“AI代写”收费场景下的法商边界厘定——以文心一言为例

高泺云[1]，朱珂怡[2]，朱珂[3]
1. 华东政法大学知识产权学院，上海；
2. 华东政法大学经济法学院，上海；
3. 华东政法大学商学院，上海

摘要：本文以百度“文心一言”为例，系统探讨了AI写作服务在收费场景下的法律与商业边界问题。随着AI写作行业爆发式增长，免费与收费并存的商业模式引发了学术代写、版权争议、虚假信息等多重合规风险。文章通过案例比较、文本分析与合规评估等方法，揭示了平台在用户协议中普遍存在的“权利让渡+风险转嫁”逻辑，并对比了中欧在AI监管路径上的差异。在此基础上，提出构建法律责任分级模型、全流程合规体系与平台治理创新方案，推动AI写作平台从“工具提供者”向“生态治理者”转型，以实现技术创新与法律合规的深度融合。

关键词：AI 写作服务；法律合规；法商融合；平台治理；商业模式

Delimiting the Legal and Commercial Boundaries in Fee-Based Scenarios of AI—Generated Content: A Case Study of ERNIE Bot

Luoyun Gao[1], Keyi Zhu[2], Zhu Ke[3]
1.Intellectual Property School, East China University of Political Science and Law, Shanghai;
2.Economic Law School, East China University of Political Science and Law, Shanghai;
3.School of Business, East China University of Political Science and Law, Shanghai

Abstract: This paper takes Baidu’s “ERNIE Bot” as a case study to systematically explore the legal and commercial boundaries of AI writing services in paid scenarios. With the explosive growth of the AI writing industry, the coexistence of free and paid business models has triggered multiple compliance risks, such as academic ghostwriting, copyright disputes, and the spread of misinformation. Through methods such as case comparisons, textual analysis, and compliance assessments, the paper reveals the prevalent logic of “rights transfer + risk shifting” in platform user agreements and contrasts the differences in AI regulatory approaches between China and Europe. Building on this analysis, the study proposes a hierarchical model of legal liability, a full-process compliance system, and innovative platform governance solutions. These recommendations aim to facilitate the transition of AI writing platforms from

* 作者简介：高泺云，华东政法大学知识产权学院硕士研究生、中国法治化营商环境研究院研究人员；朱珂怡，华东政法大学经济法学院硕士研究生；朱珂，华东政法大学商学院硕士研究生、华东政法大学丝路电商法治研究中心研究人员。

mere “tool providers” to “ecosystem governors,” thereby achieving deeper integration of technological innovation and legal compliance.

Keywords: AI Writing Services; Legal Compliance; Integration of Law and Business; Platform Governance; Business Models

1 案例简介

近年来，AI写作服务行业迎来了爆发式增长，市场热度与用户需求均呈现出前所未有的高涨态势，成为当下数字经济领域的重要增长点。然而，免费与收费并行的商业模式、学术代写、虚假广告、版权侵权等合规争议同步爆发[1]。如何平衡“免费扩张—商业变现—法律合规”成为行业痛点，亦为法商治理研究提供了鲜活样本。在此背景下，文心一言凭借其突出的行业表现，具备极强的典型性。它不仅在短时间内实现了用户规模的快速跃升，更在技术普惠层面，通过降低使用门槛让更多普通用户和中小机构享受到AI写作的便利；在商业可持续方面，探索出了适配自身发展的路径；在法律合规领域，也走在行业实践的前沿，其应对各类合规问题的举措具有参考价值。正因如此，文心一言能够为AI写作产品的商业化推进以及行业内法商融合生态的构建提供切实可行的范本，所以选择其作为案例进行深入剖析。

案例内容方面，首先系统梳理了AI写作行业的爆发动因与现状，包括用户规模激增、技术普惠趋势及主流平台的差异化商业模式，重点解构了文心一言、豆包等平台的用户协议条款，揭示普遍存在的“权利让渡+风险转嫁”逻辑及其在收费场景下的权责失衡问题。进而聚焦于收费服务中突出的法律风险，如学术代写引发的诚信危机、虚假信息生成带来的事实核查困境、训练数据版权合法性存疑、生成内容确权规则模糊，以及平台在专业服务中面临的资质缺失与职业替代争议。案例还对比了中国“渐进式治理”与欧盟“强监管合规”两种监管路径，指出国内平台在跨境服务中面临的合规挑战。

分析过程遵循“现状梳理—风险识别—治理设计—优化方案”的逻辑主线，综合运用案例比较法、文本分析法、合规评估法与商业模式解构方法。通过横向对比文心一言、DeepSeek、Kimi等平台的服务协议与盈利机制，揭示免费与收费模式下责任条款冲突的深层原因；借助司法判例与政策文本，评估平台在当前法律框架下的责任范围与免责可能性；最终从平台治理与商业模式双维度提出系统解决方案：治理层面构建风险防控体系；商业层面设计基于用户分层的变现机制、责任强化的服务增值路径与数据资产协同壁垒，推动平台由工具提供者向生态治理者转型。

