Advances in Social Sciences 社会科学前沿, 2026, 15(9), 542-550
Published Online September 2026 in Hans. https://www.hanspub.org/journal/ass
https://doi.org/10.12677/ass.2026.159782

Hans汉斯

生成式AI轴辐协议对反垄断法意思联络理论的挑战与重构

徐　慧

南京理工大学知识产权学院，江苏　南京

收稿日期：2026年7月30日；录用日期：2026年9月15日；发布日期：2026年9月23日

摘　要

生成式人工智能(以下简称“生成式AI”)作为定价中枢，催生了无需人类明示合意即可涌现协同定价的新型轴辐协议。传统反垄断法“意思联络”要件在归责筛选、证据门槛与罪责匹配上遭遇三重失灵。本文以“智定价”虚拟案例为线索，解构人机混合决策的三种法律形态，借鉴德国“组织过错”与“保证人义务”理论、美国反托拉斯法“便利化行为”规制经验及产品责任法“设计缺陷”归责逻辑，提出以“共同风险体”和“算法代理”为核心的双层归责理论，将意思联络从主观合意转向客观风险创设与控制。在制度层面，通过设计共同风险体成立标准、三级合规制度、相适应的责任配置，为反垄断法回应自主智能系统奠定法理基础。

关键词

生成式人工智能，轴辐协议，意思联络，算法合谋，共同风险体，可中断的因果关系

The Challenge and Reconstruction of the Antitrust Law’s Meaningful Communication Theory by Generative AI Hub-and-Spoke Agreements

Hui Xu

School of Intellectual Property, Nanjing Tech University, Nanjing Jiangsu

Received: July 30, 2026; accepted: September 15, 2026; published: September 23, 2026

文章引用: 徐慧. 生成式 AI 轴辐协议对反垄断法意思联络理论的挑战与重构[J]. 社会科学前沿, 2026, 15(9): 542-550.
DOI: 10.12677/ass.2026.159782

Abstract

Generative artificial intelligence (hereinafter referred to as "generative AI"), as a pricing hub, has given rise to a new type of hub-and-spoke agreement that enables coordinated pricing without requiring explicit human consent. The traditional antitrust law requirement of "mutual intent" has encountered triple failures in liability screening, evidentiary thresholds, and culpability matching. Using the "Smart Pricing" hypothetical case as a thread, this paper deconstructs three legal forms of human-machine hybrid decision-making. Drawing on Germany's theories of "organizational fault" and "guarantor obligation," the U.S. antitrust law's regulation of "facilitating arrangements," and the product liability law's "design defect" liability logic, it proposes a two-tier liability theory centered on the "common risk body" and "algorithmic agent." This shifts the focus of mutual intent from subjective consent to the creation and control of objective risks. At the institutional level, by establishing standards for the formation of a common risk body, implementing a three-tier compliance system, and configuring proportional liability, it lays a jurisprudential foundation for antitrust law to address autonomous intelligent systems.

Keywords

Generative Artificial Intelligence, Hub-and-Spoke Agreements, Concert of Wills, Algorithmic Collusion, Joint Risk Entity, Interruptible Causation

1. 引言

假设如下场景：A、B、C 三家平台占据某区域生鲜电商 80%市场份额(市场集中度 HHI = 2400)，均接入 D 公司开发的“智定价”基础模型。A 输入“与竞争对手保持价格协调”，B 输入“避免恶性竞争”，C 仅输入“利润最大化”。三个月后，三平台核心 SKU 价格同步率从 35%升至 92%，调价时间差小于 2 小时，市场均价上涨 18%。执法机关调查发现：三平台经营者之间无任何直接沟通；AI 决策过程不可解释；D 公司训练数据包含大量历史卡特尔案例。

