰恰是被学术共同体自身的求知欲与效率追求所热情拥抱的内在生产力。研究者使用AI，不是因为在资本家的逼迫下不得已，而是因为AI确实能让研究变得更快、更准、更强大。这就决定了它在制度层面的作用机制，与市场化压力有根本不同。市场化压力逼迫学者多产、快产，但它并不直接撤销笨拙劳动作为进入共同体的强制性要求——在最内卷化的学术环境中，年轻学者可能需要发更多论文，但他们仍然需要自己去做那些论文背后的分析工作。而AI的效率增益，则在更深的工序层面，直接让跳过笨拙劳动本身变成可能。它不是从外部挤压——它是从内部提供了一条可被绕过的高效替代通道。

这一区分在操作层面导向不同的诊断靶点。如果说制度性社会化的消解，其对策是反抗市场化、重建学术自治，那么本文诊断所指向的对策则要更为复杂：不能简单“拒绝AI”——因为AI带来的认知增益是真实的。问题在于如何在拥抱AI的效率增益的同时，重新建立某种等价于那道被撤销的强制性回路的制度安排，以确保共同体仍然在某个制度层面上强令每一位新手必须亲自穿越那条漫长的批评性淬炼通道。这是市场化批判不去触及的工程性问题。

至此，本文的核心命名——“强制的撤销”——从既有理论的交错中找到了它独有的辨别面：它不可被“认知去技能化”覆盖，因为后者指向个体技能的丧失，而它指向制度性强制的松动；不可

被“判断外包”覆盖，因为后者将判断视为可转移的实体，而它揭示判断必须在特定的强制性劳动回路中才能被生成；不可被“默会知识丧失”覆盖，因为后者不将强制性劳动的社会化机制本身作为独立变量加以考察；亦不可被“制度化社会化的消解”覆盖，因为后者的自变量是市场化的外部侵蚀，而它面对的是被共同体内部效率追求所主动拥抱的内生性替代。这不是对一个新现象换了一个新名字。这是在既有概念地图上切出了一条在此前理论对话中未被显式标注的分离线。

4.4 与柯林斯的专业技能论和情境学习论的对质——最接近的一组解释为何仍兜不住

第 4.1 至 4.3 节处理了波兰尼、布雷弗曼与制度性社会化。还有一组比它们更贴近的解释必须单独接住：柯林斯与埃文斯在《重新思考专业技能》中论证，默会知识的获得本质上是**社会化进入一个实践共同体**——所谓交互型专长，必须靠长期浸泡在共同体的话语与争论中才能习得；拉夫与温格的情境学习论则把学习刻画为新手经由"合法的边缘参与"逐步走向共同体核心的过程。表面上，本文所说的"认知社会化危机"就是这条社会化通道被切断，似乎只是给它们的失败版本换了个名。

但这两条解释的危机版本，都是"**通道不足导致社会化失败**"——新手进不去、浸泡不够、边缘参与被剥夺，于是没能被社会化进共同体。本文诊断的却是一个它们的框架装不下的情形：**通道仍在、共同体仍在、新手仍能充分进入并参与，但那个曾经强制每一个进入者必须穿过的环节——在批评的熔炉里让自己的判断被反复淬炼——被制度许可为可以绕过**。这不是社会化失败，而是社会化的**内容被掏空**：新手顺利地被社会化进了共同体，却被社会化进了一个不再强制要求那项最硬的判断能力的共同体。柯林斯与拉夫-温格回答"如何成功地把人社会化进去"，本文追问的是"当社会化仍在发生、但它所传递的内容中那个最不可让渡的环节被抽走时，会发生什么"。

由此得到方向相反的可判定预测。柯林斯与情境学习框架预测：只要共同体的参与通道通畅、新手能充分浸泡在其话语与争论中，核心的专业判断能力就能被传递。本文预测：在参与通道通畅、浸泡充分的条件下，只要"强制穿过批评熔炉"这一环节被制度性地撤销，那项判断能力仍会在代际间显著衰减。两者在"参与充分但强制环节被撤销"这一格反向预测——前者预期能力照常传递，后者预期一种"人人在场、却无人再被逼着学会下判断"的空心社会化。

