认知劳动的隐性拆解：生成式人工智能如何重塑知识制度的担保基础

生成式人工智能如何重塑知识制度的担保基础

陈晓艳 著 · 依王德生《SDE本体论》 · 发表于2026年7月28日 · 创新智商待独立复核

摘 要

生成式人工智能进入知识生产流程后，争论常集中于“模型能否产生新知”或“输出是否可靠”。本文把问题转向知识制度的担保基础：当模型压缩资料搜集、问题界定、论证组织和文本修订等工序时，署名者是否仍能说明关键判断由何而来、经过何种核验，并在受到质疑时重建其推理路径？本文将这种变化概括为“劳动压缩”，并提出由署名担保与责任追溯构成的“制度担保双锁”。进一步地，文章以“制度材料”描述规则被执行者理解和转译时所依赖的质量直觉，以“恐惧结构”描述学术训练可能造成的逻辑校验与惩罚预期的混合。本文不把这些概念当作已经确证的事实，而是构造一条可检验的机制链，并提出低摩擦工序链公开池、担保边界训练和透明性反馈实验三项干预。其核心贡献，是把“是否使用AI”的合规问题推进为“署名者能否对关键判断保持可追溯的认知责任”这一制度设计问题。

关键词：生成式人工智能；知识制度；署名担保；劳动压缩；恐惧结构；制度材料

1 引言

2023年以来，以大型语言模型为代表的生成式人工智能迅速进入学术写作、科研辅助和知识传播。现有争论一方面强调其在检索、重组、反例生成和文字修订上的效率，另一方面担忧依赖模型会削弱判断的独立性。本文不把AI描述为能够自动揭示真理的“AI系统”，而把它视为一种可嵌入不同认知工序、同时具有生成错误、来源不透明和诱发自动化依赖风险的工具（Bender et al., 2021）。

上述争论往往把焦点放在输出是否构成知识，却较少分析：当研究过程中的若干判断工序被快速生成的文本替代或遮蔽时，作者资格所要求的责任能力会发生什么变化。本文提出“劳动压缩”概念，指认知产物形成所需的若干中间工序被显著缩短、外包或不可见化。劳动压缩并不必然降低质量；风险发生在署名者不能辨认关键判断的来源、无法复核所依赖的证据，或不能在质疑出现时重建推理路径。此时受到影响的不是署名形式，而是署名背后的担保能力。

本文围绕三个问题展开：第一，生成式AI介入不同认知工序时，哪些条件会削弱或增强署名者的责任追溯能力？第二，为什么仅要求披露“是否使用AI”可能不足以保护知识制度的担保功能？第三，怎样设计成本可控、能够与现有评审流程比较验证的干预机制？

本文的核心假设是：AI治理的关键不只在规则文本，还在规则的接收条件。本文以“制度材料”指称作者、审稿人和编辑在解释规则时所依赖的质量直觉与风险预期；以“恐惧结构”指称逻辑不确定感与制度惩罚预期可能发生混合的认知—情绪配置。若执行者把过程暴露直接等同于能力不

足，那么“鼓励透明”的规则可能被转译成新的修辞合规。有效干预因而需要同时改变信息环境、责任标记方式和制度反馈，但这一机制仍需实证检验。

本文的研究路径如下：首先进行文献综述，厘清现有研究的三个主要脉络及其盲区；其次构建理论框架，提出“劳动压缩”“制度担保锁”“制度材料”与“恐惧结构”等核心概念及命题；再次说明本文的概念分析方法论；然后分四个核心论点展开分析，并进行反驳预期与回应；最后给出结论，阐明理论贡献、实践含义、研究局限与未来研究纲领。

2 文献综述

2.1 延展认知与分布式认知路径

关于技术如何重塑认知，一条悠久的理论传统来自延展认知假说与分布式认知模型。Clark与Chalmers（1998）在《延展心智》一文中提出，认知过程可以不限于颅骨之内，外部的物理设备——笔记本、计算器、乃至他人的心智——在满足特定条件时可以成为认知系统的组成部分。这一假说挑战了以个体大脑为边界的心智观，为理解人与技术的认知耦合提供了理论基础。Hutchins（1995）在《野外认知》中通过对航海导航团队的研究，进一步提出认知是分布式实现的：知识不在任何一个个体脑中完整存在，而是分布在人、工具与社会协作的网络之中。