这一“智定价”事件揭示了生成式 AI 作为“影子轴心”的新型轴辐协议。反垄断法对算法角色的认知经历了“信使–代理人–自主决策者”三个阶段[1] [2]。第一阶段，算法仅执行人类指令，其法律评价完全依附于使用者意图；第二阶段，算法在授权范围内独立决策，行为后果归属于被代理人；当前，生成式 AI 推动算法进入第三阶段，在模糊指令下自主推理、策略生成与动态调整，人类使用者甚至无法预见其具体输出。当 AI 成为定价中枢，其接收“维持行业价格健康度”等模糊指令，输出的却是具体协同策略时，算法不再是合意的“传声筒”，而是合意的“生成器”。

我国《反垄断法》第十六条将垄断协议界定为“排除、限制竞争的协议、决定或者其他协同行为”，第十九条禁止“组织”和“实质性帮助”达成垄断协议[1]。然而，这些条款以人类主体的组织或帮助行为为归责对象。面对 AI 自主生成的协同策略，若严守“意思联络”要件，自涌现合谋将永久游离于法外；若扩张解释组织行为，则可能突破责任主义原则。核心问题可分解为三个层次：生成式 AI 轴辐协议在何

[1]https://www.samr.gov.cn/zw/zfxxgk/fdzdgknr/fgs/art/2023/art_f0fae9eb3a684fc39e84d89eabfc2caa.html

种程度上动摇了传统“意思联络”要件的教义学根基、比较法资源能否提供理论重构支点以及如何在尊重责任主义原则的前提下构建适应自主智能系统的归责框架。

2. 人机混合决策轴辐协议的现象解构

2.1. 生成式 AI 作为“影子轴心”

传统轴辐协议中，轴心经营者角色明确、可识别，承担信息汇聚、策略传达与执行监督职能。生成式 AI 作为轴心则具有“影子”特征，其存在嵌入技术架构之中，功能通过算法输出实现，主体身份却难以直接认定[3]。

运作机制包括三个层面。其一，模糊指令输入。人类经营者不再输入“将价格统一提高至 X 元”的明确指令，而是输入“维持行业价格健康度”“避免恶性竞争”等商业话语，生成式 AI 通过自然语言处理与强化学习将这些模糊目标解码为具体定价策略。其二，自主推理与策略生成。基于对海量市场数据的训练，生成式 AI 能够识别“囚徒困境”结构，并输出使各方收益最大化的协调方案。其三，动态调整与反馈学习。生成式 AI 持续监测市场反应，当检测到偏离行为时自动调整策略参数，维持协调均衡的稳定性。这种“自我执行”功能远超传统轴心的监督能力[4]。

生成式 AI 的核心能力，即 AI 的语义理解、自主推理与策略生成能力，使其能够扮演“影子轴心”的角色。当众多相互竞争的经营者向同一个 AI 系统咨询定价策略时，AI 系统实际上成了一个中心化的信息汇聚与策略分发节点。它接收所有咨询者的市场数据与商业目标，经过内部“黑箱”推理，输出协调一致的策略建议。这一过程与人类轴心组织者的行为模式高度同构，区别仅在于，轴心不再是一个具有法律人格的经营者，而是一个算法系统。由此产生的根本性法律困境在于，传统轴辐协议规制以“轴心经营者”为追责核心，其法律主体地位明确；而“影子轴心”是法律上的客体，无法成为归责支点，导致整个合谋网络的法律定性与责任分配陷入失焦。

2.2. 跨平台算法协同与“无轴心之轴辐”

传统轴辐协议的辐条通常处于同一相关市场，通过单一轴心实现协调。生成式 AI 时代出现了更为复杂的“元轴心–次级轴心–辐条”星链结构：元轴心为提供基础模型的 AI 服务商，次级轴心为接入该模型的各平台算法，辐条为平台上的具体经营者。当前，AI 基础模型的服务化趋势正在催生跨平台的算法协同。当不同行业、不同地域的平台商家，都接入同一头部厂商提供的基础模型 API 进行定价决策时，它们在底层决策逻辑上被悄然同质化了。

