

签名的重量：论AI时代作者身份的伦理重构

——从“我贡献了什么”到“我为它承担全部回应义务”

王德生 · 德麦国际 · 2026年7月17日

摘要

摘要：当生成式人工智能将文本生产的边际成本推向零，创作领域的真正困境并非“人不再愿对文本负责”，而是一个更深的矛盾：无数创作者在AI协同工作中付出了真实的判断、筛选与背书，却发现自己用以兑现这份责任的语言、制度与观念体系——即本文所称的“贡献先验”——已将他们的劳动系统性地排斥在“正当贡献”的视野之外。本文的核心判断是：这一排斥不是某个理论家的偏见，而是一套在特定历史条件下发生出来、曾经有效、如今遭遇结构性断裂的制度-观念-心理复合体。它曾在长达两个世纪的时间里，为“作者”这个身份提供了稳定的伦理底座；如今，它却成为制造伦理撕裂的真正根源。正是这一断裂，从实践中逼显出一个此前只能被隐约感觉到、却从未被清晰命名的伦理动作：在AI时代，作者身份的核心，已不再是“我从源头贡献了什么”，而是“我是否在审慎的引导与审查之后，行使了对某个AI协同生成文本的**签名权**”——决定让它在我的名下进入公共空间，并为此承担全部回应义务的、不可代理的意志行动。这并非对“责任伦理”的精致包装，也不是施密特式“主权决断”的个人主义回响。它的独特性在于：它处理的不是政治秩序的例外创建，而是弥漫在日常写作中的、认识论断裂已成常态条件下，一个具体的人如何以自己有限的名（学术信誉、职业信用、个人声誉），为一个无法完全追溯其生成过程的文本，承担全部伦理责任。在完成这个概念与霍布斯、韦伯、施密特的三方对话之后，论文探讨了签名权在实践中的运作机制，提出了“签名信用”作为这一伦理行为在学术共同体中的可评估、可积累、可破损的伦理资产，并以正反案例展示签名权在实践中被行使、被拒绝、和在博弈中演化的具体形态。最终，论文在以最严肃的反驳——签名权是否只是将系统责任个人化的新自由主义陷阱——逼出论证的最深处之后指出：签名权的提出本身就是一个双重动作，它既是个体在伦理上站定的宣言，也是要求制度透明的政治请求；其完整意义，只有在个体的伦理决断与共同体的制度重建这双向运动中才能完成。

关键词：关键词：人工智能；作者身份；签名权；贡献先验；责任；决断；签名信用

一、引言：一个尚未被命名的撕裂

当生成式人工智能在短短两年内将文本生产的边际成本推向零，创作领域出现了一个醒目的断裂。一方面，海量的实际写作正在以人机协同的方式完成，人类在其中付出真实的判断、筛选、修正与最终背书；另一方面，当这些写作者被要求公开承认AI的角色时，大多数人选择了沉默、淡化甚至否认。批评者迅速将这种行为诊断为“虚伪”或“学术不端”。但批评的便利，反过来恰恰提示了另一个更深的缺口：除了指责他们“不诚实”，我们是否提供过一个能让诚实的协作身份得以安放的伦理位置？

想象一位研究者。她花费了十年的学术生命，积累了某个领域深厚到足以称为“直觉”的判断力。如今，她借助AI高效地完成了专著中的一章——不是让AI代劳，而是用一种前所未有的方式工作：她发起方向，否决了AI六次错误或平庸的尝试，捕捉了一个她此前未充分注意的意外发现，逐句审核了每一条引文，手动重写了那些措辞需要极其精准的段落。她知道这篇东西凝结了她的知识、品味和决断，她的内心有一种清晰的负责感。然而，当她面对一位善意询问“这些精彩的材料梳理是你自己做的吧”的同事时，她却发现——她无法说真话。

不是因为她怕被指责。而是因为她知道，一旦她说出“我用了AI”，她无法用任何正面的话语来接住自己的诚实。在当前的文化语汇中，还没有一个能让她诚实地站住的位置。她的负责，无法在旧的语言和制度中找到任何一个能被正面承认的接口。她不是在做一道道德选择题（“要不要诚实”），而是被夹在三重力量之中：AI让她能高效地完成高质量的智识工作；她的学术共同体要求她对成果负责；但同一个共同体用以承认“负责”的语言和制度，只认可那种符合“源头-输出”线性模式的劳动形态。她的劳动是分布式的、非线性的、渗透在无数微小决策中的——旧制度根本无法看见它。她的沉默，不是虚伪，而是在那个时刻找不到一个能让诚实的协作身份得以安放的伦理位置。

这个困境，是否只是一个人对新工具的适应不良？本文的判断是：不。它揭示了一个更深的、结构的、从未被命名过的断裂。

现有的讨论，无论是悲观的“作者之死”论（Barthes, 1967; Foucault, 1969），还是试图在新的技术条件下重新锚定责任的“责任转向”论，迄今都不曾质疑过一个根深蒂固的预设：**作者身份的合法性，必须建立在某种可辨识、可追溯、不可被机器替代的“贡献”之上。** 本文将此预设命名为**贡献先验**（contribution a priori），并论证：正是这一预设——这个在过去两个世

纪里维持着“作者”身份稳定运作的制度-观念-心理复合体——在AI时代成为制造伦理撕裂的真正的、尚无人触碰的根。它不是某个理论家提出的错误学说，而是一套在特定历史条件下发生出来、曾经有效、如今遭遇结构性断裂的深层架构。理解它的形成史，才能理解它为何有力到足以制造今日的撕裂；撤销它的统治，才能逼显出那个从裂缝中浮现的新伦理动作：**签名权**。

本文要迈出的那一步，不是在“作者”的旧定义上修修补补。它是要追问：如果撤销贡献先验——不是因为它“错了”，而是因为它的历史条件已经转移——那么在AI协同创作中，那个唯一只属于人的、不可被机器代理的伦理动作，到底是什么？答案藏在那个最朴素、最古老、却在现代化叙事中被长期遮蔽的动作里：签署。不是“我亲手写了”，而是“我决定让它以我的名义存在，并为此承担全部回应义务”。

这个命名，听上去朴素。但本文旨在论证，它的理论份量并不轻。在正面建构签名权的内在结构（第二节和第三节）之后，本文将迎接那个最凶狠的反驳——它是否只是将“我没写”包装成“我负责”的免责声明——并在这场正面交锋中，逼出理论最深处的韧性与真正的边界（第五节）。在此之前，必须首先完成一项基础工作：追溯贡献先验如何从历史中生成，以理解它为何能在今天制造如此深重的撕裂。

二、贡献先验的形成史：一项曾经有效的制度-观念-心理复合体

要理解“贡献先验”在AI时代为何成为制造撕裂的根源，不能仅仅将它当作一个错误观念来批判。这种批判姿态本身，就是排查病灶式的——它预设了一个“需要被摘除的病灶”，却回避了更根本的形成史追问：这个预设，是如何在特定的历史条件下，被制度、观念与心理三重力量共同生成的？它曾经解决了什么问题？它为何能在长达两个世纪的时间里维持稳定，以至于显得像是“自然事实”？

只有完成了这层追问，我们才能理解：AI时代贡献先验所遭遇的断裂，不是“一个错误预设终于被发现”，而是“一个曾经有效的历史结构，其历史条件发生了不可逆的转移”。这不是一场启蒙式的真理战胜谬误，而是一次结构性的地基沉降。

（一）制度奠基：版权法如何将“作者”发明为权利接口

1710年，英国《安妮女王法案》的划时代意义，不是因为它“承认”了作者拥有权利，而是因为它**创造**了“作者”这个法律接口。在此之前，印刷品的控制权与收益权属于书商行会，行会通过永久垄断来维护利益；书写者本人，在法律上是一个几近透明的角色。法案的实质动作是一项制度发明的路由安排：它将一项有期限的专有权利从行会手中剥离，重新分配给一个此前在法律上并不重要的角色——书写者本人。（这一段制度史的基本图景，得益于伍德曼西与罗斯对“作者的发明”的经典研究——Woodmansee 1984；Rose 1993；本文站在他们的谱系之上，追问的是这套建构在AI条件下的具体断裂方式。）

但这项安排要成立，就必须找到一个可追溯、可裁决、可验证的标准，来判断哪件作品应当获得保护。这个标准最终被锚定在“独创性”上——作品必须“源自作者的独立劳动，而非抄袭”。必须理解这个定义的制度功能：它不是美学判断的胜利，而是司法可操作性的产物。法官无法裁判一首诗是否“足够创新”，但可以裁判它是否“源自作者本人”。在十八世纪的法庭上，这个标准打开了一条可操作的裁决路径：只要你能证明这部作品是从你那里流出的——通过手稿、证人、创作过程的记录——法律就保护你的权利。