这一传统为理解人—技术耦合奠定了基础，但它主要回答认知过程可以由哪些内部和外部组件共同实现。本文追问的是另一个层次：分布式完成的认知产物凭借什么获得作者署名和共同体信任。外部工具可以增强记忆、计算和协作，也可能使部分判断路径不再被署名者掌握。关键差异不在工具是否外部化，而在作者能否核验关键节点、解释选择理由并承担纠错责任。因此，“劳动压缩”的制度后果不能从延展认知理论直接推出，需要把认知分布与作者责任的制度标准连接起来。

2.2 社会认识论与知识信任机制

另一条相关的理论脉络来自社会认识论。Goldman（1999）在《知识的社会世界》中系统考察了知识在群体层面如何被生产、传播与评估，指出个体的认知行为深度嵌入于社会制度——同行评审、学术引用、专家证词——这些制度构成了知识信任的基础设施。Longino（2002）在《知识的命运》中从女性主义社会认识论出发，强调客观性并非来自个体的中立观察，而是来自科学共同体内部的批判性对话与制度化的修正机制。Origgi（2004）则更直接地探讨了信任在知识传播中的角色：在一个认知劳动高度分工的社会中，个体必须依赖他人的证词，而信任正是证词被接受的前提条件。

社会认识论把知识信任置于证词、共同体批评和制度安排之中，为本文提供了重要基础。但AI介入还提出一个更细的问题：署名者是否具备对自己提交内容的责任能力。这里不能把“作者记得每一次措辞修改”等同于可靠担保；现代科研本来就是高度协作和工具化的。更合理的标准是，作者能否识别关键贡献来源、核验核心证据、理解主要推论，并在错误出现时参与修正。本文所说的“担保力”，因此是一组可操作的责任能力，而不是必须独自完成全部劳动。

2.3 批判性AI研究：技术对学术实践的冲击

与本文最直接相关的，是关于模型局限、AI系统功能性声明以及出版责任的研究与规范。Bender等（2021）指出，大型语言模型生成流畅文本并不等于理解，且输出继承训练数据和规模化部署的风险。Raji等（2022）进一步提醒，讨论AI伦理之前必须检验系统是否实现了被宣称的功能。出版伦理规范则明确区分工具与作者：AI工具不能承担作者责任，使用者必须披露相关使用并对准确性、完整性、原创性和引用负责（COPE，2023；ICMJE，2026）。这些来源支持“责任不能转交给模型”，但尚未回答如何测量作者对具体判断节点的实际掌握程度。

现有规范主要依赖使用披露和人类最终负责原则，这是必要起点，却可能把责任处理成一次性声明。本文的推进在于把责任拆解为可检查的过程能力：来源辨认、关键证据核验、推理重建和错误修正。AI并非创造了署名责任问题；它使长期隐含在作者制度中的担保劳动变得更需要显式设计。解决方向也不是排斥AI，而是建立能够区分合理辅助、未经核验的替代以及实质性共同推理的过程记录。

2.4 文献缺口的诊断

综上，延展认知解释认知如何分布，社会认识论解释信任如何由共同体实践维持，批判性AI研究和出版规范说明模型输出不能自动获得作者资格。尚待展开的问题是：AI对中间工序的压缩怎样影响作者的来源辨认、证据核验和推理重建能力；正式披露规则又为何可能在执行中退化为模板化声明。本文以“劳动压缩—制度担保—制度材料—恐惧结构”构成一条待检验的机制链，并据此提出干预与证伪方案。

3 理论框架

3.1 核心概念定义

劳动压缩：指AI使资料搜集、信息过滤、问题界定、论证组织或文字修订等工序显著缩短、外包或不可见化。其关键变量不是速度本身，而是过程可追溯性：署名者是否能够区分模型建议与自身判断，核验关键材料，并说明主要推论在何种证据和选择下形成。劳动压缩既可能释放认知资源，也可能造成责任追溯缺口，效果取决于任务类型、使用方式和核验机制。