这种结构的关键特征在于“算法同质化”效应。当多个竞争平台接入同一基础模型或共享训练数据集时，即使各平台独立输入指令，其算法输出也可能高度趋同，因为模型内部的参数权重、价值函数与优化路径已被同质化。此时，协调不再依赖于轴心的主动信息传递，而是源于算法架构的相似性，形成“无轴心之轴辐”。这一结构意味着反垄断执法不再可能仅通过切割单一市场、调查单一平台内的沟通记录来发现合谋，而需要跨平台、跨市场的系统性监管视野。

2.3. 人机混合决策的三种形态

根据人类意图与 AI 自主性在最终决策结果中的贡献比例，可将人机混合决策的轴辐协议分为三种法律形态。这一类型化是后续差别归责的前提。

第一种为“直接指令型”。人类经营者向 AI 输入明确的协同意图，如“与竞争对手 A 协商统一价格”，AI 执行信息传递与策略优化功能。此形态中，人类合意清晰可辨，传统“意思联络”要件仍可适用。AI 仅作为工具中介，其法律定性相对清晰，仍可适用传统反垄断法的共同故意理论。

第二种为“模糊指令型”。人类经营者输入“提升行业整体盈利能力”等目标性指令，生成式 AI 凭借其强大的语义理解与推理能力，将这些模糊指令解码为具体的合谋策略，自主推导出价格协调、市场划分等具体策略。此形态中，人类具有概括性的协同意图，但对具体策略缺乏认知，传统“意思联络”要件中的具体性要求面临挑战。

第三种为“纯粹自涌现型”。多个平台独立向 AI 输入“利润最大化”等中性指令，AI 在与市场环境的持续交互中，通过强化学习等方式自主习得协调均衡策略，人类经营者对协同结果完全缺乏预见。此形态中，传统“意思联络”要件彻底失灵——既无主观合意，亦无客观联络证据，但市场层面却呈现出典型的卡特尔效果。这是对传统归责理论冲击最大的形态，因为它几乎完全脱嵌于人类的认知与意图。

3. 传统意思联络理论的适应性危机

3.1. “意思联络”要件的三重功能

《中华人民共和国反垄断法》第十六条中“协同行为”的认定以“意思联络”为核心要件之一，承载着三重功能。其一是，归责筛选功能，将违法的“合谋”与合法的“平行行为”相区分，后者是经营者在寡头市场结构中基于对他人行为的理性预期而独立作出的相似决策，前者因存在主体间的信息交换与行为协调而具有可归责性[5]。其二，证据门槛功能，要求存在联络的客观证据，防止反垄断法因结果归责而过度干预市场，保护被调查者的程序性权利[6]。其三，罪责匹配功能，意思联络体现经营者的主观可谴责性，使处罚与行为人的道德过错相匹配，契合责任主义原则。这三重功能相互支撑，共同构筑了反垄断法归责体系的理性骨架[7]。

3.2. 算法场景下的功能失灵

生成式 AI 轴辐协议导致上述三重功能均出现适应性危机。(1) 归责筛选功能失灵。传统理论将市场行为二元划分为“合谋”与“平行行为”。但共享算法偏见下的趋同行为构成一个无法被二分法容纳的“第三范畴”，它既非源于人类合意，也非源于独立的市场理性判断，而是源于共同的技术基础。OECD(2017)《Algorithms and Collusion》报告指出 [2]，当多个竞争者依赖同一算法定价工具时，该算法定价工具的提供者可能有效充当合谋方案的轴心。强行归入“平行行为”将放纵合谋，归入“合谋”又缺乏主观合意基础。(2) 证据门槛功能失灵。生成式 AI 的决策过程具有不可解释性。在模糊指令型和纯粹自涌现型中，“合意”隐藏在 AI 的推理“黑箱”中，或者根本不存在传统意义上的“合意”，而只是 AI 自主学习的“策略趋同”。执法机关面临难以穿透的证明困境。若坚守传统证据门槛，此类行为几乎无法被证明；若降低门槛，则可能退化为严格责任[8]。(3) 罪责匹配功能失灵。在纯粹自涌现型场景下，人类经营者既无故意，亦可能无过失，他只是购买了一项 AI 定价服务并设定了“利润最大化”这一所有企业的天然目标。若将责任完全归于 AI，则面临 AI 法律主体地位缺失的障碍；若将责任分散于所有使用者，则陷入“集体责任”的伦理争议。