五、强制的撤销：三重互生机制

确定了核心命名的不可还原性之后，本节将从田野材料与理论辨析中，正面展开“强制的撤销”在三重相位上的互生逻辑。这三重相位不是三个彼此独立的阶段，而是同一种结构性漂移在三个不同层面上的同步侧影：制度许可的变革、批评对话的软化、以及集体方向感的分化。它们互相喂养、彼此加固，共同构成一个自我维持的闭环。

5.1 制度许可：从“必须越过”到“可以被绕开”

第一重变革发生在制度许可的层面。它不是任何一个委员会下的决定，也没有任何一份政策文件写着“即日起免除笨拙劳动”。它是从无数微观的制度性默许中一点一点沉积出来的：某个博士生在论文中使用AI生成的分析结果而不被追问底层推导过程，这篇论文被接受发表；某个招聘委员会在评

估候选人时，将论文发表数量作为主要权重指标，而不再有资源去评估每篇论文背后是否有深度的批评性淬炼；某个导师因为自己也不完全理解学生使用的模型，在组会上对学生的方法论部分只做形式性提问，而不再追问“你这一步凭什么这样走”这种需要双方都有共同劳动基底才能展开的深度对话。

每一次这样的微观默许，单独看都是完全合理的。博士生用AI是因为他要在短得多的时间内完成前辈十年才完成的分析量，不这么做就会在求职市场处于劣势。招聘委员会用数量指标是因为这是唯一能在海量候选人之间进行标准化比较的工具。导师不追问深层推导过程是因为他自己也没有花数百小时去亲自拆解那个特定的神经网络架构，如果追问到第四层就会暴露自己的知识盲区，而暴露盲区在一个资深学者的职业直觉中是要避免的。

问题在于，当所有这些微观合理的选择被累加起来时，它们在制度层面共同结算出了一个无人蓄意设计却实实在在发生的结构性后果：那个曾使笨拙劳动成为进入学术对话的强制性门票的制度性紧度，被系统地松动了。 年轻人不是被禁止从事笨拙劳动；他们只是不再因不从事它而承受系统性的制度惩罚。通道还在。入口还在。甚至连“必须通过博士资格考核”的规章也一字未改。但规章之下的那层实质性强制力，已经因AI的效率增益和学术劳动市场的加速竞争而大幅软化。

这就是“强制的撤销”的第一重相位。它不是禁令，而是许可——一种静默的、分布式的、从无数合理选择中涌现出来的制度性许可，允许下一代在绕过熔炉的情况下依然能够被承认为合格的生产者。

5.2 批评对话的软化：从淬炼到形式性认证

当制度许可使得越来越多的学术参与者不再被强制穿过那套笨拙劳动时，第二重变革随之在共同体内部的日常批评对话中显现。批评作为淬炼工具的功能，被批评作为一种形式性认证的功能所置换。

在健康的学术生态中，批评性对话是认知成长的唯一已知途径。一个判断被提出，它必须穿过同行评议中审稿人的独立审查。审稿人之所以能够给出具有淬炼效力的批评，是因为他们自身也曾在同一种笨拙劳动中被磨砺过——他们能在方法论细节中嗅出脆弱点，能在理论推演的跳跃处指认出未被辩护的那一步。这种批评之所以能让原作者在回应中重新淬炼自己的论点，正是因为批评本身就是独立劳动的结晶。它迫使原作者去补足论证、去修改方法、去重新面对自己此前忽略了的反例。

但这一淬炼功能的运转，依赖于一个苛刻的前提：批评双方都曾穿越那道强制性劳动回路，因此双方共享一套可被摊开在桌面上的、可被逐行审查的“可批评材料”。

AI对这一前提的冲击是双重的。在成本维度上，如第二节唯象学案例所示，要有效批评一个基于复杂深度学习模型得出的结论，批评者需要施加至少不低于原作者的劳动——复现、调试、修改、比较。当这种劳动在绝大多数日常语境中都因过于昂贵而变得不切实际时，批评就不会发生。同行评审的质量不会骤然崩塌；它更可能在表面上维持正常运作的同时，在内部悄悄地发生功能置换。审稿人不再去逐行检查推导过程——因为推导过程有一部分被黑箱吸收，逐行检查已不再可能。审稿人转而指