制度担保双锁：指知识制度使一个判断获得可问责性的两类机制。第一是署名担保，即署名者承诺认可作品的核心内容并承担公开责任；第二是责任追溯，即受到质疑时能够识别贡献来源、说明关键证据与推论、启动纠错。两者不是要求作者亲历全部技术操作，而是要求其对足以改变结论的关键节点保持理解、核验和修正能力。该概念与现有作者资格规范相衔接，但把“负责”进一步转化为可测试的能力集合。

制度材料：与正式的制度规则相区别，指作者、审稿人、编辑和管理者在解释规则时所依赖的质量直觉、职业风险预期与惯例。制度材料不决定规则条文，却可能影响同一条规则被理解为实质透明、能力不足、风险信号或形式要求。它是解释规则执行差异的中介概念，而不是无法观察的本体实体，可通过情境实验、审稿文本、决策记录和访谈进行测量。

恐惧结构：指学术训练中可能形成的一种认知—情绪配置，即“论证存在需要检查的不确定性”与“暴露不确定性将受到评价惩罚”在主观体验上难以区分。本文不把它等同于临床恐惧，也不预设单一神经传感器。它的经验含义是：当证据质量相同，仅改变公开暴露和评价后果时，研究者的警觉、回避或修订行为是否系统变化。该概念连接质量判断与制度风险，但必须与一般焦虑、声誉管理和理性风险规避区分。

3.2 命题

基于以上概念，本文建立以下核心命题：

P1：劳动压缩对知识制度的影响取决于过程可追溯性，而非AI产出质量单一变量。当署名者无法辨认关键判断来源、核验核心证据或重建主要推论时，署名担保和责任追溯能力将下降。

P2：生成式AI使原本隐含在作者制度中的担保劳动成为需要显式维护的对象；但担保缺口并非AI使用的必然后果，过程记录、独立核验和实质理解可以缓冲该效应。

P3：AI使用披露规则的效力受制度材料调节。若审稿人把过程透明稳定地解释为能力不足，作者将更可能进行最低限度披露或修辞性合规；若透明记录获得中性或正向反馈，披露质量更可能提高。

P4：逻辑不确定感与惩罚预期的混合可能中介制度材料对行为的影响。该中介是否成立，应通过操纵证据质量与评价风险的实验设计加以区分，而不能由概念定义直接推出。

P5：低摩擦过程公开、担保边界训练和制度反馈可能改善责任追溯能力；若这些干预只增加表格负担而不能提高来源辨认、证据核验和推理重建表现，则本文的干预机制不获支持。

3.3 跨学科理论资源

本文的分析框架调用三个远端学科的理论资源，各有其有效性边界。

第一，延展认知假说。延展认知提供了理解“认知过程可以越出个体边界”的基本立场，使本文得以将AI视为嵌入认知劳动加工流中的功能组件，而不是外在的工具。但延展认知的有效性在此处仅止于认知过程层面，本文的推进在于将分析从认知过程延伸到知识的社会担保结构——这是延展认知话语未曾触及的层面。

第二，科学的社会研究传统。默顿（Merton，1942/1973）提出“有组织的怀疑主义”（organized skepticism）作为科学的制度规范，强调科学知识区别于其他认知产物的关键在其可以被共同体持续检验的体制安排。布迪厄（Bourdieu，1975）则揭示科学场域是竞争性的空间，研究者争夺的是“科学资本”——定义什么算好问题的权力。科学的社会研究传统提供了理解知识制度运行逻辑的宏观视野，但它将“个体的署名劳动”视为场域运作的给定条件，未将其作为独立变量解剖。AI的劳动压缩恰恰是在这个“给定条件”层面上动了手术。本文的“担保锁”分析框架，正是旨在打开这个黑箱。