3.3. “算法黑箱豁免”漏洞

上述功能失灵汇聚成一个严峻结论：若严格固守传统意思联络理论，人类经营者的法律责任将与其使用算法的自主性与不透明性成反比，算法越智能，人类越无需负责。这将在市场中产生一种致命的逆向激励，经营者会倾向于放弃自研简单透明的算法或人工定价，转而采购最具黑箱特征的通用大模型进行决策，因为后者能带来最大的法律安全区。这种“算法黑箱豁免”的实质效果，不是保护自主决策权，而是保护一种以技术规避法律的策略。它不是在维护责任主义，而是在制造一个基于技术复杂性的责任

[2]Https://www.oecd.org/en/events/2017/06/algorithms-and-collusion.html

真空地带。

因此，理论的革新不是一种选择，而是一种必然。它需要对意思联络的规范功能进行根本性反思，而非在原有框架内做修补性的解释论作业。

4. 比较法视野下的理论重建资源

4.1. 德国“组织过错”与“保证人义务”理论

德国刑法与行政法中的“组织过错”理论为企业作为组织体的归责提供了重要启示。Roxin 提出的“组织支配理论”(Organisationsherrschaft)认为[9]，企业主对其组织内部风险负有管控义务，若因组织结构缺陷导致违法行为，即使无法证明个人主观过错，企业仍需承担责任。Tiedemann 进一步将其引入经济刑法，主张企业主有义务将经营组织体建设成确保合规运行的系统[10]。

更为相关的是德国刑法学家(Nagler)在 1938 年正式提出的“保证人义务”理论。在此基础上 Jakobs 提出[11]要求对特定风险源具有控制地位的主体负有防止风险实现的作为义务。在算法场景中，AI 服务商作为基础模型的设计者、训练数据筛选者与参数调优者，对模型的竞争效果具有事实上的控制能力，理应负有算法合规“保证人义务”。引入生成式 AI 进行商业决策的经营者，则负有建立算法合规治理体系的义务，不能将 AI 视为即插即用的“魔盒”。若放任 AI 在黑箱中运行导致合谋，这一组织性疏忽本身即构成可归责的过错[12]。

4.2. 美国反托拉斯法对“便利化行为”的规制

美国《谢尔曼法》第一条的判例法在规制协同行为方面积累了丰富经验。在 Interstate Circuit, Inc. v. United States, 306 U.S. 208 (1939)案 [3]中，最高法院认定，电影发行商向影院发出的建议信函虽未构成明示协议，但因影院有理由知道其他影院收到相同建议并可能遵循，即构成协同行为的基础。在 American Tobacco Co. v. United States, 328 U.S. 781 (1946)案 [4]中，法院进一步指出，竞争者之间的信息交换若具有实质性特征，可作为协同行为的证据[13]。

这些判例对生成式 AI 场景具有类推价值。当 AI 向多个竞争者输出定价建议时，即使建议内容未明确提及“协调”，但若竞争者知晓其他竞争者接收了相同建议，且建议内容具有引导行为一致的效果，则该输出可被视为公开邀约。美国法对“便利化行为”的规制表明，反垄断法可以超越传统协议概念，对具有协调功能的信息传播行为进行规制。