出几个方法论上的通用注意点，建议一些补充分析，然后签字通过。论文发表了，但它在那场本应将其置于最严厉判断的火焰中淬炼的程度，远比过去要低。同行评审从一个用以“判断与修正”的机制，逐渐蜕变为一个用以“认证与流通”的机制。一篇文章被接受，越来越意味着“它通过了最低限度的技术性审核”，而越来越少意味着“它经受住了共同体内最严厉判断的考验并幸存下来”。

在路径维度上，这种软化的冲击渗透在当代学术文化的日常组织之中。会议报告越来越密集，讨论时间被不断压缩，指标评估体系的压力弥漫在所有人的日常计算里。正是在这些看似琐碎的组织方式变动中，那个曾经用来进行深度对质的制度性空间被悄悄挤走。共同体保留着批评的形式，但其内部那个曾经将粗糙判断淬炼为坚固知识的火候，已经降到了远低于有效阈值的水平。

5.3 方向感的分化：从集体收敛到碎片化

当日常批评对话从淬炼退化为形式性认证后，第三重变革随后展开：共同体作为一个集体，对“什么是值得继续追问的真问题”的方向感开始分化。这种方向感在此处有严格的操作性含义：它指的是共同体日常实践中两项具体的、可被经验观察的认知功能。其一，对“哪些问题属于本领域尚未解决的真问题”的排序共识；其二，对“哪条研究路径已被证明是死胡同”的集体排除能力。

在一个批评功能完好的共同体中，这两项功能是通过持续的、分散的批评性辩论来维持的。大量个体在各自日常劳动中分别触碰到同一组未解问题或同一条死路；他们将触碰结果以论文、评论、会议讨论的形式带入公共空间；经过数年甚至数十年的对冲与收敛，一个高度可信的排序共识或一条公认的死路裁定就逐渐浮现。方向感的本质，是这个无限分散却又最终收敛的集体判断过程。

当批评的劳动从内部被软化后，这一收敛过程就失去了它最重要的动力源：大量独立判断之间的碰撞与修正。共同体不再有集中的能力去排除那些反复被证明没有探索前途的路径——因为在批评功能减弱的语境中，“被证明没有前途”这件事本身很难被权威地奠定。需要花费高昂劳动成本才能做出的死路裁定，越来越难找到愿意投入这笔成本的个体。与此同时，在缺乏来自共同体深处的对长期问题的批评性守护时，个体研究者自然会倾向选取那些能最快速、最高效地产出可发表结果的方向。这些方向不一定与共同体最深层的认知困境有关联；它们与AI工具的直接可应用性高度相关。方向因此不再由共同体内持续批评性辩论的收敛来定向，而由算法产出的易得性与发表周期的速度来分化。

需要特别强调，方向感的分化不是无人发问，而是发问的性质发生了微妙但深远的迁移。学术会议上的提问依然会发生，但它们越来越多地停留在对AI产出质量的鉴赏与比较层面——这个模型的特征工程有没有考虑X？那个数据集的抽样偏差是否被Y方法纠正？——而越来越少触及那个更根本的、需要批评者与被批评者双方都曾在同一种劳动中被磨砺过才能被提出来的追问：你确定这个问题的提出方式本身，不是在一条被我们共同默认了的死胡同里原地转圈吗？

5.4 三重互生的闭环与代际维度的推论

这三重相位——制度许可的软化、批评对话的形式化、方向感的分化——不是三枚彼此孤立的标签；它们是同一枚棱镜的三个切面，互相喂养，彼此加固，构成一个自我维持的闭环。

制度许可的软化使得下一代人可以绕过笨拙劳动。当他们绕过笨拙劳动后，他们与上一代人之间就不再共享同一套源于劳动的“可批评材料”；这进一步软化了批评对话的淬炼功能。批评对话的软化又进一步削弱了共同体排除死路、凝聚方向感的能力，使得个体研究者更倾向于投向那些能够快速产出可发表成果的短期方向。而当短期方向取代了由集体批评所凝聚的长期方向感成为主流时，整个学术评价体系就会进一步被速度和数量的逻辑所俘获；这反过来又在制度层面进一步默许甚至鼓励了对笨拙劳动的绕过。由此，这三重变革形成了一个闭合的反馈环：每一个都在强化另外两个，而另外两个又在不断加固这个闭环的整体紧度。