案例强调，在AI写作迈向深度收费时代的背景下，合规能力已从成本项转变为核心竞争力，平台需通过技术赋能、规则重构与利益平衡，将外部法律要求内化为商业模式的核心组成部分，从而构建既鼓励创新又防控风险的可持续发展生态。

2 问题的提出：AI写作服务的崛起与文心一言的战略转型

2.1 行业爆发与免费化浪潮

2.1.1 用户规模激增与技术普惠

2025年，AI写作服务行业迎来爆发式增长。QuestMobile数据显示，豆包MAU达1.15亿，星火AI凭借教育场景深度整合实现DAU2700万，Kimi平台则通过按量计费模式吸引中小企业用户。这些数据背后，是生成式AI技术从“工具”向“基础设施”的转型。文心一言通过开源策略和成本优化，免费开放ERNIE系列模型免费开放及将推理成本下降70%，推动服务全面免费化，用户规模在半年内

从2亿跃升至3亿，印证了“技术普惠”对市场渗透的加速作用。北京某高校学生利用文心一言免费版完成毕业论文初稿，通过调整关键词和参数，在三天内生成超过200页的学术文本，这一案例被《中国青年报》报道，引发社会对AI辅助写作效率的广泛讨论。

2.1.2 免费模式的商业逻辑重构

免费化并非单纯获客手段，而是构建生态壁垒的关键。各平台免费化策略呈现差异化路径：豆包免费版每日限5次文档处理，相当于一本《三体》的文本量，适合学生群体；星火AI通过作业批改等基础功能免费吸引教育用户，再以竞赛解答等深度推理服务引导付费，2025年Q2教育行业ARPU值达8200元；Kimi平台则采用“基础免费+API按量计费”，输入价格4元/百万tokens，输出16元/百万tokens，电商旺季商品描述生成成本可控制千元内。百度财报显示，文心一言通过“开源引流+云服务变现”模式，带动智能云业务同比增长26%。企业客户ARPU值（平均收入）显著提升，例如千帆平台接入中国石化、上海交大等8大行业客户，提供算力套餐和定制化模型服务。以长安汽车为例，其研发部门通过文心一言的代码生成功能，将新车机系统开发周期缩短40%，相关成本节约超1200万元。这种“基础服务免费+增值服务收费”的分层变现体系，既降低了用户准入门槛，又通过高价值客户实现商业可持续性。

2.1.3 政策监管与行业规范并行

行业爆发伴随监管升级。国家网信办《AI生成内容标识办法》要求所有生成内容必须添加显式标识和隐式标识，武汉首例AI图片侵权案判决进一步明确：AI生成内容若具备独创性表达，可受著作权法保护。2025年3月，北京互联网法院审理某自媒体使用AI生成财经报告案时，首次要求平台提交内容生成全流程日志，包括用户输入关键词、模型版本号及生成时间戳，这一判例推动行业建立“创作可追溯”机制。政策与司法的双重约束，推动行业从“野蛮生长”转向“合规创新”，为免费化浪潮划定法律边界。