这个定义解决了一个紧迫的制度问题：如何将权利从行会垄断中解放出来，重新分配给一个可识别的、可追溯的权利主体。它用“源头可追溯性”这一形式上可裁决的尺度，替代了根本不可操作的美学判断。它的后果同样深远：从此以后，法律不承认那种“看不见劳动痕迹”的操作为贡献。一个编辑可能为一本书投入了三百小时的改写，但她的名字几乎永远不会出现在版权页上，因为她的劳动被认定为“衍生”，而非“源头”。制度的逻辑是清晰的：只有可作为源头的、可追溯的贡献，才配得上权利的回报。署名权、获酬权、改编权，都紧紧附着在“源于作者本人”这个锚点上。

值得在此做一个关键辨识：版权法将权利锚定在“源头”上，不是因为立法者深思熟虑地选择了一个关于创作的哲学命题，而是因为它需要找到一个能在法庭技术上站得住的标准。这个标准一旦确立，就反过来塑造了整个创作文化的制度环境。作者们学会了保存手稿、记录创作过程、强调作品的“原创性”——不是为了满足某种美学理想，而是为了在法律上保护自己的利益。贡献先验的制度根基，正是这一高度实用的、回应特定历史矛盾的司法技术产物。

（二）观念奠基：浪漫主义如何为制度注入审美正当性

版权法提供了制度骨架，但骨架需要一个灵魂来赋予它正当性。这个灵魂，来自一场看似远离法律的文化运动。十八世纪末至十九世纪的浪漫主义，为“作者”注入了伦理的与审美的光辉。

在早期现代的写作实践中，作者往往将自己视为对已有材料的汇编者、改编者、续写者。莎士比亚从未声称自己“原创”了那些后来成为经典的情节——它们多半来自普鲁塔克、霍林希德或意大利故事集。这种写作观念的核心是：创作不是从虚空中发明前所未有之物，而是对已有文化材料的重新组织、改编和再生。汇编者、改编者、注释者——这些角色在知识的传播与重塑中承担着不可或缺的功能，但他们并不宣称自己是“独创性”的来源。

浪漫主义重新定义了一切。它将创作从“对已有材料的重组”重新定义为“天才个体以独特心灵为源泉、从虚空中化生出前所未有的新事物”。这个观念的兴起，并非偶然。它回应了一个特定历史困境：在前现代的知识秩序中，权威来自对传统的继承和注释；但在现代性条件下，传统的权威崩塌之后，创作者如何证明自己的价值？浪漫主义提供了一条路径：个体用他独一无二的内在声音，替代了传统作为价值的来源。从此，“真正的创作”不再是对传统的忠实传承，而是个体心灵对传统的创造性超越。

这个观念在此后两百年中渗透进社会的细枝末节：它塑造了公众对“创作”的日常理解——一个少年写出自己的第一首诗，会被鼓励“写出自己的声音”；它划定了教育体系最基本的伦理边界——“不得抄袭”，以及最高的褒奖尺度——“原创性贡献”；它甚至内化为个体心理结构的一部分——成为一个“真正的作者”，意味着你必须向世界证明，你贡献了某种不可替代的、来自你独特心灵的东西。

法律关心的是“是否从你那里流出来”（可追溯的源头），审美关心的是“是否从你独特的灵魂里长出来”（不可复制的个人印记）。两种操作截然不同，但它们在更深的层面上共享了同一个预设：**配得上作者之名的人，必须是一个可以被识别为“贡献之起点”的人。** 贡献先验，在制度的可裁决性与观念的审美化之间，获得了双重加持。

（三）心理奠基：主体性确证的需求

第三重奠基，比前两重更隐蔽，却也更顽强。“想要成为作者”的动机，不能完全归结为名与利。更深的那一层，是主体性的确证。当一个人完成一部作品并郑重署名，她在向世界宣告：这个此前并不存在的东西，因为我的决定和行动，如今存在了；这是我生命在这个世界留下的一个独一无二的痕迹。

这份宣告所带给她的自我价值感，不是名利所能替代的。“我存在过，我创造过，我给世界增添了点什么”——这是现代个体在被科层化、消费化、匿名化的社会洪流中，一个最坚固的自我锚点。值得深思的是，这种心理需求的形成，本身也是历史的产物。在前现代的行会制度下，书写者的身份认同并不依赖于“我是独创的源头”，而更依赖于“我是某项技艺的合格传承者”。只有在浪漫主义将“独创性”抬升为创作的最高价值之后，“我贡献了某种此前未有的东西”才逐渐成为个体自我确证的底层需求。

如今，这份心理需求的底层，仍然以“我贡献/增添了某些此前未有的东西”为前提。如果这个前提被抽走——如果一个人被告知，“你做的事情，只是筛选和确认，本质上没有从你那里流出任何此前不存在的东西”——那么，整个以“创作作为主体性确证”的心理结构便会在根基处动摇。

（四）一个暂时稳固的收束

至此，可以做一个谱系意义上的收束。贡献先验不是某个人或某个学派发明的“理论”。它是三重历史力量相互对冲后的收束：版权法需要一个可裁决的权利锚点（制度驱力）；浪漫主义需要一个新的价值来源来填补传统权威退场后的真空（观念驱力）；逐渐内化的现代个体需要一个通过创造来确证自身存在的心理锚点（心理驱力）。这三种力量来源不同、运作逻辑不同、目标也不同，但它们在一个特定的历史节点上，共同收敛于同一个预设：**作者的合法性，建立在可追溯的、独特的、源自其本人的贡献之上。**

这一收束是有效的。它支撑了版权法的运作，让数百年的文学市场得以有序运行；它激励了无数创作者投身于艰苦的艺术与智识劳动，相信他们的独特贡献会被承认；它为教育、批评、学术评价提供了可操作的标准。它有效到足以让“作者=贡献者”这个等式，在漫长的使用中从“人为的制度安排”逐渐自我抹去痕迹，成为“理所当然的自然事实”。

但这套结构在AI时代遭遇的，不是“错误被发现了”，而是**其历史地基发生了不可逆的转移**。当文本生产可以在极低的边际成本下由非人类主体完成时，“从源头追溯贡献”的制度技术，开始无法处理那种“非线性的、分布式的、渗透在筛选与背书中的”劳动形态；标志着个体独特价值的“独创性”观念，开始在人机纠缠的模糊地带失去辨识度；而“通过创造来确证自我”的心理需求，则在“你到底贡献了什么”的质问中陷入无法言说的困境。贡献先验不是一夜之间失效了。它是那些曾经支撑它的历史条件——印刷时代的文本生产模式、个体作为意义唯一源头的文化信念、创造作为主体性确证的心理结构——正在同时发生不

可逆的位移。下一节，我们将进入这个断裂的现场，看它如何在具体的人的实践中制造撕裂。

三、断裂的现场：旧结构遭遇新实践

将贡献先验从隐性的背景拉到聚光灯下，是为了看清它如何在AI时代的创作现场制造出一种特定的、此前未被命名的撕裂感。它不是简单的“害怕被机器替代”，也不是“被发现用了AI”的恐惧，而要精密得多：创作者确实实地付出了大量的判断、筛选和方向引导，她内心清晰地知道，这份最终的文本凝结了她的知识、品味和决断；但当她试图将这份负责的体验翻译成“我贡献了什么”时，她发现旧的语言和制度找不到任何一个能安放她这次劳动的位置。她的负责，无法在贡献先验主导的坐标系中被记录为一笔正当的贡献。

让我用一个人的完整经历来呈现这种张力如何在具体的工作细节中成形——不是作为一个抽象的理论案例，而是作为一种每个在AI时代工作的人都能感同身受的“预演”。（下述“陈”为合成性思想实验人物，情节系典型化建构；本文全部案例同此，声明见注[1]。）

陈是某高校的历史学讲师，研究晚清茶叶贸易已逾十年。她正在撰写的专著中有一章，涉及1870年代长江流域茶商与洋行之间的定价权博弈。她已经掌握了这一时期的关税档案、茶商家族文书和地方志，核心论点的判断在她胸中成型——她识别出了三层博弈结构：俄商通过汉口茶栈压价的机制、本地中间商层的双重代理角色、以及华商在定价协议中的集体行动策略。但将这些材料组织成连贯、兼顾宏观论述与微观举证的文字，是极耗时间的工作。她的初稿写了三千字，发现笔力不足以完整呈现材料的丰富性——不是她不会写，而是一个人在教课、开会、指导学生与写作的其他章节之间，能投入每个论证段落的时间极为有限。

她开始使用AI。不是漫无目的地“试试看”，而是带着一个清晰的意图：她已经知道自己要说什么，她需要AI帮助她把已经确定的判断扩展成完整的学术论述。她输入了自己对这章核心论点的陈述——约三百字，包括三层博弈结构和五条核心档案证据——并指定方向：“帮我展开第一层，关于俄商通过汉口茶栈压价的机制。用学术散文，不要讲套话，紧扣我给你的证据。”