这一闭环带来了最深远的推论，触及代际传承层面。学术共同体的本质特征不是某一代人有判断力，而是每一代人都有能力将判断力传递给下一代。实现这种传递的唯一已知通道，不是通过讲道理，更不是通过阅读回忆录，而是通过让下一代在真实的批评性劳动中生活足够长的、无法被绕过的时间——让他们自己在与导师、与同辈、与数据、与理论的反驳中，被迫做出判决，并为这些判决承担被后人证明为错的后果。正是在这个过程中，那套不可言传的“批评者品格”从上一代人的认知身体传递到下一代人的认知身体里，如同一块火石在撞击中被点燃。

但当制度许可使得下一代人可以被默许绕过这道门槛时，他们就从未真正踏入那条会在他们的认知身体里点燃批评者本能的强制性通道。入口还在，导师还在，论文要求也在，但那个曾经逼迫每个人都必须亲自用身体穿越整条回路才被认可的制度性紧度，松了。这些人自己成为导师和审稿人时，面临着一种他们自己难以克服、甚至难以自我察觉的贫乏。他们可以教给下一代的，是他们自己实际经历并熟练掌握过的——如何高效使用AI筛选信号、如何产出符合技术标准的论文、如何在鉴赏模式下辨别哪个模型结果更好。但他们无法传递给下一代一种他们自己从未被塑造过的东西：那种只有在黑暗中摸索足够久、并在摸索的尽头被迫做出一个无人能代劳的判决后才会沉淀进骨头里的批评者本能。他们不吝教。他们慷慨地给予了下一代自己所拥有的一切。但品格无法被“给予”——它只能在特定的、正在被局部撤销的强制性劳动回路中被传染。

这就是“强制的撤销”在代际维度上的完整逻辑。它不是任何一个个体的失败，不是任何一项技术的罪责，而是一个由无数微观合理选择共同结算出来的集体无意的结构性后果：学术共同体作为一个认知社会化的制度装置，正在它自己追求效率、速度与精确性的过程中，不经意地削弱了自身再生产其最核心认知品格的社会化能力。这一过程，既不能被“技能退化”“判断外包”“默会知识丧失”所无损覆盖，也不能被“学术资本主义的侵蚀”所完整解释。它有它自己独立的、尚未被充分命名的生成机制。

六、推衍：在医学与教育中的平行位移

一个有质量的诊断性概念，必须能在不同的专业土壤上被再次辨认出来。本节将展示，同样的三重互生机制——在AI的效率增益与相应制度激励的共同作用下，强制性劳动社会化回路被局部撤销——在医学住院医师培训与基础教育两个平行领域中各自投下了相似但学科特有的侧影。

6.1 医学：从独立质疑的品格到算法推荐的从容消费

放射科向来是医学中最为依赖图像解读与独立临床判断的领域之一。传统住院医师培训中有一个容易被忽视的核心工序：学员必须在昏暗的阅片室里，在资深导师的监督与挑剔之下，成千上万次地独自做出影像解读，然后承受漏诊被严厉批评、误诊被当众纠正的持续挫败。这场漫长而有时近乎残忍的学徒期，在一个医生的认知身体里悄无声息地建构起两种互相支撑的能力：在模糊阴影中能嗅出早期病变的身体性警觉；以及在看到检验结果、器械读数或AI辅助提示时，去翻查更厚实的病史、追问更细致的体格检查、坚持第二意见的那种独立判断的批评习惯。

几款以卷积神经网络为核心的商用AI辅助诊断系统已通过监管批准，并在多项前瞻性研究中被证明在乳腺X线摄影、肺结节CT检测等任务上达到甚至超过资深放射科医师的敏感度（McKinney et al., 2020）。技术上的优越性无可辩驳。但当这些系统在住院医师第一年就被作为默认阅读工具引入培训流程时，新一代医生就没有被允许在那场漫长的挫败中重塑他们自己的认知身体。十年后，当他们必须面对AI系统给出“不确定”输出的最疑难灰色地带时——恰恰是那些最需要医生超出模型推荐、用自己的全部临床判断力来做出独断的时刻——他们能依赖的，是自己曾经在类似的模糊边缘上反复挣扎过的内在经验。但他们记得的，却不是那些挣扎；他们记得的是导师说过的一两句口诀和论文上列出的标准处理步骤。