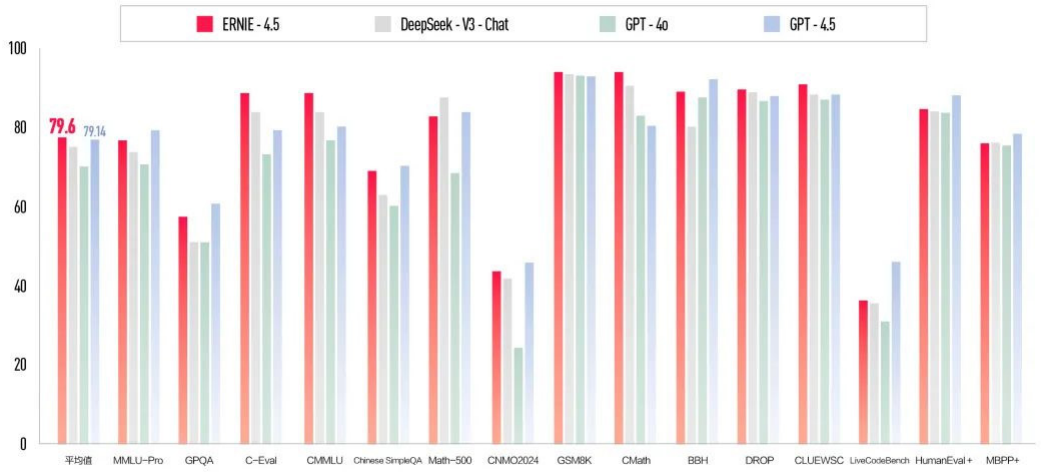
2.2 文心一言的定位演进

2025年3月16日，百度发布文心大模型4.5和X1，并在文心一言官网免费向用户开放，企业用户和开发者可通过百度智能云千帆大模型平台便捷调用文心4.5API。旗下产品矩阵也将陆续接入两款新型大模型，3月17日，搜索智能助手文小言已接入两款模型及DeepSeek-R1满血版，支持多种模型自动调度。如图1所示，文心4.5为自研原生多模态基础大模型，多项基准测试结果优于GPT4.5和DeepSeek-V3。

原生多模态能力：具备对文本、图像、音视频等混合数据的综合处理能力，语言能力包括理解、生成、逻辑和记忆显著增强，尤其是去幻觉、逻辑推理以及代码能力。例如能够综合理解图片中的文字/表格，提取重点并给予分析，对网络梗图能进行理解和逻辑解释。如图2、图3所示，作为能力更全面的深度思考模型，模型兼备准确、创意和文采，在中文知识问答、文学创作、文稿写作、日常对话、逻辑推理、复杂计算及工具调用上表现尤为出色。

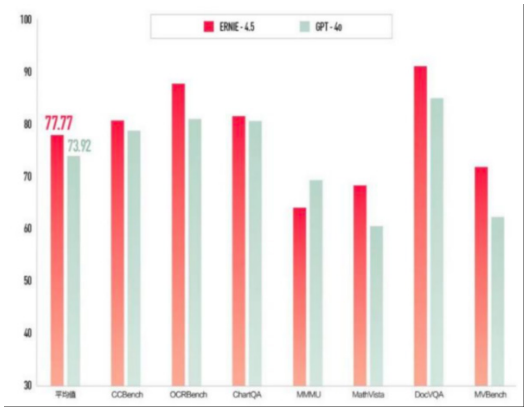
技术升级：1）FlashMask动态注意力掩码：提升长文本处理能力和多轮对话上下文理解交互表现。2）多模态异构专家扩展：平衡不同模态梯度贡献，解决训练梯度冲突问题，提升多模态融合能力。3）时空维度表征压缩：降低图片和视频的计算复杂度，提升长视频语义提取及多模态数据训练效率。4）基于知识点的大规模数据构建：通过知识分层采样、跨模态压缩融合及定向合成技术提升模型知识密度，降低模型幻觉。5）基于自反馈的Post-training：自反馈迭代提升学习系统稳定性。

成本优势：如表1所示，文心4.5API调用输入价格0.004元/千tokens，输出价格0.016元/千tokens，仅为GPT4.5的不到1%。此外，公司开源战略持续推进，计划于6月30日开源文心大模型，开发者可进行定制化开发和应用。



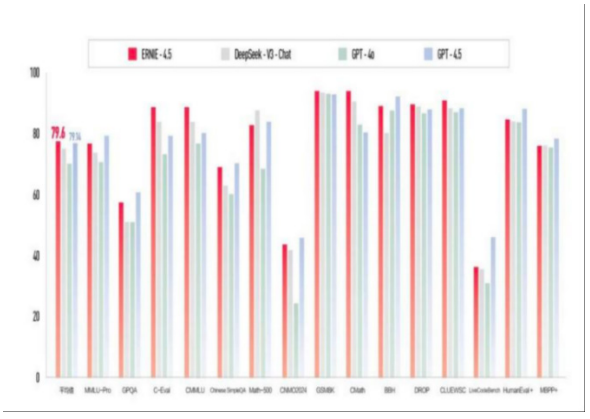
资料来源：百度，交银国际

图1. 文心 4.5 多项基准测试结果优于 GPT4.5 和 DeepSeek-V3



数据来源：百度公众号，华龙证券研究所

图2. 文心大模型 4.5 多模态能力



数据来源：百度公众号，华龙证券研究所

图3. 文心大模型 4.5 文本能力

表1. 文心 4.5 和 X1、GPT 4.5 价格对比

	文心 4.5	文心 X1	GPT 4.5
输入	0.004 元 / 千 tokens	0.002 元 / 千 tokens	0.540 元 / 千 tokens
输出	0.016 元 / 千 tokens	0.008 元 / 千 tokens	0.108 元 / 千 tokens
触发搜索	0.004 元 / 次	NA	NA

资料来源：百度，公开资料，交银国际 *汇率按 7.2 计算

2.2.1 技术开源：从封闭到开放的生态重构

文心一言的战略转型始于技术架构的全面开源。2025年，百度发布文心大模型4.5Turbo，并开源飞桨框架3.0，支持全球开发者自由迭代。其ERNIE-4.5-0.3B轻量模型以3亿参数实现中文场景推理精度达92%，单卡显存占用低至2.1GB，使中小企业首次拥有私有化大模型的可行性。这种开放策