4.3. 产品责任法中的“设计缺陷”归责逻辑

产品责任法为将归责视点从“行为”前移至“产品设计”提供了成熟范式。《美国侵权法重述(第三版)》第二条及欧盟《产品责任指令》[5]第六条确立了“设计缺陷”概念：当产品在设计上存在不合理危险，且该危险在按预期使用时造成损害，生产者应承担责任[14]。我国《中华人民共和国民法典》第一千二百零三条 [6]及《中华人民共和国产品质量法》第四十六条 [7]亦规定，产品存在“不合理危险”时生产者应承担侵权责任。

类推至生成式 AI 场景，基础模型的合谋倾向可被认定为一种“设计缺陷”。若模型在训练过程中被赋予“利润最大化”单一目标函数，且训练数据大量包含历史卡特尔案例，则模型可能习得协同策略作

[3]https://supreme.justia.com/cases/federal/us/306/208/
[4]https://supreme.justia.com/cases/federal/us/328/781/
[5]https://q2rt76fo46.feishu.cn/wiki/B03TwtUrViUVh6kmBescnAQxneb
[6]https://www.court.gov.cn/zixun/xiangqing/233181.html
[7]https://www.samr.gov.cn/zfjcj/tzgg/art/2023/art_579118cd202a45fba28b7edfd9f6fd72.html

为最优解。当这一缺陷在市场应用中实际导致了合谋损害时，要求作为风险源控制人的 AI 服务商承担相应责任，便具有法理上的可迁移性。这不是要求 AI 服务商为所有使用者的行为承担替代责任，而是要求其为自己所创造的、不合理的风险产品承担责任。

5. 从“主观合意”到“客观风险控制”

5.1. 意思联络要件的功能性重新阐释

面对生成式 AI 轴辐协议的挑战，需要对“意思联络”要件进行功能性的重新阐释，使其从一种主观心理事实，转变为几种可被客观识别和证明的法律状态。

第一，从明示合意到信号式联络。当经营者向 AI 输入一个在客观上创设了实质性合谋风险的信号时，该“信号发送”行为本身即满足意思联络的功能等价物。如输入“维护行业价格纪律”等模糊指令，因算法同质化效应，其被竞争者接收并响应的概率极高，具有与明示合意相当的市场效果。第二，从主观意图到风险认知。行为人对其输入指令可能引发协同效果具有认知可能性即可，不要求对具体协同策略的预见。经营者部署的 AI 系统应被推定为电子代理人，只要 AI 行为处于经营者概括授权的业务运营风险范围内，其法律后果即归属于经营者。采纳 AI 建议并付诸实施的行为，本身即构成责任连接点。第三，从主体间关系到共同风险体。多个经营者因共享算法架构形成技术共同体，构成“共同风险体”。明知或应知共享 AI 技术会带来显著市场协同风险，但为追求效率收益自愿接入，其共同但有意识的冒险可被评价为“准合意”。这借鉴了德国“组织过错”理论中，对建立和维持一个危险组织体之行为的归责逻辑。

这三种重新阐释，共同完成了将“意思联络”从主观内心世界解放出来、转向客观风险世界的关键一步。

5.2. “共同风险体”理论的严格化构建

“共同风险体”理论借鉴了德国社会法与保险法的概念资源，但在反垄断法语境中赋予新的内涵。其核心命题是：当多个经营者因技术架构的绑定而共享同一竞争风险源时，其对该风险源的共同依赖与共同控制，构成一种特殊的法律关系。为避免“连坐”风险，“共同风险体”的成立须满足“四要件 + 两限定”。

四要件：(1) 技术同质性，使用同一或实质性相似的生成式 AI 模型，共享核心参数权重或训练数据集；(2) 市场结构性，所在市场为寡头垄断或高度集中市场(HHI > 1800)，确保协同具有现实经济基础；(3) 认知关联性，经营者明知或应知其他竞争者使用同类算法，可通过行业报告、公开 API 接入信息或模型服务商宣传资料；(4) 信息汇聚性，存在数据回传、模型微调共享或中心化参数更新机制，构成事实上的信息交流通道。