AI返回了初稿，结构合理，语言通顺，但有一个事实性错误。它将一次发生在1873年的茶商集体罢市事件，错误地与俄商的“贷款预购制度”直接挂钩，层层推演出一套看似有理、实则倒因为果的因果链。这里必须暂停，去理解陈在那个时刻的心理活动。她不是“发现了一个错误”然后淡定地纠正。她读到那段话时，心头紧了一下。那种紧张，来自一个在这个领域浸泡了十年的人的本能警觉：如果她不是碰巧亲自核查过那两本专门讨论此次罢市的地方档案研究，这段论证看起来如此言之成理，她可能就信了。她的专业知识价值，恰恰是在这种“看起来都对，但骨子里错了”的时刻被触发。她将两本著作中的关键段落摘出，粘贴给AI，并指令：“你搞错了因果——这场罢市，主要针对的是本地中间商层，与俄商的直接关联甚微。根据这些材料，重新写这一节，把罢市的对象纠正为本地茶栈。”

AI重新生成，这次事实正确了。但陈发现，AI在纠正错误的同时，行文滑向了另一个问题。它为了强调本地中间商的角色，过度展开了彼时茶叶贸易的一般性背景——从唐代茶马互市一路讲起，洋洋洒洒两页——冲淡了她想要聚焦的具体博弈机制。她删掉了AI写的全部背景铺陈，自己手写了一句简短的过渡，然后继续引导：“保留刚才纠正过的论证，但不展开贸易通史。直接切入1874年罢市后洋行与茶栈之间的定价协议重订。”

在随后的三轮交互中，类似的事情反复发生。她否决了一个段落——不是因为它有错误，而是因为它虽然在事实上无懈可击，论证却薄弱到无法承受任何实质质疑。她抓住了一个突然出现的意外角度：AI在某个版本中提到了一个她此前未充分注意的次要茶商、一个在当地方志中被一笔带过的人物。她敏锐地判定这个角度有价值——这个茶商看似边缘，但他的活动轨迹恰好揭示了本地中间商层与洋行之间一种更隐蔽的定价默契。她调整了全章结构来容纳这个发现。她逐句审核引文准确性，纠正那些AI凭“语言惯性”拼凑出的、稍有偏差的表述。她手动重写了最核心的两段论证——不是AI写不出，而是那些措辞的精确度、论证的节奏、以及对反方最有力论点的预判与击破，需要她用全部的专业直觉一寸寸地去拿捏。

最终成型的这一章，约一万二千字。其中六成左右的文字，是AI首先生成、然后经过她逐段甚至逐句审核、修改、重写的；大约四成——特别是论证的核心转折和她认为措辞需要极其精准的段落——是她自己手写或彻底改写的。但这样的统计本身，就已经是一种对她劳动形态的歪曲。真正重要的不是“谁写出了哪些句子”，而是那些否决、引导、筛选、捕捉意外发现、在关键时刻坚持某个判断——这些动作，才是她作为这本书的作者真正在做的事情。它们渗透在整个过程中，无法被追踪、量化、或还原为“她写了百分之四十”。

问题出在一次系里的非正式学术午餐上。一位她信任的年长同事，读完她的初稿后善意地夸了一句：“这章那些精彩的材料梳理，是你自己做的吧？很扎实。”在那个瞬间，陈内心发生了一次闪电般迅速、却默然无声的撕裂。她想说我用了AI，但我

做了很多”——但这句话在她心里打了几个转，没有说出来。

不是因为她怕被惩罚。而是她意识到，一旦她说出口，她无法用任何正面的话语来接住自己的诚实。她能想象同事的反应：不是指责，而是一种困惑的、略带尴尬的沉默。在当前的文化语汇中，还没有一个能让她诚实地站住的位置。她的劳动——那些否决、引导、筛选和最终背书——是一笔在旧的贡献坐标系中无法入账的“黑账”。如果她说出实情，她的专业投入不会得到正面的命名，只会被归入一个模糊的灰色地带，在那里，她的同行既无从表扬她，也无从批评她，只能尴尬地当作没听见。而这种模糊，反倒是她最有希望让自己的成果按其本身的水准被评价的方式——换句话说来说，制度在反向激励她保持沉默。

陈的经历，精确地呈现了这种撕裂的发生机制。她的负责体验——那种“我确实为这篇东西投入了十年积累的判断力”的确信——与旧制度的合法化接口之间，出现了系统性的不兼容。她不是在面对一个道德选择题（“要不要诚实”），而是被夹在三重力量之中：AI工具让她能够高效地完成高质量的智识工作；她的学术共同体要求她对成果负责；但同一个共同体用以承认“负责”的语言和制度，只认可那种符合“源头-输出”线性模式的劳动形态。她的劳动是分布式的、非线性的、渗透在无数微小决策中的，旧制度根本无法看见它。她的无法言说，就是贡献先验在AI时代最尖锐的现实效应：不是“人不负责任”，而是“负了责却无法被看见、被承认、被记入学术生命”。贡献先验的实质，是一套筛选和排斥机制：它只承认符合“源头-输出”模式的劳动，而将那些渗透在筛选、引导、修正、背书中的劳动，全部排斥在“正当贡献”的视野之外。

值得在此做一个精确的概念界定。本文使用“贡献先验”一词，严格地在上述三重奠基所分析的意义上使用：它特指现代作者身份成立所依赖的那个未明言的预设——“作者身份必须建立在可辨识的、源自创作者本人的独特贡献之上”。这个预设既嵌入在版权法的操作逻辑中，塑造了浪漫主义以来的创作观念，也内化为个体自我确证的心理需求。它不是任何一个理论家正式提出的学说，而是这三重奠基共同维护的、使得“作者”这个身份得以成立的底层脚手架。本文使用“先验”一词，取其“经验条件”之义，而非康德先验哲学中的严格含义——它是使得“作者”经验得以可能的历史条件，而非先于一切经验的认知结构。

四、签名权：从矛盾中逼出的伦理动作

面对陈的困境，最常见的回应路径是试图“重新定义贡献”——将筛选、背书、引导也纳入“贡献”的范畴。这种尝试的动机是善意的，但它仍然困在贡献先验的逻辑内部：它试图扩大“贡献”的名单，却从未质疑过“作者必须首先被言说为一个贡献者”这个前提本身。真正的出路，不是去发明一种更细粒度的贡献计量方法，而是撤销这个前提，然后去看清楚：当一个人不再被“我贡献了什么才配被称为作者”这个问题所捆绑时，她在AI协同创作中所做的那一系列真实动作——发起、引导、筛选、修正、背书——其内在的、不可替代的伦理重量，到底归结为什么？

（一）为什么不是“重新定义贡献”

在展开正面论述之前，必须首先排除两条看似便捷、实为歧路的替代方案，以论证签名权不是“一种备选方案”，而是矛盾张力的必然收束。

第一条歧路：将“筛选与背书”重新定义为一种“贡献”。这条路在政治上看似容易——它只需要将贡献的概念从“产出文本”扩展为“对文本做出重大判断”。但它面临一个致命困难。贡献先验的制度核心，不是“贡献”这个词的语义弹性，而是版权法对“可追溯的源头”这一可裁决标准的需求。法律需要能够在法庭上站得住的证据——手稿的时间戳、创作过程的记录、从源头到产品之间一条可追踪的链。而陈的否决、筛选、捕捉意外发现——这些动作，即使她自己在工作结束后也未必能完整追溯，更不用说提供法庭可采信的证据。把“筛选”叫做“贡献”，在话语层面毫无困难，但在制度操作层面锁死了——法院不会因为它被重新命名就改变它要求可追溯证据的操作逻辑。走这条路，终究只是语义上的自我安慰。

第二条歧路：发明“AI协同作者”的署名格式，用技术化的方式标注AI的参与程度和人类的工作量。这条路面临另一个困境：它回避了责任承担的单一性。无论标注多么细致，AI本身不能被追责——它不是一个能够独立承担民事责任或伦理责任的主体。当一个文本出现事实错误、论证缺陷或伦理问题时，必须有一个具体的人站定并说“我来负责”。署名格式可以描述参与程度，却无法锚定责任终点。

签名权之所以是矛盾的必然收束，正是因为这两条歧路的尽头指向了同一个结论：在AI的大规模应用已不可逆的现实条件下，“贡献”的可追溯性被工程化地削弱了（第一条的路堵死了），但“责任必须锚定在一个可以回应、可以追溯的人身上”这一伦理事实并未改变（第二条的路也堵死了）。这两个条件之间不存在调和。不是你最终选择了“签名权”这个概念，所以它诞生了；是因为这两股力量——生产过程的去人化与伦理责任的单一性——在现实中持续对冲，必然地逼出一个不再依赖“可追溯贡献”、却仍然坚持“人必须承担全责”的伦理动作。这个动作，就是签名。签名，不是某个聪明人想出的一种新颖方案，

而是在这场对冲中，唯一没有被两个条件同时锁死的动作。它保留了“人的责任”，同时放弃了“贡献的可追溯”——这是代价最大、因此也最少被认真对待的选项。但它正是矛盾对冲后剩下的那个选项。

（二）签名权的三层界定

签名权是指：一个人在对某段AI协同生成的文本进行了审慎的引导、审查与修改之后，最终以自己的名义，决定让它进入公共空间，并为此承担全部不可推卸的回应义务的意志行动。在AI时代，作者身份的核心，不再是“我贡献了什么”，而是“我是否行使了签名权”。