第三，恐惧条件化、再巩固与消退研究为理解威胁预期如何随经验更新提供了有限的类比资源（LeDoux, 2000; Nader et al., 2000; Rescorla & Wagner, 1972）。这些研究主要来自个体层面的实验范式，不能直接证明学术评价规则会产生同构的神经机制，更不能据此推断群体会在某个临界比例后发生“再条件化”。本文只保留一个可检验的行为假设：如果研究者反复观察到透明披露未受惩

罚，并获得稳定、可信的制度反馈，其披露预期和行为可能更新。该假设需要由行为实验和纵向制度研究验证。

4 方法论

4.1 研究设计

本文采用概念分析为主、制度后果推演为辅的研究方法，属于理论性研究的范畴。选题的决定基于以下判断：AI对知识制度的冲击，其最核心的机制——“劳动压缩如何传导至担保结构”——在当前的实证研究中尚缺乏被系统概念化的工具。本文的任务不是提供该机制的因果检验（那是后续实证研究的任务），而是提出一套经得起后续实证检验的概念框架，为设计这样的实证研究提供理论资源。

4.2 分析策略

本文的分析分为四步。第一，拆解署名担保与责任追溯，并把“负责”转化为来源辨认、证据核验、推理重建与错误修正四类能力。第二，比较AI介入问题界定、论证组织和终稿修订时，对上述能力可能产生的不同影响。第三，引入“制度材料”与“恐惧结构”解释透明规则为何可能被执行者重新转译，同时明确一般焦虑、理性风险规避和形式合规等竞争解释。第四，提出三项可做小规模试验的干预，并以过程透明度、责任追溯测试和审稿后果作为结果变量。具体样本量、期刊数量和观察期限应依据功效分析与实际评审周期预注册，不把三至五年等数字设为普遍规律。

4.3 方法局限

本文是一项理论研究，局限主要有三点。第一，“劳动压缩导致担保能力变化”尚未获得因果检验，责任追溯能力的测量工具仍需开发和验证。第二，“恐惧结构”跨越认知心理与制度社会学，目前只能提出行为层面的中介假设，不能以神经科学类比替代直接证据。第三，三项干预可能带来额外负担、策略性填写和隐私风险，群体层面的效果不能由个体消退学习直接外推。上述局限不是附带声明，而是决定本文哪些结论只能作为研究命题。

5 分析与讨论

5.1 担保的双锁：被暴露的隐式结构

本节把署名担保与责任追溯拆解为可观察的制度功能。二者在AI出现以前已经受到协作分工、工具依赖和荣誉署名等问题的挑战；生成式AI的意义在于提高了完整文本先于责任核验形成的概率，使既有缺口更加突出，而不是凭空创造一种全新的责任危机。

署名既分配贡献与声誉，也表示对作品的公开认可和责任承诺（Biagioli，2003）。本文把后一个维度称为“署名担保”：作者应理解作品的核心主张，确认关键材料得到适当核验，能够识别实质性贡献的来源，并同意在错误被发现时参与修正。担保不要求作者独自完成全部劳动，也不能用主观的“我亲历过”替代可核查的责任标准。

责任追溯是署名担保受到质疑后的操作能力。它包括定位支撑结论的关键证据、说明重要方法和推论由谁完成及如何复核、辨认模型介入的高影响节点，并启动更正、撤回或重新分析。责任追溯不要求重现全部心理过程，而要求保留足以判断结论可靠性和分配纠错责任的记录。

在AI普及之前，署名与责任的若干条件往往被默认满足，但并非所有作者都亲历全部工序，协作研究、统计软件和专业分工早已使劳动呈现分布式特征。因此，本文不把“亲自完成全过程”浪漫化为传统常态。真正需要维护的是关键节点的责任覆盖：贡献是否被正确标记，核心证据是否由有能力的人核验，作者是否理解足以影响结论的推论，以及质疑发生时能否定位和修正错误。

生成式AI扩大了这一问题的规模和速度：流畅文本可以在来源、证据和推理尚未被作者掌握时进入成稿，使署名形式与责任能力更容易分离。但劳动压缩也可能把机械性工序让渡给工具，使研究者把注意力用于更高层次的判断。本文因此不宣称AI必然“抽空署名”，而预测风险集中在高影响判断被模型替代、缺乏独立核验且没有过程记录的使用情形。