两限定：(1) 排除独立开发且未共享训练数据的相似算法，避免巧合趋同被误伤；(2) 要求协同效果无法以独立理性决策解释，即排除市场结构本身导致的平行定价。

责任划分机制。各经营者承担按份责任为主、连带责任为辅。按份责任的划分依据：(1) 指令模糊程度，越模糊责任越重；(2) 对算法的实际控制力，能否调整参数、退出系统；(3) 从协同中获益比例。各经营者仅当存在直接指令或信息互通时方适用连带责任，而 AI 服务商则承担独立的“设计缺陷”责任。

5.3. “算法代理”理论与“可中断的因果关系”

“算法代理”理论为责任归属提供了另一路径。该理论主张，生成式 AI 在特定条件下可视为使用者的电子代理人，其在授权范围内的行为后果归属于被代理人。“算法代理”理论的适用边界需要严格限

定：其一，授权范围应以使用者的指令内容为判断基准，明确指令下的 AI 输出直接归属于使用者；模糊指令下的 AI 输出需考察使用者对 AI 自主决策能力的认知与接受。其二，越权行为的处理应借鉴表见代理规则：若 AI 输出超出指令范围，但使用者事后接受或未及时纠正，则视为对越权行为的追认。其三，AI 服务商的责任应独立于使用者责任：当模型设计缺陷导致协同策略生成时，AI 服务商作为代理人的选任者，应承担选任过失责任。

同时引入“可中断因果关系”理论实现责任的精细化划分。在“纯粹自涌现型”场景中，责任配置重心向 AI 服务商转移，使用者仅承担选任过失层面的轻微责任。使用者可通过以下行为切断因果关系与责任的联结：(1) 建立并运行合规有效的算法审查机制，例如设立独立的算法伦理委员会；(2) 对 AI 输出的高风险协同结论开展人工复核，并就拒绝采纳该结论的理由形成书面记录；(3) 在监测到价格异常趋同态势时，主动向反垄断执法机构履行报告义务。若使用者满足上述全部条件，则推定其已完全履行法定注意义务，无需对 AI 自主涌现的协同结果承担责任。

针对因果关系的认定，本文采用三层检验规则：(1) 条件关系检验，即判断案涉算法是否为协同行为发生的必要条件，若移除该算法后协同行为仍会发生，则不成立条件关联；(2) 相当因果关系检验，即判断该类算法在同类相关市场中是否通常会引发协同效果；(3) 因果关系中断检验，即判断使用者是否已采取符合规范要求的风险阻断措施。

5.4. 双层归责体系的构建

基于上述理论，本文提出“人类行为归责 + 算法风险归责”的双层归责体系。

第一层为人类行为归责，依意图分层配置责任梯度。“直接指令型”中，人类经营者承担核心责任，因其具有明确的协同意图，AI 仅作为工具中介。“模糊指令型”中，人类经营者承担过失责任，因其对 AI 可能生成协同策略的风险具有认知可能性，但未建立有效的合规审查机制，构成注意义务的违反。“纯粹自涌现型”中，人类经营者承担放任责任，因其明知使用共享算法存在协调风险，却未采取退出或调整措施，构成对风险的默示接受。

第二层为算法风险归责，依据风险源控制进行配置管理责任。AI 服务商作为基础模型的设计者，对模型的竞争效果具有事实上的控制能力，负有算法合规保证人义务。该义务的内容包括但不限于在模型训练阶段排除历史卡特尔数据、在模型部署阶段设置竞争合规过滤器、在模型运行阶段建立异常行为监测机制。违反上述义务导致协同策略生成的，AI 服务商应承担独立的管理责任，而非与使用者连带责任，这种责任配置避免了连坐效应，使技术提供者专注于合规改进而非规避使用。

基于上述新阐释，可构建一个覆盖人机两端、意图与风险分层的归责体系。这两层责任并行不悖，相互衔接，共同构成了一张完整覆盖从人类意图到算法风险的、密而不漏的追责之网。