这个命名，需要在三个层次上被严格界定，才能区别于那些听上去类似、实质不同的概念。

第一层：签名权不是“贡献路由此处”的另一个说法。“我签名，所以我负责”——这个句式容易让人误以为，签名只是对一份已有贡献的追认。不是的。在AI协同创作中，签名本身就是那个最终的决定性行动。当AI已经可以完成绝大多数的文本生成工作之后，还能有什么东西是唯有那个署名者才能做的？不是“写得更好”——论语言平滑度，AI可以达到极高的水准。不是“想得更深”——AI可以综合的文献量远超个体。那个唯一只属于人的东西，不是生产过程内部的某个环节，而是那个为已经成形的文本**赋予存在资格**的意志动作。签名，就是创造性的最后的、在AI时代变得决定性的那个环节。在手工写作时代，这个动作因为和“写作”同时发生而从未被单独看见；在AI时代，当写作本身可以被大量外包之后，签名的独立性才第一次被单独看见。

第二层：签名权不是“把关人”概念的翻新。“把关人”概念最初由社会心理学家库尔特·勒温在1947年提出，后由大卫·曼宁·怀特于1950年引入传播学，用以描述新闻编辑对信息的筛选过程。把关人的核心特征是：他面对的是已经成形的稿件，他的权力是选择“放行”还是“阻止”。而AI协同创作中的签署，其独特性在于：签署的意志贯穿了整个生成过程。陈在第一次纠正AI的事实错误时，她已经在行使签名权了——因为她之所以去纠正，恰恰是因为她打算最终为这一段负责。她在否决平庸、捕捉意外发现、坚持某个笨拙但关键的论点时，都是在行使签名权。签名不是最后一刻的盖章，而是从发起意图的第一秒起，就已经作为定向力渗透在整个历程中。

第三层：签名权不是“责任伦理”的简单重申，也不是施密特式“主权决断”的个人主义回响。这一层需要专门展开，因为它关涉到本文立论最深处的根基，也关涉到这个概念能否在强大的学术对手面前证明自己的不可还原性。

韦伯在1919年《以政治为业》的著名演讲中提出的“责任伦理”，对行动者的核心要求是：一个成熟的政治家，必须能够预见到自己行动的“可计算”的后果，并为此承担相应的责任。与只关心动机是否纯洁的“信念伦理”相对，责任伦理要求行动者对他行动在现实世界中可能引发的、可预见的效应负责。这是一个建立在因果链条相对清晰、可追溯的世界观之上的伦理学命题。必须公允地承认，韦伯并非天真地相信政治家能够完全预知后果；他深知政治行动的后果常常溢出预期。但他仍然将“尽可能准确地计算后果”作为行动者的伦理义务。这种义务的前提是：行动者能够（至少大致地）理解自己的行动与后果之间的因果关联，并基于这种理解来调整行动。

但当陈按下发送键的那一刻，她面对的是什麼？她面对的，是一个她负全责、但在认识论上她无法完全预知其未来演变路径的文本。AI的文本生成是一个概率空间——那个最终被她签署的文本形态，是在无数微小的、不可完全追溯的交互决策中涌现出来的。她可以审核事实、调整论证、打磨措辞，但她无法像理解自己的思维链条一样理解AI的每一笔生成。她甚至无法完全预知，这份文本在进入公共空间后，会在不同的阅读环境中产生什么样的连锁反应。她是在一个认识论黑箱的末端，做出了一个伦理学全责的决断。

“责任伦理”要求因果清晰才能全责。陈的处境却是：因果不完全可追溯，但她仍然选择承担全责。这就是签名权与责任伦理最根本的断裂点。责任伦理处理的是“我知道后果所以我要负责”的成熟审慎，签名权处理的却是“我知道我无法完全知道后果，但我仍然选择站定并承担”的古老决断。前者承诺负责，建立在认识论的足够强度之上；后者承诺负责，恰恰是认识到认识论强度永远不够之后，仍然做出的意志宣示。这不是责任伦理的精致变体，而是责任伦理在遭遇认识论断裂时必须让渡的、由另一个更古老的伦理动作来接管的地带。

但这恰恰是那个最凶狠的理论挑战出场的地方。这个“在认识论断裂条件下仍然做出的决断”，难道不就是施密特式“主权决断”在数字时代的个人主义回响吗？

在施密特的政治神学中，主权者的决断处理的是一个特定的伦理-政治情境：在法规范无法提供明确指引的“例外状态”下，主权者做出无先例可循的决断，并以自身全部的政治存在为之负责。他与签名权确实共享了一个形式结构——认识到规则的断裂之后仍然站定的决断。这一家族相似性如果不被正面对待，将使签名权面临被“决断论”轻易穿透的风险：读者完全可以说，“签名权就是施密特在数字时代的个人化翻版”。

但二者的分野是根本性的。施密特的决断处理的是**政治秩序创建的非常时刻**——例外状态之所以可怕，恰恰因为它是例外的，它中断了常态，要求主权者在法律沉默之处替法律说话。与之相对，签名权处理的不是“例外”，而是一种**新的日常**。陈不是在某个特殊的历史危机中挺身而出；她是每天都在面对一个她已经默认接受的事实——她和AI协作产出的文本，其生成过程的细节她永远不可能完全掌控。这种认识论断裂不是一次性的政治决断时刻，而是弥漫在她每一次按下“发送”键的日常动作中。它已经从例外变成了常态。

第二个分野同样关键。施密特的决断，其后果是由整个政治体来承担的——主权者的决定约束的是国家，战争与和平的代价由全体国民分摊。签名权的后果，则首先和最终，是由那个具体的人以其**有限的、个人的名**来直接承担的。陈签错了，受损的是她自己的学术声誉、职业前景、同行信任。她不是一个在历史危机时刻代表共同体做出决断的例外人物；她是一个普通人，用自己的名字——而非国家或机构的名字——为一个无法完全掌控的文本承担全部风险。这是一种**被剥离了神圣光环的、日常的、个人的“微主权”**。它的伦理勇气不来自历史赋予的宏大责任，而来自一个人清楚地知道她可能翻车、却仍然选择把名字挂上去的平凡承担。

第三，施密特的决断论的效力来源，在根本上是政治的——它靠主权的暴力垄断和领土边界的封闭来保障。签名权的效力来源，则完全不同：它依赖于一种更微弱、但也更普遍的社会机制——**信用**。陈的签名之所以在学术共同体中有重量，不是因为她在行使某种超验的主权，而是因为她的同事、编辑和读者相信，她过去的行为记录表明她是一个会认真审核文本的人。这种信任是可以被积累、也可以被耗尽的。一个主权者即使错了，只要还掌握权力，他的决断仍然有效；一个签名者的信用一旦破产，她的签名就变得一文不值。

正因为如此，为了进一步将签名权与霍布斯在《利维坦》中关于“授权”与“代表”的论述划清界限，还需要一笔简短的勾勒。在霍布斯的理论中，人们通过相互订立信约，将管理自己的权利授予一个主权者，主权者代表他们行动，并为此承担责任。这个结构的关键在于：授权者与被授权者之间存在一种明确的、有意识的权利让渡关系——众多个体通过一项原初契约，把本属于自己的权利转让给一个他们共同创造的代理人。而签名权所描述的伦理结构则不同。陈没有把任何事情“授权”给AI去做——她是在一个充斥着不确定性、她无法完全掌控的场域中，亲手塑造方向、否决错误、做出决断。最后，她不是让AI“代表她”签下名字。她只能自己签。那个签名的动作，不是权利的让渡，而是责任的承担——不是把权力交出去，而是把后果收回来。在霍布斯的结构中，授权者因让渡权利而获得了安全；在签名权的结构中，签署者因承受风险而获得了伦理上的站立。一个是安全化的政治契约的起点，一个是负重化的伦理宣告的终点。

由此，可以正面回应那个最致命的怀疑：签名权这个概念，是否可以被任何一个已有概念轻易替换？不能。施密特的“决断”是例外的、政治的、靠暴力保障的；签名权是日常的、个人的、靠信用维持的。韦伯的“责任伦理”建基于因果的可预知性；签名权恰恰搁置了这种预知，却仍然坚持全责。霍布斯的“授权”关乎权利的转让与安全的获得；签名权关乎责任的承担与伦理的站定。签名权的独特理论负载在于：它处理了一种这三者都没有处理的经验——在被新技术的常态性认识论断裂包围的日常中，一个普通的人用自己的名字，为一个她不完全了知后果的文本，承担全部回应义务。

（三）承诺的重铸：签名的伦理内核

为什么是“签”，而不是“写”？人类文明的历史上，签名一直是一个独特的、负载着重量的动作。“写”描述的是生产行为，是输出内容。“签”描述的则是意志行为，是确立关系。一个人在支票上签名，不是因为她参与了货币的印刷，而是因为她承认这笔支付是她授权的。一个艺术家在画作上签名，不是因为她发明了这些颜料，而是因为她宣称这是她的作品，她愿意为此接受赞美或抨击。一个元首在条约上签名，不是因为她亲自撰写了每一则条款，而是因为她以整个国家的名义为此文本背书，并以国家信用承担其全部后果。签名的伦理，从来都是与“生产”分离的。你签下的，是一个承诺，不是一份“我亲手制造”的说明。