5.2 劳动压缩如何拆锁：三道工序的微观解剖

为了看清失重如何发生，我们需要推开微观之窗，追踪劳动压缩在认知劳动的三个关键时间节点上如何作用。

第一类压缩发生在问题界定阶段。模型可以迅速生成综述脉络、候选问题和概念边界，从而降低探索成本；与此同时，使用者可能接受一套看似完整却未核验覆盖范围的地图。可检验的风险不是“没有亲自走过全部道路”，而是面对边界案例时无法解释为何排除某些条件，或不能识别模型遗漏的重要文献。后续研究可比较AI辅助组与独立检索组在问题边界解释、反例识别和来源准确性上的表现。

第二类压缩发生在论证组织阶段。模型善于发现局部不连贯、补充过渡和生成反驳，但流畅性可能被误当成证据充分性。模型也并非只能处理既有框架内部的问题；其能力取决于提示、材料和核验协议。因此更准确的风险是：当使用者只要求模型优化现有论证而不进行框架级比较时，局部自洽会提高，而基础假设仍可能未经检验。应分别测量文本流畅度、论证有效性、证据质量和框架替代方案数量，避免把它们合并为“质量”。

第三类压缩发生在终稿修订阶段。多轮人机改写可能降低判断来源的辨认度，但这种影响并非只能依靠记忆判断。版本记录、提示日志、引用管理和贡献标注可以保留来源链。责任追溯的重点也不是逐字判断某句话由谁首先写出，而是识别哪些事实、推论和价值选择由模型引入，作者是否核验并愿意承担修正责任。因而“来源辨认度”应以关键判断节点为单位，而不是以句子所有权为单位。

三类压缩的共同风险，是高影响判断在进入正式评审前已经失去清晰的责任覆盖。制度收到的稿件可能在格式、语言和引用上完整，却无法从成品判断作者是否理解关键推论。由此产生的是一个测量缺口，而不是已经被证明的普遍性担保衰减。本文的实证任务，是检验过程可追溯性是否能够独立预测错误发现、答辩表现、纠错速度和撤稿风险。

若大量作品出现责任覆盖不足，同行评审可能更依赖容易观察的表面特征，如表达流畅度、格式完整度和与既有范式的一致性。但本文不能据此断言知识制度已经“空转”，也不能预设精致度与知识进展无关。更严格的命题是：当表面质量相近时，责任追溯能力较低的作品是否表现出更高的不可

复核错误、更慢的纠错速度或更差的跨情境稳健性。只有这些差异获得重复验证，才能支持担保基础受损的制度判断。

5.3 为什么规则不够：制度材料与对群体质量直觉的预翻译

面对上述风险，期刊与机构通常采用AI使用披露、作者责任声明和保密要求。COPE与ICMJE明确要求人类作者承担准确性、完整性、原创性和引用责任，并说明AI工具不能列为作者（COPE, 2023; ICMJE, 2026）。这些规则具有必要性，但其效果取决于披露是否具体、能否核验，以及执行者如何理解透明信息。本文关注的不是“规则必然无效”，而是规则在何种条件下可能退化为模板化合规。

制度材料不是身体中不可见的实体，而是规则解释的稳定倾向。例如，审稿人可能把过程透明理解为严谨，也可能把同样的信息理解为能力不足；作者可能把披露理解为责任说明，也可能把它压缩成最低限度的免责声明。这些差异可通过配对情境实验、审稿文本编码和决策结果测量，而不能仅凭经验感受推断。

AI的高质量文本反馈可能通过三条路径影响制度材料。第一，成品可见而过程不可见，使研究者高估同行产出的线性和成熟程度。第二，修订速度提高可能增加比较压力，但“速度提高必然造成持续恐惧扫描”尚无证据。第三，如果制度只惩罚错误而不区分诚实纠错与未经核验的自信表达，作者可能倾向于隐藏不确定性。这三条路径都存在竞争解释，必须分别检验，而不能归并为一个确定发生的恐惧机制。