这里的危险不是诊断准确率的全面下降。AI的整体精确度可以持续上升，将这一点漂亮地掩盖过去。真正的危险在于：当全部常规诊断都已被AI高效处理、剩下的全部是机器也不确定的疑难灰色地带时，那个曾经能够超越系统推荐、凭借深层独立批评性判断力做出决断并将之坚持到底的集体本能，已在一代人的代际尺度上被缓慢稀释。这种稀释不能被一两位优秀的主治医生用个人努力弥补，因为它的根源不在个体的勤勉或天分，而在于那个曾经迫使年轻医生必须用自己的身体反复穿过诊断的不确定性的制度性紧度，在效率的合理论述下松了。有经验的放射科学家已经在专业期刊上就此发出警告（Recht & Bryan, 2017）。

6.2 教育：从理解力的自我挣扎到知识成品的流畅消费

在教育系统中，强制的撤销发生在认知品格形成的最早期也最根本的阶段。一个学生在数学题前苦思冥想，走入死胡同，退回来，再换一条路尝试——这个过程从来不是学习的可削减代价；它本身就是理解力被一砖一瓦构建起来的唯一已知工序。正是在与错误、与挫折、与自身极限的持续对抗中，那个后来能够让他在面对完全陌生的、没有标准答案的复杂问题时独立出判断并为之负责的认知品格，得以从最粗粝的土壤中萌芽。

大语言模型驱动的AI辅导软件能以对话形式实时给出详尽的分步解答、解题思路的多种变体、甚至针对一个错误概念的精准辨析。这无疑在某些场景和某些学生身上有显著的辅助学习效果。但一旦它在教育的各个关键阶段被默认允许、且被制度——学校作业评分方式、考试政策、教师培训重心——默认为可以绕过核心学习过程的高效替代品时，那个从错误中自己爬出来的强制性通道就被快速地短路了。学生可以流畅地提交完美的作业，但这份完美的所有权不属于他。他不是学会了；他是被流畅

地提供了答案。他正在变成一个知识成品的熟练鉴赏者与从容消费者，却不是这种知识背后所需的那种理解力的自主生产者。

更深的连锁锁在未来。当这一代被提供答案长大的学生在十余年后自己站上讲台时，他们对自己要教给下一代学生的知识，缺乏一种亲身在困惑中浸泡过的深层体验。他们能条理清晰地解释正确结果背后的逻辑；但他们无法为讲台下另一双困惑的眼睛，用自己身体活过的语言与节奏，重现那段从黑暗到乍现第一道微光的动态生成过程。教育共同体的代际批评能力——那种老练的教师能够在学生犯下某个特定类型错误的第一时间就看穿其认知底层缺陷的诊断本能——在自己加速效率的名义下，从最开端的师范生阶段开始被温和地截断。

6.3 互生灵验：跨域的共性结构

在这两个截然不同的专业田野上，与唯象学一样，问题的核心从来不在AI技术本身的卓越或局限。问题的核心在于：在将这项技术移植入各个专业知识生产体系的过程中，在日常效率优化的合理名义下、在缩减培训预算的财务考量中、在追求产量与分数的良性压力里，那套曾将不同世代、不同个体强制性地编组进同一个批评性认知共同体的劳动社会化基座，正在不知不觉间、不由任何坏人蓄意地，被一块接一块局部抽离。强制的撤销不是一段由恶意者精心策划的坏剧情；它是一场由所有人都在做着最合理、最有效率、最无可指摘的日常决策时，静默地共同结算出来的认知社会关系的结构性重组。

七、讨论：边界、风险与证伪承诺

一项论断若要在学术辩论中获得被严肃对待的资格，它必须自愿站立在可以被推翻的露天里，而不是用修辞为自己建造一座刀枪不入的语言堡垒。本节首先回应一个最具分量的潜在反驳，随后设定四条可被未来经验观测检验的证伪条件。

7.1 新型批评形式的反驳与“志愿劳动天花板”