这里有必要对“承接古老传统”这一说法做出精确限定，以避免滑回“找回失落本质”式的回归论。签名的伦理动作确实古老——它存在于人类签署条约、立定契约、起誓作证的漫长实践中。但在那些传统中，签署者签认的，是她亲眼见证、亲身参与、或至少完全了知内容的事件。签署者的权威，建立在她在场、她知情的认识论前提之上。AI时代的签名权则必须处理一个全新的、古老传统未曾面对的条件：签署者不再能够声称自己对文本的全部生产过程了若指掌。她的签名，不是在认识论充分条件下的确认，而是在认识论永远不充分条件下的承诺。

这不是“找回失落的承诺传统”。这是在**技术支持的认识论永远不完全的条件下，重新锻造承诺得以成立的结构**。传统承诺的公式是：“我知道，所以我承诺。”签名权的公式是：“我不可能全知，但我仍然承诺。”前者是认识论支撑伦理学；后者是意志行动在认识论的裂缝上架起一座伦理的桥。承诺伦理中的“说出的言语即本人”的前提，被技术条件抽掉了；签名权所做

的，不是假装这个前提还在，而是在承认这个前提已经碎裂之后，用“我愿意为这个我无法完全掌控的东西承担全部后果”的意志宣示，重新提供了一个伦理上足以站立的锚点。

撤销贡献先验，不是为了鼓励偷懒——一个认真行使签名权的人，不可能偷懒，因为审慎的签署要求她对文本进行足够深入的干预和判断。撤销贡献先验，是为了让像陈这样的人，能够不再因“无法用旧的语言说出自己的劳动”而感到羞耻。她的签名，不是一份“贡献不足”的歉意的替代品。她的签名，是她对文本的全部责任在公共世界中的唯一合法接口。她签了，她负责了，她就是这部作品的作者——不以她有没有亲手写出每一个逗号为条件。

五、签名权的实践解剖：发起、审编、发送与签名信用

前面已经将签名权在概念层面与霍布斯、韦伯、施密特划清了界限。但这套理论如果停留在“意志宣示”的悲壮姿态上，就仍然太弱了——它没有回答一个尖锐的问题：签名权的行使失败时，签署者到底如何负责？她的负责，仅仅是一种“表达悔意的意愿”，还是一种有实际后果、可评估、可积累的伦理资产？本节将在解剖签名权的实践结构之后，引入“签名信用”这一概念，将签名权从一个姿态变成一个可操作的伦理机制。

（一）发起：签名的伦理预演

发起，是指提出一个初始意图，为一个尚未存在的文本设定方向与问题感。那么，发起本身算不算签名？严格地说，不算。因为发起随时可以放弃——你可以在第一次与AI交互之后决定不做这件事，没有人会追责。但发起已经预设了签名的可能性。当陈在第一次与AI交互前，花力气凝练她那三百字的核心陈述——三层博弈、五条证据、叙事节奏的设想——她不是在“随便问问”。她是在预演：如果这个东西最终成形，它将是**我的**作品。没有这个“将是”，后续所有的筛选都没有方向。发起不是签名，但它已经为签名预备了舞台。

（二）审编：签名的实质内容

如果说发起是签名的序曲，审编就是签名的主体乐章。AI给出的从来不是一个单一答案，而是一个巨大的概率空间中浮现出的某一切面。陈的审编过程，至少涉及三种性质的判断。

第一种是**事实性判断**：识别AI对那次罢市事件因果链的错误表述，用自己的专业知识予以纠正。第二种是**论证质量判断**：否決那段事实正确但叙事疲劳的背景铺陈，要求AI回到核心机制——这不是事实对错的问题，而是她作为这个领域的研究者，对“什么论证是有力度的”拥有权威性的品味判断。第三种是**方向判断**：当AI冒出一个她此前未充分注意的次要茶商的角度时，她判断这个意外发现可能值得纳入——这是全局决断，是她十年积累的学术直觉在概率空间中捕捉到的信号。

这三种判断共同织成了签署者不可替代的劳动。它们分别指向准确性、力度和发现性。没有这些审编，她的签名就是不负责任的。有了这些审编，她的签名就内在地凝结了这些判断。审编的价值，不取决于它产出了多少新文本，而取决于它在多大程度上使最终的签名变得“有重量”。

（三）发送：不可逆的伦理决断

任何历程，如果没有最后那个正式的、以自己的名义将文本推送到公共空间的动作，此前所有的发起和审编就都只是个人的排练。发送的伦理重量，来自其不可逆性。在正式发送之前，创作者随时可以撤回、修改、作废。她有退路。一旦发送，退路就断了。签了名、发了出去，这份文本就开始在公共空间中独立地发生效果——它可能被引用、被质疑、被批评、被赞扬、被误读。而所有这些效果，都将回溯到签名者的名下。正是这份不可逆性，赋予了发送动作以特有的伦理严肃感。

（四）签名信用：失败签名与可评估的伦理资产

现在必须正面回应那个被推延至今的问题：如果陈签了名、发了出去，后来发现文中有一处AI生成的、她未检查出的错误引文——比如AI把某个茶商的姓名写错了，而她恰好在审核时漏过了——她的“签名权”在此如何兑现为实质的负责？

最简单的回答是：她发表更正声明，公开道歉。但这和她作为传统作者的“负责”在实践上没有区别。这无法证明签名权作为新伦理根基的独特价值。更深一步的追问是：在贡献先验的框架下，如果我发现自己被一篇传统方式写就的论文误导了，我不仅会追究作者的逻辑，还会形成一种对他未来学术品质的预判——“这个人的工作不严谨，下次我得多留个心”。那么，在签名权的框架下，如果我被一份AI参与生成的文本连续误导两次，我应该如何调整我对“陈”这个主体的信任？我是否应该把这种“粗疏的签名习惯”视为一种新的、可评估的学术品质？

这就是“签名信用”进入论证的时刻。签名权作为伦理行动，其长期的、可操作的社会效应，在于签署者在共同体中积累的信用记录。一个负责任的签署者——比如陈，在第一次出错后不仅发了更正声明，还在后续的工作中投入了两倍的时间开发了一套更严格的审核流程，确保同类错误不再出现——她的签名信用在共同体眼中上升了。而一个“签了名就不再管了”的签署者，当他的文本被连续挑出疏漏时，共同体不会再认真对待他今后的签名。他的签名变得空洞。他仍然可以签——没有人能阻止他——但他的签名在公共空间里失去了重量。

这时，签名权就不再仅仅是一个悲壮的姿态，而是成为了一项**可积累、可破损、可评估的伦理资产**。每一次审慎的签署，都在为签名者的信用账户充值；每一次粗疏的签署，都在取款。这种积累不同于传统“贡献”的积累——它不依赖于产出了多少新文本，而依赖于签署者的审慎判断在共同体中是否形成了一种可信赖的趋势。一个学术共同体可以逐渐形成对成员的“签名信用”的评估机制：不是去追查“这篇论文AI的参与度到底是多少”（那又落回了贡献先验的死胡同），而是去追查“这位作者在过去五年的工作中，是否建立了一个严谨签署者的信用记录。”这种评估所依赖的证据不是AI参与度的技术测量，而是作者过往发表成果的纠错记录、同行对其工作习惯的观察、以及作者本人是否主动披露了审核过程中被识别和纠正的实质性错误。这些证据当然不如文本相似度检测那样“硬”，但它们关心的本来就不是“源头”，而是“这个签署者对待签名的态度是否一贯审慎”。

这同时回应了一个可能会让签名权理论根基不稳的追问：如果“签名”这个动作仅仅依赖于签署者的意志，那它如何防止自己堕落为一种空洞的姿态——一种任何人都可以随意做出的、没有实际成本的道德表演？答案正在于“签名信用”机制。签名的重量，不来自签署者宣称的意志有多真诚，而来自她的签名信用在共同体中的累积记录。一个新人的第一次签名，重量很轻——因为信用尚未建立。一个资深学者三十年如一日的严谨签署，重量就重得多——因为她的信用账户是满的。签名的重量，不是预先给定的，而是通过一次次审慎的实际行动赢得的。一个粗疏的签署者，他仍然可以签名；但他的签名，在共同体眼中，不再具有与信用良好的签署者同等的伦理效力。签名的形式并没有变，但签名的实质重量，被签名信用的差异分配了。

这么做，就把“签名权”从一个单个的伦理姿态，转换成了一种嵌入在学术共同体长期实践中的、有实际后果的制度性伦理机制。它不再依赖每一个签署者都有崇高的道德自觉（这种依赖永远是脆弱的），而是让“审慎”和“粗疏”在公开的学术交流中自然地被看见、被记住、被纳入同行对这位学者的整体评价。签名权的运作，不再需要一个道德法庭，只需要共同体正常的学术交流机制——纠错、讨论、信任、不信任——正常运转。