因此，本文的非直觉判断应被表述为条件命题：透明规则的执行结果可能由执行者既有的质量直觉和风险预期重新塑造。若同一份工序记录在不同评价环境中引发系统性不同的质量判断，就支持“制度材料的预翻译”假设；若评价差异完全由记录本身的质量解释，则没有必要引入这一新概念。

5.4 从输入端解耦：三项可检验的制度设计

若规则效果部分取决于执行者接收到的经验信号，干预就应同时改变可见信息、责任能力和制度反馈。以下三项设计不是已经证明有效的政策处方，而是可在小范围、设置对照并评估负担后实施的实验方案。

第一项设计是低摩擦工序链公开池。在获得参与者同意并保护未发表材料的前提下，为研究者提供记录问题变化、失败路径、关键提示和核验结果的轻量空间。其目的不是公开所有草稿，而是降低同行只能看见最终成品造成的信息偏差。结果指标包括参与率、隐私事件、记录真实性、对同行过程的估计偏差以及后续责任追溯表现。若公开池只增加展示性记录或泄密风险而不改善这些指标，应停止或修改。

第二项设计是担保边界训练。研究者在AI辅助后标记关键判断节点：哪些来自原始材料，哪些由模型建议，哪些经过独立核验，哪些仍属未解决的不确定性。训练目标不是追踪每个句子的所有权，而是使责任覆盖与结论影响相匹配。可通过盲测比较训练前后在来源辨认、证据核验、口头答辩和错误修正中的表现；若只提高填写熟练度，则不能视为有效。

第三项设计是透明性反馈实验。期刊可允许作者自愿提交结构化工序附录，并把它与正文审稿适度分离，以减少“披露即自证不足”的风险。本文不建议未经验证就把透明标记纳入编辑优先级或审稿人分配，因为这可能制造新的声誉指标和不公平优势。更稳妥的做法是随机或准实验比较不同反馈方式对披露质量、评审负担、错误发现和后续纠错的影响，再决定是否形成制度激励。

三项设计的共同逻辑，是为执行者提供可核验的过程信息，并检验稳定的非惩罚性反馈能否改变透明行为。本文不把这种变化描述为群体神经系统的消退或再条件化，也不预设存在统一临界点。支持性证据应表现为：在控制个体风险偏好和作品质量后，干预改善责任追溯能力，并且效果能够跨期维持；否则应接受更简单的合规负担或自我选择解释。

5.5 反驳预期与回应

本节主动列出两种可能针对本文框架的强力反驳，并予以回应。

反驳一：过度诊断——“学者写论文一直都在使用工具，从计算器到统计软件，都没摧毁过署名担保，为什么AI就让担保崩塌？这不是对技术的妖魔化吗？”

这一反驳成立了一半。计算器和统计软件同样可能遮蔽关键假设，传统工具并不天然保留完整责任链；AI与它们之间也不是绝对断裂。生成式AI更值得单独研究，是因为它能够同时介入问题界定、论证生成和文字呈现，并以低成本产出高度流畅的完整文本。本文的预测不是“使用AI就掏空署名”，而是：当工具介入高影响判断、作者缺乏独立核验且过程记录不足时，责任追溯表现会下降。若实证比较没有发现这种交互效应，劳动压缩理论就需要收缩。

反驳二：可操作性疑虑——“期刊流程已经极度超载，再要求工序链标注、信用累计，不是把制度复杂度推向崩溃吗？”

这一疑虑构成制度设计的硬约束。工序记录若过长、不可核验或直接影响录用，可能迅速演变为形式主义和新的不平等来源。因此本文只主张自愿、小规模、具备对照条件的试验：限制记录对象为影响结论的关键节点，测量作者与审稿人的时间成本，并设置隐私保护、退出机制和独立抽查。若透明附录不能提高错误识别或责任追溯，却显著增加负担和策略性填写，应停止推广，而不是用更多规则修补失败。