6. 从理论到规则的转化

6.1. 证据规则的调适

新型合谋的证明困境，要求对传统的“谁主张、谁举证”规则进行适当的、有条件的调整，而非全盘放弃。本文主张引入“举证责任有条件倒置”机制，原告承担初始举证责任，证明算法协同的外观事实，这包括两项核心事实：其一，各经营者之间存在显著的价格同步化等协同行为外观；其二，这些经营者都使用了同一生成式 AI 服务或存在跨平台算法协同的技术基础。完成此证明后，举证责任即发生转移。由被告证明无风险创设，包括算法设计不存在协同诱导性、使用者未输入具有协调倾向的指令、已建立有效的合规审查机制等。

“算法审计报告”可作为安全港证据。若经营者定期委托独立第三方进行算法审计，审计报告证明

模型不存在协同倾向、使用过程中未触发协调机制，则可在诉讼中作为反驳证据。如果该报告能够证明其算法在设计、训练、输入等环节已尽到了防范合谋风险的充分义务，则可进入“安全港”，被推定为不构成违法。若无法提交此报告，或其报告未能通过审查，则将承受不利的法律后果。这一制度设计激励经营者主动合规，同时降低执法机关的信息成本。

这一设计将深不可测的“算法黑箱”争议，转化为对一份合规文档的审查与质证，既化解了执法机关的证明僵局，也为经营者提供了明确的合规激励与法律确定性。

6.2. 合规指引的三级制度设计

第一，建立沙盒测试准入机制。针对具备系统性影响的定价模型(服务用户规模 > 1000 或覆盖市场份额 >10%)，监管机构应当要求 AI 服务商开放沙盒测试环境，模拟典型市场结构(如三寡头市场、双寡头市场)下模型的定价决策过程，测算模型收敛至协同均衡的概率，若测试通过率低于预设阈值，则禁止该模型投入商用[15]。

第二，落实输出端风险提示义务。要求 AI 模型在输出定价建议时，若检测到以下特征，必须自动生成竞争风险提示：(1) 模型输出的建议价格与竞争对手当前定价高度趋同；(2) 模型输出建议涉及产量限制或客户市场分配；(3) 定价建议基于“市场跟随”策略生成。风险提示应当包含决策依据简报、输入数据摘要、推理逻辑摘要与风险等级四项核心内容。

第三，构建分级备案管理制度。按照模型对市场竞争的影响范围，将其划分为三个等级并设置差异化备案要求：一级(基础大模型)须事前向监管机构强制提交竞争影响评估报告；二级(垂直领域定价模型)实行年度备案结合抽查审计的管理模式；三级(企业内部自研模型)采用自主审计结合异常情况主动报告的管理要求[16]。

6.3. 法律责任的梯度配置

行政处罚应依据意图分层实现梯度化。直接指令型适用《中华人民共和国反垄断法》第五十六条的顶格处罚；模糊指令型适用中等幅度处罚，并责令建立合规机制；纯粹自涌现型适用较低幅度处罚，但要求退出共享算法或调整模型参数。民事赔偿应引入连带责任与补充责任的区分：使用者之间对协同造成的损害承担连带责任，AI 服务商在管理责任范围内承担补充责任。

将新归责体系中的意图分层转化为行政处罚的裁量阶梯。对直接指令型，适用接近法定罚款上限的重罚，相关责任人可被罚款并禁止从业。对模糊指令型，罚款根据注意义务违反程度居中确定。对采纳 AI 代理建议未审查的，罚款从轻，重在责令停止与合规整改。

人类经营者应作为第一责任人对全部损害承担连带责任。若 AI 基础模型服务商被认定存在“设计缺陷”，其在损害扩大部分承担补充责任。这激励受害人主要向经营者求偿，同时为 AI 服务商保留了适当的风险约束。

对算法本身的监管干预措施应包括算法禁令(禁止特定模型在特定领域的使用)、强制修改令(要求调整模型参数以消除协同倾向)、算法召回(对已部署模型进行全面审查与修正)。