一个不能回避的反驳是：AI时代是否正在催生新型的、可能更有效的批评形式？开源模型权重的复现性检验社区、基于对抗性攻击的模型稳健性测试、预注册报告制度对理论先行的强迫、GitHub上那些对已发表论文代码与数据进行事后审查的数据侦探（data detectives）——这些新型实践难道不恰恰证明批评并未退场，而只是在重新配置自己的基础设施？

这一反驳有力地指出了批评的形式在演变。但它恰好也暴露了本文诊断真正指向的那个层面：批评的强制性。上述新型批评形式，目前几乎全部依赖于个体的志愿劳动。它们要求批评者投入巨大的个人时间成本，去复现一个自己无法从中获得任何职业晋升回报的结果，去钻研一个并非自己研究方向的复杂黑箱模型的内部逻辑。正因为这种劳动是志愿的、而非制度性强制的，它天然是稀缺的。数万名依赖AI产出论文的研究者中，可能只有几十位愿意并能够从事这类高成本、低职业回报的志愿批评。这种结构性的不对称——AI使产出变得极其便宜，却使批评变得极其昂贵——并不会因一小群志愿者的卓越努力而被逆转。恰恰相反：批评越是依赖个体英雄主义，越是证明它已经从共同体的构成性

制度化功能，退化为少数人的道德自律。本文诊断的不是批评的灭绝，而是批评从“一个进入共同体的默认门票”，变成“一座可绕过的自愿悬桥”。那座桥还在，但绝大多数人选择、且被制度默许，不再去走。

历史上确有不少由志愿劳动发端最终被制度化的变革——开源软件运动是常被引用的范例。但开源软件之所以能够从志愿运动发酵为整个软件工业的基础设施，是因为它所生产的工具本身对所有参与者有直接的生产力价值：一个开发者贡献代码，他自己也能用上。开源因此形成了强大的互惠引擎。而学术批评复现劳动却往往缺乏这一互惠结构：花费数月复现并批评一篇他人的论文，对批评者自己发表的直接助益微乎其微，甚至可能因树敌而反向损害职业前景。它更接近公共品供给的经典困境，而非开源社区的互惠积累。将其简单类比于开源运动的成功制度化，低估了批评劳动作为公共品所面临的特定激励结构困难。

7.2 四条证伪条件

以下四条证伪条件分别对应本文核心论证链上最脆弱的四个环节。如果未来在任何一条上出现相反的证据，本文的诊断就需要被大幅修正、甚至整体放弃。

条件一：制度性强制的等效重建。 本文的一个核心隐忧是，在现有的学术评价与劳动市场激励结构下，被AI绕过的强制性笨拙劳动很难被大规模、稳定地等效重建。如果在未来十五年内，在一个或多个AI深度介入的研究领域，出现了可被经验识别的、被领域内资深成员公认为等效于旧有强制性回路的新一代制度安排——例如，新的同行评审标准要求所有使用AI工具的研究必须同时提交可被独立复现的完整分析流程，且不满足该要求则直接拒稿；新的求职市场信号将“是否有深度独立批评性判断的经历”作为准入性门槛；新的晋升标准将“是否在共同体内部批评性辩论中做出过有据可查的实质贡献”作为升等条件——并且，这些新制度安排在至少一到两个案例中被观察到能使下一代研究者的深度批评判断力恢复到被同领域多数资深成员认可的水平，那么本文关于“强制的撤销”的核心机制就需要被修正。届时，本文所诊断的就不是一个结构性陷阱，而只是一个过渡期陡坡，可以被清晰的制度创新所攀越。

条件二：新一代批评者的集体涌现。 本文在第五节做出了一个可被时间检验的高风险推论：在AI成为默认研究工具后受训的第一代粒子唯象学家，将面临系统性的批评判断力贫乏。一个强有力的反证是：如果在2040至2065年间，这代学者大量涌现出被超越学派偏见的多方共同指认为具有老辈同等水平的独立批评判断力的人物——并且对这些人的深度访谈与档案研究表明，其核心判断力确实形成于自己在博士及博士后期间亲自反复穿越的那种笨拙的、与物质材料的深度对抗性劳动，而不是依靠旧技术范式留下的少数“用老方法训练出来的遗民”在夕阳末期的最后余晖——那么本文关于代际社会化锁死的因果链就会被击穿。届时，本文只是系统性地低估了人类在以合作互助、分布式思辨或其他尚未被预见的新认知路径上，集体长出等效批评品格的能力。