6 结论

6.1 核心发现

本文提出，生成式AI对知识制度的深层影响之一，是它可能使认知工序的压缩速度超过作者责任机制的更新速度。由此形成的风险不是署名自动失效，而是署名形式与来源辨认、证据核验、推理重建和错误修正能力发生分离。正式披露规则是必要条件，却可能因模板化执行而失去辨识力。本文以“制度材料”和“恐惧结构”解释这种执行差异，并提出三项可检验的制度实验。全文结论保持条件性：只有在对照研究中证明过程可追溯性能够独立预测责任表现，这套框架才获得经验支持。

6.2 理论贡献

本文有三项理论贡献。第一，以“劳动压缩”描述AI对认知工序时间、可见性和责任覆盖的共同改变，并将其与一般工具使用区分为可检验的程度差异，而非本体断裂。第二，提出“制度担保双锁”，把作者责任从道德宣言转化为来源辨认、证据核验、推理重建和错误修正四类能力。第三，以“制度材料”分析透明规则在执行端的转译，并以“恐惧结构”提出逻辑不确定感与惩罚预期可能混合的中介假设。四个概念保留了文章的原创性，同时分别配置了竞争解释和放弃条件。

6.3 实践与政策含义

实践上，本文建议先进行低成本试验，而不是立即建立新的强制体系。期刊可试行结构化、可选的工序附录，比较其对错误发现、审稿时间和纠错速度的影响；研究生课程可训练学生标记高影响判断的来源、核验状态和不确定性，并以口头答辩和复核任务评价效果；科研机构可在严格保护未发表材料的前提下研究过程透明度与成果稳健性的关联。参与机构数量、样本量和观察期应由功效分析、评审周期与伦理要求决定，不预设“五至七家期刊、三年实验”为普遍最优方案。

6.4 研究局限

本文完全属于理论建构，至少存在四项局限。第一，劳动压缩与责任追溯之间缺乏直接因果证据，测量工具尚未经过信效度检验。第二，“制度材料”和“恐惧结构”可能与一般风险规避、职业社会化和印象管理重叠，新概念必须证明增量解释力。第三，过程记录可能泄露未发表思想、敏感数据和提示内容，也可能对资源较少的研究者造成更高负担。第四，本文主要以期刊论文为场景，实验室科学、人文阐释和政策研究中的责任分配并不相同，不能直接推广。

6.5 未来研究方向

后续研究可沿五条路径推进：第一，开发并验证责任追溯能力量表与行为任务；第二，用随机或准实验比较不同AI介入程度、核验协议和过程记录对答辩与纠错表现的影响；第三，研究审稿人如何解释相同的透明信息，并检验制度材料是否具有跨机构和跨文化差异；第四，评估工序公开池和结构化附录的真实负担、隐私风险与策略性填写；第五，追踪不同作者资格和评价制度怎样塑造暴露不确定性的预期。所有时间窗和效果阈值均应事前注册并进行敏感性分析。

附论·最近邻切割：制度担保双锁不是延展认知、不是社会认识论、不是作者身份规范

「劳动压缩—制度担保双锁—制度材料—恐惧结构」这条机制链，逐环都紧邻一个成熟理论。链条要成立，得证明它不是这些理论的串联。

第一个邻居是延展认知/分布式认知——认知任务分布到 AI 工具中。本文的分离线：延展认知关心认知能力的构成（思维如何跨工具分布），它对「谁该为分布式认知的产物负责」保持中立；本文的「制度担保双锁」（署名担保 + 责任追溯）恰恰锁定这个被延展认知搁置的问题——当资料搜集、问题界定、论证组织、文本修订这些工序被压缩进 AI，署名者是否还能重建关键判断的来路并为之担

责。延展认知问「认知延展到哪」，双锁问「认知延展之后，责任还锁得住吗」。可裁决：本文预测存在一类产物，认知上高度依赖 AI（延展成立）、但署名者无法重建其关键判断的核验路径（担保失锁）——这类「认知延展而担保失锁」的产物的存在，是延展认知框架不预测、双锁专门捕捉的。