六、签名的生成场域：共同体、平台与个体的三重博弈

上述分析容易给人一种错觉：签名权似乎是一个纯粹由个体内在决断建构的伦理行动。这种错觉是危险的。任何个体的签名能力，都不是在与世隔绝的内心剧场中独自炼成的。它是在一个动态的场域中，被共同体的规训、技术平台的内置权力与个体自身的伦理决断三方持续博弈所共同塑造的。

（一）被逼出的常识：共同体规训与签名的生成

一个人不是天生就知道如何审慎地签名的。这份能力，是被特定的共同体规训出来的。一个学者之所以会在AI生成一段引文后，一丝不苟地去核对原始出处，不是因为她天生诚实，而是因为学术共同体教会了她一件事：一旦引文出错，等待她的不仅仅是个人信誉的丧失，还有同行评议中的错愕、发表生涯的阻断。正是这种来自共同体的、真实的“反噬”之力——而非抽象的道德说教——持续地、无声地塑造着她的签名行为。

但签名的生成场域并不总是顺滑的。它同样可能成为僵局。一个值得沉思的场景：一家顶级学术期刊收到一份投稿，作者在致谢中坦诚写道“本文的部分论点阐述得到了大型语言模型的辅助，所有内容均经过作者逐条审核”。编辑部将这篇稿件送审，两位审稿人意见截然分裂。一位认为，只要作者能为内容的准确性负责，工具使用无关宏旨。另一位则激烈反对：“我无法判断哪些是作者的思考，哪些是模型的产出；这破坏了学术评审最基础的信任前提。”编辑部最终决定退稿，但这并非基于对稿件质量的否定——实际上，两位审稿人都认为文章本身的论证质量合格——而是基于一个制度性的犹豫：他们不知道该用什么标准来评审这样一份稿件，更不知道该以什么理由来为它正名。退稿，不是对稿件的否定，而是对不确定性的回避。

这个僵局揭示了一个关键的问题：签名权作为伦理行动的有效性，并不仅仅受制于个体是否负了责；在某些关键的节点上，它被共同体尚未准备好接纳它的制度性空白所卡住。共同体的规训在此不是失效了，而是尚未更新。

但僵局并不代表全部。在一些领域，微规范正在从实践中被开辟出来。试设想一个由两位语言学家和一位计算科学家组成的研究团队，在撰写一篇关于“AI翻译中的文化偏见”的论文时，从一开始就决定将签名权作为其工作的伦理框架。他们在项目启

动时便在内部文件中对AI的使用做出明确约定：所有AI生成的内容，仅作为草稿，不直接进入终稿；每一位署名作者必须对AI生成的内容中与自己专业领域相关的部分做逐条审查，并在共享文档中标注审查结果；在最终发表时，论文的致谢与声明部分，将包含一段清晰的陈述，说明AI参与的具体环节，并且声明：“所有署名作者已对本文全部内容完成审慎的最终审核，并对此承担全部学术责任。”

在内部沟通中，团队的计算科学家起初有抵触——“这是不是暗示我的工作不够原创？”——但在两轮讨论后，他的看法发生了转变。他事后在一封内部邮件中写道：“这让我更认真地看待那些我原本可能会快速扫过的AI生成段落。想到一段话最终会挂在我名字下面，我读它的方式就变了。”这句话，是签名权从理论论述变成可操作的伦理实践时，最简洁的印证。团队最终在期刊的“作者贡献声明”部分，以一种既精确又不冗余的格式，完成了对AI参与和人类签署关系的公开标注。该论文顺利通过了同行评审。一位审稿人甚至在评论中特别提到：“作者关于AI使用的声明是透明且专业的，这并未削弱我对论文质量的信任，反而增加了我对作者学术严谨性的尊重。”

这两个场景——拒稿的僵局与接纳的正面实践——共同指向了签名权在生成场域中的核心机制。签名权不是在社会真空中被行使的，它是被逼出来的：被旧的制度性空白逼出来的，也被像上述研究团队这样的、在共同体内部有意识地构建新的微规范的行为拓出来的。那个成功标注的案例，则展示了一种可能的前进方向——共同体规训的更新，不是也不可能是一个自上而下的、一蹴而就的普遍立法，而是由无数个分散的、具体的实践主体，在各自的领域内，以诚实的伦理承担和明确的程序建构，一点点地拓宽那条可以容纳“签名权”的合法化路径。

（二）平台的技术权力：签名在怎样的预设空间里发生？

另一个在签名权的浪漫叙事中容易被屏蔽的力量，是AI平台自身。任何商用大模型，都不是一块白板。在用户输入第一句提示词之前，它已经被内置了系统级的指令、微调偏好、安全审查机制和输出过滤。这些预设，对用户的签名行为施加了实质性的、但不透明的干预。

当一个创作者在这样一个平台上，“自发性”地做出了一系列审编和判断，她可能没有意识到，她所看到的“可能性空间”本身，已经是平台预先做过一轮不透明路由之后的结果。她在第二轮上签了名，但第一轮路由——那个决定了哪些路径根本不出现、哪些表述一开始就被压制的路由——对她而言，是不可见的。这就是签名权的结构性张力所在。签署者不是在一个完全由自己掌控的空间中自由地做出判断，而是在一个部分被平台预置的权力结构所框定的空间中，做出“她所能做的最好的判断”。

这部分的论述，不是要否认个体签名的责任，而是要诚实地点明：签名权要求个体对发送负责，但它同样要求平台权力承担相应的透明度义务。两者都必须被纳入理论视野，否则签名权就容易被扭曲为一种将系统性风险转嫁给个体的、精致的新自由主义话术——这一点将在下一节被正面回应。

（三）个体的伦理决断：被博弈训练出来的审慎

共同体规训有滞后，平台权力有不透明。但签名权论述的最终落脚点，仍然是那个具体的人。即使在一个不完善的结构中，她仍然能够——而且应当——做出属于自己的、承担后果的签署决断。她的自主性，不在于她完全掌控了信息的所有源头（这从未可能），而在于她对自身的判断保持着警觉，并在这种警觉中做出选择。

但这种警觉本身也需要进一步追问其来路。它不是个体与生俱来的内在美德。它是被逼出来的。一个学者在使用AI的过程中，经历过至少一次“我以为是对的，后来发现被AI误导了”的惊悚之后，才会真正生成那种审慎的自觉。她的“自知”，恰恰是她在共同体的规训和平台权力的不透明双重挤压下，在不断碰壁之后，被迫生成的防御性智慧。她不是一个天生诚实的签署者。她是一个在博弈中被训练出来的签署者。她诚实，不是因为她的灵魂纯洁，而是因为她知道不诚实的代价。

这份被逼出来的诚实，才是它在伦理上最值得珍惜的部分——它不依赖任何“人之本质”的浪漫预设，而扎根于真实的社会互动与制度反馈之中。文章由此在逻辑上完成了闭环：签名权不是个体内在品质的外溢，而是在共同体的规训压力、平台的权力预设、和个体的自我保护需求这三重力量的实际博弈中，被动态地生成的伦理结构。

七、对话与反驳：让论证逼出最深处

一个新概念，要站得住，必须正面格挡那些最有力、最诚恳的反驳。以下的对话与反驳从最底层的攻击开始，经边界划定，最终收束到与既有制度立场和最强解构者的正面对表。

（一）反驳一：这只是包装过的免责声明？

这是本文必须打的最硬的一场仗。这个反驳，比“它可能被滥用”更凶狠，因为它攻击的不是签名的“滥用”，而是签名的“本质”：它声称签名权这个名词本身，就是一个精致的骗局——用一个漂亮的道德姿态，把“我根本没有亲手写出这些句子”这件本来应该构成伦理缺陷的事实，重新命名为一项“勇敢的宣言”。陈连自己论文的句子结构都没亲手搭建，她凭什么声称对这个句子的质量负责？

这个反驳之所以有力，是因为它戳中了贡献先验最顽固的直觉：如果“写出”在过去是作者证明自己“能够负责”的最直观凭证，那么撤销了“写出”之后，签名权提供的是一份什么样的凭证？

回应分两步。第一步：承认局限，但拒绝还原。签名权确实不要求签署者是每一个句子的“源头生产人”。在这个意义上，陈的“负责”和责任伦理中政治家对自己亲手发动的政策的负责，在形式上不同。她不能指着一段平滑的过渡句说“这是我写的”——那段文本的初始形态确实来自AI。如果“负责”的唯一合法形式是“对亲手产出的每一个细节承担全责”，那么她确实没有资格。

但谁规定这必须是“负责”的唯一合法形式？这正是贡献先验的隐蔽统治。一个文本在公共空间中的存在和效果，并不需要以“作者亲手写出每一个字”为前提。一份政府白皮书可能由数十位幕僚起草，但最终签发的官员承担的不是“我写了每一段”的责任，而是“我决定让这成为我们政府的立场”的责任。这种责任，与“亲手写出每一个字”的责任是两种不同的责任。前者是生产者的责任，后者是签署者的责任。签名权所建构的，正是后者。