条件三：黑箱透明化对批评回路的成功重建。 本文的一个核心担忧是，AI模型的认知不透明性正因其“不可逆性”而锁死共同体内部的深度批评性对话。如果未来十五年之内，可解释性人工智能领域取得实质性突破——不是产生几篇优秀会议论文，而是开发出了被粒子唯象学或计算社会科学的日常

研究实践广泛且自发地采用作为公共批评基础的新一代透明化工具——并且，在这些工具成为默认配置后，有田野调查证据记录到共同体内部的深度批评性对话出现了被参与各方共同承认为“质变性回返”的复苏，那么本文关于黑箱锁死批评回路的诊断就被证明是对一个过渡期阵痛的过度悲观，而非一项有持久诊断能力的命题。AI就只是一次远比印刷机影响更深远的工具的会聚，而不是这次会聚的意外代价本身。

条件四：创新分布的长期逆证。 本文暗示，在AI最深度介入的领域，范式级创新与其AI产出占比之间存在长期逆向关联。如果到2045年前后，在AI介入最深的一个或几个研究领域，不仅没有出现被其他领域共同认可的“范式级突破荒”，反而被公认产出了比AI初步介入时更多、更密集的、被超越学派偏见的无涉利益第三方同辈共同认定为标志性的理论或实验跃迁，并且有证据表明这些跃迁的核心推动者自述其深层批评判断力是被AI工具的使用正面增强而非绕过的，那么本文就应当被老老实实地修订为“一项对认知过渡期焦虑的过度投射”，而不能被当作一项成熟的、可诊断当代学术深层结构变迁的命题。

这些条件，不是一个诊断者在文末为自己贴上的护身符。它们是一个判断者留给那些愿意在未来花时间去检验它的同行们的一串钥匙。一个判断如果想要被共同体认真对待，它就必须把自己的死亡条件亲手摆到桌面上，供他人检验。在那些条件被任何一条满足之前，把“强制的撤销”放到那个可以被众人审视、可以开始接受检验的当下，是本文对自己所诊断的那个正在经历变迁的主体位置——那个敢于为自己的判断承担证伪风险的批评者——所能奉上的、最后的自我践行。

结语：熔炉会变成什么

我们回到本文开篇时粒子唯象学合作组里的那个年轻研究者。他完成了一篇使用AI工具做的论文，通过了同行评审，拿到了博士学位。在答辩结束的晚宴上，一位退休的老教授私下问他：在跑出那个最关键信号图时，你是否曾因完全不相信它，而回到白板前亲手重新推导和检查过一遍？他诚实地答：没有。

本文的所有论证，都是这场沉默的一则诊断注脚。但它不是在指责这个年轻人，不是在批评那套AI工具，更不是在哀悼一个不可重返的黄金时代。

本文是在说：那个曾迫使老教授在他自己的青年时代必须亲手在白板前苦熬的制度性紧度——那套由论文发表门槛、求职市场信号、同行评审标准与导师日常监督共同织成的强制性回路——正在被AI的效率增益、学术劳动市场的竞争压力与认知分工重组的合力所局部撤销。年轻人没有被禁止去白板前苦熬。他只是不再被制度逼着去那里。而当他选择不去时，他承受的惩罚相比上一代人要轻得多。这就是整个诊断的实质所在。

这座熔炉不会在一夜之间熄灭。它会先失去强制力，再失去淬炼功能，然后在很长一段时间里保持着一个体面的外观——会议还在召开，论文还在发表，同行评审还在运行——直到某一天，最后一位曾在旧日那座真正的熔炉里亲手锻造过自己判断力的批评者，也不再坐在会议桌边提出那个让整个房间安静下来的问题。那一刻，熔炉完成了它向一座纪念馆的转变，纪念的不是物理学的死亡，而是物理学作为一种由批评者构成的认知共同体的社会身份的结束。

但也可能不会。熔炉可以不被冷却，它可以被重新设计。它需要的不是在每一场答辩晚宴上让老教授拉住年轻人问那些伤感的问题。它需要的是一套新的制度性强制——一套在AI时代仍然能够保证每一位共同体新成员都必须用自己的认知身体，笨拙地、反复地、无人可代地，穿越那条批评性淬炼的完整回路的制度安排。