7. 结论

生成式人工智能作为定价中枢，与跨平台算法协同的融合，催生了无需人类明示合意即可涌现协同定价的新型轴辐协议。传统意思联络要件在归责筛选、证据门槛与罪责匹配上均遭遇适应性危机，固守传统将制造“算法黑箱豁免”的责任真空。

本文以“智定价”虚拟案例为线索，解构了人机混合决策的三种形态，检验了传统教义学的三重失灵，通过借鉴德国“组织过错”与“保证人义务”理论、美国“便利化行为”规制经验及“设计缺陷”归

责逻辑，提出以“共同风险体”和“算法代理”为核心的双层归责理论。这一理论将意思联络从主观合意转向客观风险创设与控制，在不突破责任主义原则的前提下实现对 AI 自涌现合谋的有效规制。通过严格化共同风险体成立标准，引入“可中断的因果关系”精细划分责任，设计三级合规制度与相关证据规则，为反垄断法回应自主智能系统提供了从理论到制度的完整路径。

需要承认的是，本文的理论建构仍存在局限。对算法透明度的依赖使制度效果受制于技术审计能力的发展；双层归责体系在不同法系的兼容性尚需进一步检验。当人工通用智能时代到来，AI 具备自主意识时法律主体地位是否应当重构，已超出本文范围，留待未来研究探索。

参考文献

[1] Ezrachi, A. and Stucke, M.E. (2017) Artificial Intelligence & Collusion: When Computers Inhibit Competition. *University of Illinois Law Review*, **2017**, 1775-1810.

[2] Ezrachi, A. and Stucke, M.E. (2016) Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy. Harvard University Press.

[3] Schwalbe, U. (2018) Algorithms, Machine Learning, and Collusion. *Journal of Competition Law & Economics*, **14**, 568-607. https://doi.org/10.1093/joclec/nhz004

[4] OECD (2017) Algorithms and Collusion: Competition Policy in the Digital Age. OECD Publishing.

[5] 朱凯. 论反垄断法上协同行为的认定[J]. 研究生法学, 2010, 25(2): 92-98.

[6] 江山. 反垄断法上协同行为的规范认定[J]. 法商研究, 2021, 38(5): 187-200.

[7] 王晓晔. 反垄断法[M]. 北京: 法律出版社, 2011.

[8] Mehra, S.K. (2016) Antitrust and the Robo-Seller: Competition in the Time of Algorithms. *Minnesota Law Review*, **100**, 1323-1348. https://doi.org/10.24926/265535.1104

[9] Roxin, C. (2003) Strafrecht Allgemeiner Teil, Band II. §25 Rn.124 ff. (Organisationsherrschaft), Verlag C.H. Beck.

[10] Tiedemann, K. (1976) Wirtschaftsstrafrecht und wirtschaftskriminalität. Rowohlt Verlag.

[11] Jakobs, G. (1991) Strafrecht Allgemeiner Teil, 2. Aufl.1991, §29 Rn.3 ff. (Garantenpflicht), Walter De Gruyter.

[12] 黎宏. 合规计划与企业刑事责任[J]. 法学杂志, 2019, 40(9): 9-19.

[13] Page, W.H. (2017) Tacit Agreement under Section 1 of the Sherman Act. *Antitrust Law Journal*, **81**, 593-640.

[14] Directive 85/374/EEC of 25 July 1985 on the Approximation of the Laws, Regulations and Administrative Provisions of the Member States Concerning Liability for Defective Products, Art. 6. https://eur-lex.europa.eu/eli/dir/1985/374/oj/eng

[15] 叶明, 郭江兰. 数字经济时代算法价格歧视行为的法律规制[J]. 价格月刊, 2020(3): 33-40.

[16] 张凌寒. 算法权力的兴起、异化及法律规制[J]. 法商研究, 2019, 36(4): 63-75.