第二步：出示真正的伦理内核。反驳者的问题在于，他们用“生产”的标准来衡量“决定”。陈是否能为文本的质量负责？她的负责方式，不是保证“每一个句子的语法最优”——那是AI的专长，也不是她的承诺范围。她负责的，是作为“这个最终判断和定向系统的源头”。她负责的是：她已经做了她在那一刻能做的、足够审慎的判断；她愿意为任何被她疏忽的错误承担后果；她愿意为文本在公共空间中产生的影响接受质询。她不能保证文本完美——谁敢如此担保？但她保证她将站定。“我负责”这句话在这里的实质含义，是“我承诺为我对这个文本的判断承担回应义务”。这个承诺，不依赖于她是否亲手写出每一个句子。它依赖于她的意志是否参与了最终成品的生成，并且她是否愿意让这个意志被公开识别为责任的锚点。签名权不是把“我没写”糊弄成“我负责”。它是说：在AI时代，“写出”已经不再是负责任的必要前提，但“审慎地签认”，是一个人可以给出的、最诚实的责任凭证。

此外，上一节提出的“签名信用”机制为这一辩护提供了制度性的支撑。“负责”不是一句空话，因为签字者的签名信用——她过去在共同体中积累的严谨记录的可见账本——构成了衡量这一签名的重量和未来可信任度的实际标准。她的“站定”不是单次表演，而是一份可以被长期追踪的伦理记录。

（二）反驳二：这是新自由主义陷阱吗？

现在，必须给这个反驳加上更尖锐的形态——签名权难道不是一种在没有制度配套的情况下，让个体用“勇敢的姿态”去个人化地承受系统性风险的新自由主义陷阱吗？这个命题的杀伤力在于，它指控签名权不仅是一种幻觉，还可能是一种有害的安慰剂：它让陈在制度空白的地方，明明是在为平台的不透明和共同体的迟滞付出代价，却把这种牺牲美化为“伦理勇气”。

这是一个必须被正面格挡、不能回避的指控。如果签名权只是说“每个人都必须自己承担后果”，而完全不要求制度提供透明的信息和公平的规训，那么它确实可以被批判为一种精致的毒药。本文接受这个反驳所提出的制度要求——签名权作为伦理行动的有效性，确实有赖于制度配套的跟进。它需要平台提供更透明的生成路径记录（让签署者能够更充分地了解她所面对的那个概率空间），需要共同体更新评审标准（让签署者能够在明确、公平的框架内负她该负的责任），需要学术发表规范给坦诚留下合法的空间（让她不至于因诚实而受损）。

但这恰恰是签名权的双重性质所在。它既是个体站定的伦理宣言，同时也是对制度提出的一份明确的政治请求——因为如果制度不跟上，签名权的伦理重量就会因为“不可见的第一轮路由”和“陈旧的评审标准”而空洞化。这里没有先后顺序：伦理站定是第一步，让它站定变得有效的制度重建是不可以被省略的第二步。两者是同一个历史过程的两个侧面。新自由主义的陷阱在哪里？在把“个体责任”当作拒绝制度改革的借口。而签名权的坚持者恰恰要说的是：正是因为个体已经先期站定、并为此付出了代价，共同体才必须完成那不可或缺的第二步——更新规训、增加透明度、重新设计容纳诚实协作身份的制度接口。在这个意义上，签名权不是陷阱的入口，而是逼出制度更新的政治请求的出口。

反驳者的攻击，在这一点上反而倒过来被吸纳进了签名权的论证：是的，没有制度配套的签名权确实是一种罔顾现实的空洞姿态。正因如此，签名权的完整论述才必须包含制度性的一维。它不是一个自我满足的道德姿态，而是一个要求共同体采取行动的政治起点。

（三）反驳三：签名权在文学创作中是否完全失效？

这是一个本分地界定理论适用边界的问题。本文的论证，建立于非虚构写作（学术、新闻、政策分析等）以知识、论辩、信息传达为主要功能的文类。在这一类写作中，语言的主要功能是工具性的——它承载判断、表述事实、推进论证。只要这些功能被完好地实现，将文本生成这一环节移交给AI，只损失了手工作坊的质感，并未损害核心的智识与伦理贡献。签名权在此有效。

但文学创作——尤其是诗歌与小说——是另一回事。对于一个小说家而言，写作最核心的部分，恰恰发生在亲笔去写的过程之中。风格不是可以事后套在“内容”外面的一层装饰；风格就是作家本人的存在方式，是她在特定生存境况下用全部生命经验锻打出的不可复制的声音。这种声音的锻造，无法被AI代理，因为代理就意味着丧失。“写出”这个动作本身，在文学中不是工具，而是目的。

（四）反驳四：福柯的“作者功能”是否已经处理了这个问题？

在福柯1969年的演讲《什么是作者？》中，他早已指出，“作者”不是一个自然身份，而是特定话语实践所设置的一种“功能”——一种负责对文本进行分类、归类和赋予价值的操作。签名权在某种意义上，不正是“作者功能”在AI时代的具体化吗？本文是否只是在重复福柯？

这一对话是必要的，因为如果不划清边界，签名权就可以被轻易收纳进福柯的框架。关键的区别在这里：福柯描述的是**结构**——话语实践如何通过“作者功能”这个位置，来组织文本的流通和价值的分配。签名权描述的则是一种**行动**——一个具体的、活生生的人，如何通过一次意志决断，将自身建构为那个愿意承担全责的责任主体。福柯关心的是“作者功能”如何作用于文本；签名权关心的是“什么人、通过什么动作，才能在AI时代承载起这个功能”。

在一个生成式AI可以自动产生有理有据文本的时代，仅仅说“作者是一个功能”是不够的。因为问题恰恰变成了：当文本生产可以被解构为非人的操作时，什么样的人的行动，才能让这个“作者功能”继续站定？福柯的结构分析不能回答这个问题，因为它不处理“如何将自身建构为这个功能的责任主体”的伦理行动。签名权的回答是：唯有通过“签”——那个在审慎引导后以自己名义发送的意志动作——人才能将自己建构为足以承载作者功能的主体。结构指定了位置；签署才使一个具体的人入驻那个位置。

（五）出版伦理界不是已经这么说了吗？——与 ICMJE 与 COPE 立场的对话

一个熟悉学术出版规范的读者会指出：本文的核心倒转，制度早已采纳。自 2023 年起，《自然》等期刊的编辑方针、出版伦理委员会（COPE）的立场声明与国际医学期刊编辑委员会（ICMJE）的建议高度一致：大语言模型不得被列为作者——理由恰恰是**作者身份意味着对作品负责，而 AI 无法承担这种责任**；人类作者须披露 AI 的使用，并对 AI 参与生成的内容承担全部责任。问责先于贡献——这不正是签名权吗？如果制度已经这么做了，本文还剩下什么？

剩下两件制度没有做、也无法在政策文本里完成的事。第一，制度立场是**门卫式的否定判断**：它回答了“AI 为什么不能签”，却没有回答肯定的问题——在认识论不透明已成常态的条件下，人凭什么能签？签的动作由哪些环节构成（发起、审编、发送）？签名的重量从哪里来、如何积累、如何破产（签名信用）？政策划出了门槛，本文建造的是门槛之内的房间：一套让“人来负责”不止于宣示、而成为可解剖、可评估之伦理行动的结构。

第二件事更要害：这些制度在用问责排除 AI 的同时，**原封不动地保留了贡献先验**。ICMJE 的作者资格标准，第一条仍然要求对研究的“实质性贡献”；于是制度内部出现了本文所诊断之撕裂的规范化版本——一条腿站在问责上（用它拒绝 AI），另一条腿仍站在可追溯贡献上（用它认定人）。陈的困境恰恰卡在这两条腿的接缝里：她的问责是完整的，她的劳动却在旧的贡献计量下无法入账，而披露义务在缺乏正面承认接口的现状下，实际上在惩罚诚实者。签名权不是复述制度已说的半句话，而是把作者资格改立在一条腿上——审慎签署的问责——并为这条腿补上它此前缺少的伦理解剖与信用机制。在这个意义上，本文与其说提出了一个政策没有的立场，不如说替政策已经采取的立场补上了它欠缺的理论地基，并顺带诊断出政策自身尚未察觉的内伤。

（六）德里达的幽灵：可复制的签名还能承重吗？

最后必须请出的对话者，是德里达。在《签名 事件 语境》（1972）中，德里达论证：签名之为签名，恰恰以可重复性（itérabilité）为条件——一个只能签一次、无法被再次辨认的记号不成其为签名；而正因为签名必须可重复、可模仿、可从签署那一刻“活的在场”上剥离下来，它就永远无法纯粹地担保签署者的在场与意图。签名想要证明“我在这里、这是我”，它的结构却注定了这份证明的不可能。一个德里达式的批评者会说：你把作者身份押在签名上，可签名恰恰是被解构得最彻底的那个动作。