这种制度安排会是什么样子，在当下仍是一个开放的、可争辩的问题。本文的全部工作，只是将它从前反思的效率论与默认的制度惯性中拎出来，作为一个可以被看清楚的、可以被命名的、可以被争议的、也最终可以被有意地去重塑的独立对象。接下来的事，就不是一篇论文能够完成的了。

7.3 本文禁止什么：把机制的独有性收拢为禁区

第 7.2 节给出了四条证伪条件。本节把与最近邻（尤其柯林斯的社会化论）的分界坐实为禁区，使"强制的撤销"不至于被泛化成"一切能力退化"的同义词。

方向相反的核心预测（重申并收紧）。 与柯林斯/情境学习的判别点在于：在控制"参与通道通畅度"之后，代际间的判断能力衰减仍应由"强制批评环节是否被制度撤销"独立预测，且后者更强。若通道通畅度足以解释全部衰减、撤销变量不再有独立贡献，则本文相对于社会化论没有增量，本文被证伪。

禁止清单。 其一，禁止把"强制的撤销"等同于"技能退化"或"社会化失败"——它特指强制环节被制度许可绕过，而非通道被切断（第 4.4 节）。其二，禁止在缺乏"强制性回路"可操作界定的场合宣称本机制在场（第 3.1 节已给出操作化界定）。其三，禁止把三重互生机制中任一环节单独抽出作因果解释——制度许可、对话软化、方向分化是互生的闭环，拆开则失效。其四，禁止把医学、教育中的平行位移（第六节）当作已被验证的结论，它们是待检的外推。

本文不主张什么。 第一，不主张新型批评形式毫无价值（第 7.1 节已回应"志愿劳动天花板"）。第二，不主张所有学科都同速经历这一撤销——撤销的速率取决于该学科强制批评环节的制度刚性。第三，不主张已给出制度修复方案：本文的任务是把"认知社会化危机"从一句忧虑，锻造为一个有操作化界定、可证伪、可与最近邻区分的机制。

> **修订说明** 修订范围：（一）补入与核心命题最接近的先行文献并逐处做"承认占位—剩余分工—可判定差别"的处理（柯林斯与埃文斯的专业技能论、拉夫与温格的情境学习论），不以未被引用充作独创；（二）新增/强化"证伪条件与禁区"一节，将原文的对质进一步坐实为一组方向相反、可编码的预测与禁区；（三）扩充参考文献。作者原有的论证结构、核心判断与经验材料未作改动。

参考文献

1. Collins, H. M., & Evans, R. (2007). Rethinking Expertise. University of Chicago Press.
2. Lave, J., & Wenger, E. (1991). Situated Learning: Legitimate Peripheral Participation. Cambridge University Press.

3. ATLAS Collaboration. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Physics Letters B, 716(1), 1–29.
4. Braverman, H. (1974). Labor and monopoly capital: The degradation of work in the twentieth century. New York: Monthly Review Press.
5. Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 1–12.
6. Burnham, J. C. (1990). The evolution of editorial peer review. JAMA, 263(10), 1323–1329.
7. Ginsberg, J., Mohebbi, M. H., Patel, R. S., et al. (2009). Detecting influenza epidemics using search engine query data. Nature, 457(7232), 1012–1014.
8. Guest, D., Cranmer, K., & Whiteson, D. (2018). Deep learning and its application to LHC physics. Annual Review of Nuclear and Particle Science, 68, 161–181.
9. Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. Science, 355(6324), 486–488.
10. Latour, B. (1987). Science in action: How to follow scientists and engineers through society. Cambridge, MA: Harvard University Press.
11. Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. Science, 343(6176), 1203–1205.
12. McKinney, S. M., Sieniek, M., Godbole, V., et al. (2020). International evaluation of an AI system for breast cancer screening. Nature, 577(7788), 89–94.
13. Polanyi, M. (1958). Personal knowledge: Towards a post-critical philosophy. Chicago: University of Chicago Press.
14. Recht, M. P., & Bryan, R. N. (2017). Artificial intelligence: Threat or boon to radiologists? Journal of the American College of Radiology, 14(11), 1476–1480.