第二个邻居是戈德曼的社会认识论——知识的可靠性由共同体的信任实践维持。切开：社会认识论关心信任如何在共同体中被生产和维持（谁可信、证言如何传递），它的落点是信任的社会机制；本文关心的是署名担保这一具体制度装置在劳动压缩下的失效条件——不是泛泛的信任，而是「署名者能否对关键判断保持可追溯的认知责任」这一可操作的能力。社会认识论说「知识靠共同体信任」，双锁说「这里有一个具体的制度锁（署名=担保），AI 正在悄悄撬开它，而撬开的过程可被追踪」。可裁决线：本文预测署名担保的失效不是均匀的，而集中在被 AI 压缩的特定工序（问题界定、证据核验）上，且失效程度随该工序被 AI 替代的深度而增加。若担保失效与工序无关、均匀分布，则本文对「劳动压缩沿工序差异地侵蚀担保」的刻画错误。

第三个邻居，最贴近的，是出版伦理中的作者身份规范（ICMJE 作者标准、AI 披露规则）。有人会说：这不就是「要求披露是否使用 AI」的老问题吗？本文的分离动作最关键：现行规范停在「是否使用 AI」的二元披露；本文指认这种披露的结构性不足——一份「用了 AI」的声明并不能告诉任何人署名者是否仍能重建关键判断，而一份「没用 AI」的声明也不排除作者本就说不清自己的判断从何而来。真正的问题不是「用没用」，而是「署名者能否对关键判断保持可追溯的认知责任」。本文由此把合规问题（用没用 AI）推进为制度设计问题（如何让署名担保在劳动压缩下仍然锁得住）。可裁决：本文预测，仅要求「是否使用 AI」的披露，与署名者实际的判断可追溯能力不相关——即披露了「使用」的作者未必更不可追溯，披露「未使用」的也未必更可追溯。若二元披露与可追溯能力强相关，则现行规范已足够，本文的「双锁/制度材料/恐惧结构」框架多余。

附论·可裁决预测：担保双锁在哪里松动，就在哪里可测

原稿已构造「劳动压缩—制度担保—制度材料—恐惧结构」的机制链并提出三项干预与证伪方案。此处收束为三条最硬的裁决线。

预测一（工序差异侵蚀）：署名担保的失效集中在被 AI 压缩的特定工序（问题界定、证据核验），且失效程度随该工序被替代的深度增加。若担保失效与工序无关、均匀分布，则「劳动压缩沿工序差异地侵蚀担保」被证伪。

预测二（二元披露的无效性）：仅要求「是否使用 AI」的披露，与署名者实际的判断可追溯能力不相关。若二元披露与可追溯能力强相关，则现行披露规范已足够，本文框架多余。

预测三（恐惧结构的可分离性）：「恐惧结构」预测——学术训练造成的逻辑校验能力与惩罚预期在个体身上混合，使得对 AI 的规避既出于质量考量也出于风险规避。可操作化：分离「因判断质量顾虑而不用 AI」与「因惩罚预期而不用 AI」两类动机，本文预测二者可被区分且在不同制度环境下比例不同。若二者不可分离、始终同步，则「恐惧结构」作为一个独立机制的刻画不成立。

参考文献

[1] Goldman, A. I. (1999). Knowledge in a Social World. Oxford: Oxford University Press.

[2] Clark, A., & Chalmers, D. (1998). The extended mind. Analysis, 58(1), 7-19.

[3] International Committee of Medical Journal Editors (ICMJE). (2023). Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals.

[4] Hutchins, E. (1995). Cognition in the Wild. Cambridge, MA: MIT Press.

关于本文的创新智商标注

本文在原始投稿基础上，由编辑依王德生《SDE本体论》与 SDE 创新智商评估规程做了「最近邻切割」与「可裁决预测」两节加固，意在抬高其不可还原性（I）与差异锐度（D）。依评分规程铁律，任何人不得为自己参与修改的文本认证分数，故本文所涉创新智商为**修改设计目标（134–138 区间）**，非认证分；认证分须由独立评分者盖出处另行给出。