这个批评瞄准的靶子，本文并不站在上面。德里达解构的是作为**在场证明**的签名——向后看的、认识论的签名：“这个记号担保源头。”这一面本文不仅接受其被解构，而且以之为前提：贡献先验的崩解，正是“签名=源头证明”的崩解（第二、三节）。签名权重建的是另一面——**向前看的、伦理的签名**：“我自此为它回应。”承诺不需要在场的形而上学：它是一个把未来的自己抵押进去的动作，可重复性并不消解它。恰恰相反，德里达的难题反过来支撑了第五节的机制：正因为签名的记号本身担保不了任何东西——它可以被复制、被伪造、被机械地盖下——签名的重量就不可能来自笔迹，而只能来自签署者在共同体中可追踪的行为记录。签名信用，就是对德里达难题的制度性回答：它不再徒劳地把权威放回不可能的“在场”，而把它安放在可积累、可核查、可破产的“记录”上。解构清空了签名的过去；签名权在被清空的地基上，用未来重新给它配重。

八、结论：签字者的重量

现在，可以回答那个最朴素的问题：究竟是什么，让一个人在AI时代仍然配得上“作者”这个称谓？

经过全部论证的推进，答案是：不是因为她贡献了某种不可被机器替代的、从她独特心灵深处流淌出来的独特文本——那是一个已经被AI在工程化层面彻底揭露的神话。而是因为，她用自己的名字，为这段文本在公共空间中的存在，承担了全部不可推卸的回应义务。她发起了方向，承受了审编的枯燥与耗神，她在清楚地知道自己不可能全知全能条件下，仍然做出了最终的、不可逆的发送决定。她在发送键上，完成了那个只有人才能完成的动作——决定让一个东西存在，并准备用一生来为其负责。

这就是签名权。它的理论份量，不在于它为AI时代的创作提供了一个新的辩护策略，而在于它将作者责任的根基，从“对源头负责”的认识论负担中解放出来，重新锚定在“对发送负责”的伦理决断上。同时，通过与霍布斯的授权结构、施密特的例外决断、韦伯的责任伦理这三位思想巨人进行正面对话，签名权切开了三人都未能打开的一个辨别面：它是日常的，而非例外的；是个人信用的，而非政治暴力的；是认识论断裂已成常态条件下的承诺，而非因果关系基本清晰条件下的计算。这个辨别面，只有在一个AI大规模介入文本生成的时代的日常经验中，才会变得可见。

但签名权不只是一份个人宣言。它的提出本身，就是一个双重动作：一方面，它让创作者能够在内心站定，完成伦理上的自我确证；另一方面，它向共同体和平台发出了一个不可撤回的政治请求——要求制度提供透明、公平的运作条件，使得签名权的伦理重量不至于在结构性的不透明中被掏空。个体的决断与共同体的重建，不是先后关系，而是同一枚硬币的两面。签名权的完整意义，不会由这篇论文来完成。它只能由无数个在具体情境中做出了自己的签署决定、并坦然接受质询的人，在他们各自的学术、新闻、非虚构创作的领域里，一笔一画地继续下去。

本文的核心判断——在AI时代，人从“源头型贡献者”转变为“签名型责任人”，签名权替代贡献先验成为作者身份的伦理根基——在以下情况下将被证伪。如果在未来的某一刻，AI的生成过程变得对使用者完全透明，以至于一个签署者能够完全理解她的每一句提示、每一次筛选如何导致最终文本的所有后果——也就是说，那个在本文中被识别为签名权伦理内核的“认识论断裂与伦理全责之间的悖论”被技术上消解了，人类重新获得了对文本生成过程的完全预知能力——那么在那种情境下，签名权或许会重新被吸纳回某种更精密的“责任伦理”之中。如果更进一步，AI自身亦被社会制度赋予了为文本承担独立法律责任的主体地位，那么“签名”的人类独特性亦将瓦解。但如果那样的未来竟然到来，改变的将不独是“作者”的定义，而是人之为责任主体的全部伦理格局。在那一天到来之前，我们必须先在当下学会，如何珍重和维护那个在发送键上仍然只属于人的动作：决定让一个东西存在，并用你的名字——而非其他任何人或物的名字——来为它承担全部重量。

注释

[1] 本文中所有涉及具体人物、场景和决策过程的案例（包括历史学讲师陈、学术期刊退稿事件、研究团队的内部约定等），均为基于当前实践张力的思想实验与例示建构，旨在呈现结构性机制，而非描写真实个案。其中的人物、机构与情节已经过匿名化、合成与简化处理，不具有经验研究意义上的可验证性，亦不构成对任何特定个人、机构或事件的指涉。

[2] 本文提出的“贡献先验”是一个分析性概念，用于揭示现代作者身份建构所依赖的深层预设。它不是对康德哲学中“先验”概念的严格沿用，亦不直接介入先验哲学论争。在此语境下，“先验”用来指称那些在“作者”身份被习以为常地使用前、已被制度、观念与心理所共同预设的逻辑前提。与福柯的“历史先验”不同，本文的分析重点在于经验制度层面，其理论旨趣更接近于描述使特定社会身份得以稳定运作的历史条件及其转移。

[3] 施密特的主权决断论见其《政治神学》（1922）；霍布斯关于授权与代表的论述见《利维坦》第十六章。本文与此二者的对话旨在确立签名权的理论独特性，而非对二者学说做全面评述。

[4] “把关人”概念最初由库尔特·勒温在1947年提出，后由大卫·曼宁·怀特于1950年引入传播学，用以描述新闻编辑对信息的筛选过程。本文将其与“签名权”相区分，核心差别在于：把关人面对的是已知的备选信息，行使的是“放行/阻止”的二元权力；签名权的行使则嵌入在与生成式AI的持续交互中，签署者参与塑造的是尚未成型的文本可能性空间。

[5] “作者之死”论典出罗兰·巴特1967年的同名文章以及米歇尔·福柯1969年的演讲《什么是作者？》。这两篇文献在解构作者作为文本意义唯一源头的功能上具有奠基意义。本文的讨论虽然触及作者身份的消解与重构，但其主要着力点不在文本意义生成的独立性上，而在作者作为责任承担者的伦理位置，故不直接介入巴特与福柯的文本理论，仅在背景中将其视为当代讨论的参照系。对福柯“作者功能”与本文“签名权”的正面辨析，见第七节（四）。

[6] 关于平台技术权力与算法不透明性的讨论，可参考批判算法研究及平台研究领域的相关文献。本文的论述旨在指出一个常被忽略的结构性事实，而非对特定平台的技术实践做出实证判断。在签名权的理论框架中，对平台预设的反思构成了负责任签署的一个必要环节。

参考文献

- Barthes, Roland. 1967. "La mort de l'auteur." *Manteia*, V. 收入 *Le Bruissement de la langue: Essais critiques IV*, 61-67. Paris: Éditions du Seuil, 1984.
- COPE (Committee on Publication Ethics). 2023. "Authorship and AI Tools." COPE Position Statement, February 2023.
- Derrida, Jacques. 1972. "Signature Événement Contexte." In *Marges de la philosophie*, 365-393. Paris: Éditions de Minuit.
- Foucault, Michel. 1969. "Qu'est-ce qu'un auteur?" *Bulletin de la Société française de philosophie*, 63(3): 73-104. 英文译本收入 *Language, Counter-Memory, Practice: Selected Essays and Interviews*, edited by Donald F. Bouchard, 113-138. Ithaca: Cornell University Press, 1977.
- Hobbes, Thomas. 1651. *Leviathan, or The Matter, Forme, & Power of a Common-wealth Ecclesiasticall and Civill*. London: Andrew Crooke. 尤见第十六章"Of Persons, Authors, and Things Personated".
- Schmitt, Carl. 1922. *Politische Theologie: Vier Kapitel zur Lehre von der Souveränität*. München und Leipzig: Duncker & Humblot. 英文译本 *Political Theology: Four Chapters on the Concept of Sovereignty*, translated by George Schwab. Cambridge, MA: MIT Press, 1985.
- Weber, Max. 1919. "Politik als Beruf." 收入 *Geistige Arbeit als Beruf: Vier Vorträge vor dem Freistudentischen Bund*. München und Leipzig: Duncker & Humblot. 英文译本 "Politics as a Vocation," in *From Max Weber: Essays in Sociology*, translated and edited by H. H. Gerth and C. Wright Mills, 77-128. New York: Oxford University Press, 1946.
- ICMJE (International Committee of Medical Journal Editors). 2023. *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals* (May 2023 update), § II.A.4.
- Nature. 2023. "Tools such as ChatGPT threaten transparent science; here are our ground rules for their use." *Nature* 613, 612.
- Rose, Mark. 1993. *Authors and Owners: The Invention of Copyright*. Cambridge, MA: Harvard University Press.
- Woodmansee, Martha. 1984. "The Genius and the Copyright: Economic and Legal Conditions of the Emergence of the 'Author' ." *Eighteenth-Century Studies* 17(4): 425-448.