略不仅降低模型训练成本，更通过千帆平台接入上百主流模型，形成“MaaS”生态。例如，中国建材集团利用百度AIPaaS构建智能物流系统，实现全国3000个仓库的动态调拨，年运输成本降低18%。开源策略还催生“模型黑市”现象，部分开发者将定制模型以高价转售，百度随即推出“模型认证”计划，通过数字签名验证模型来源，维护生态健康。

2.2.2 分层变现：C端流量与B端价值的双轮驱动

定位演进的核心在于商业化路径的多元化。C端通过文库AI和深度搜索功能，构建“创作+分发”一体化平台；B端则推出企业级知识增强工具链，如代码生成、多模态检索，服务长安汽车、地平线等龙头企业。财报显示，2025年Q2大模型调用量同比增长33倍，商业化转化效率显著提升，印证“技术-产品-市场”的闭环成型。文心一言还创新“基础费+效果费”混合定价模型，企业客户ARPU值达12.8万元/年。

2.2.3 法律合规：平台责任与用户义务的嵌入

战略转型中，文心一言将法律合规深度融入产品设计。用户协议明确标识义务，并通过技术手段确保可追溯性。2025年4月，百度联合中国法学会发布《AI写作服务合规指南》，首次将“创作链条可视化”纳入行业标准，要求平台记录用户调校行为，建立版权证据链。在司法实践中，武汉中院判例要求AI生成内容需保留创作过程证据，文心一言通过服务协议和工具链，帮助用户满足法律要求。2025年4月，百度联合中国法学会发布《AI写作服务合规指南》，首次将“创作链条可视化”纳入行业标准，要求平台记录用户2000次以上调校行为，建立可追溯的版权证据链。

AI写作服务的崛起，是技术进步、商业模式创新与政策规范共同作用的结果。文心 X1 为首个自主运用工具的深度思考模型，基于关键技术包括递进式强化学习训练方法、基于思维链和行动链的端到端训练和多元统一的奖励系统，在知识问答、文学创作、逻辑推理等方面表现优异，增加多模态支持，并能进行工具调用如 AI 绘图、代码解释器、网页链接读取等。文心一言通过免费化扩大用户基础，以开源生态重构行业格局，最终在法律框架内实现商业价值与社会责任的平衡。其技术开源、分层变现、法律合规的三重定位，不仅为AI写作产品商业化提供主流路径，更为法商融合的可持续生态构建了范本。选择文心一言作为案例，正因其代表

表2. 文心一言迭代情况

文心一言（百度）	推出时间	简介
ERNIE 1.0	2019 年 3 月	知识增强的文心大模型 ERNIE 1.0，文心一言的早期版本之一。
ERNIE 2.0	2019 年 7 月	学习大规模语料中的词法、语法、语义等知识。
ERNIE 3.0	2021 年 7 月	发布文心大模型 3.0（ERNIE 3.0），首次在千亿级预训练模型中引入大规模知识图谱。
文心一言 3.0	2023 年 2 月	官宣新一代大语言模型文心一言（英文名：ERNIE Bot）。
文心一言 3.5	2023 年 6 月	新增了插件机制，通过插件方式扩增了大模型的能力边界。
文心一言 4.0	2023 年 10 月	理解、生成、逻辑、记忆四大能力提升，重构了原来的搜索模式。
文心一言 4.0 Turbo	2024 年 6 月	上下文输入从 4.0 版的 2K tokens 升级到了 128K tokens；AI 生图分辨率提升，从 512x512 提升至 1024x1024；智能体技术，包括理解、规划、反思和进化。
文心一言 App 4.0.0 版本	2024 年 9 月	文心一言 App 升级为“文小言”，提供了问问题、陪聊天、写文章、画图片和下任务五大核心场景能力。
文心大模型 4.5	2025 年 3 月	首个原生多模态大模型，API 调用价格仅为 GPT4.5 的 1%。
文心大模型 X1	2025 年 3 月	性能上对标 DeepSeek-R1 的深度思考模型，同时还支持多模态、多工具调用能力，API 调用价格约为 R1 的一半。
文心大模型 4.5 Turbo	2025 年 4 月	文心 4.5 Turbo 的多项基准测试成绩显著优于 GPT 4o，平均分达到 77.68，超过 GPT 4o 的 72.76。而对比文心 4.5，文心 4.5 Turbo 速度更快、价格下降 80%，每百万 token 的输入价格仅为 0.8 元，输出价格 3.2 元，仅为 DeepSeek-V3 的 40%。
文心大模型 X1 Turbo	2025 年 4 月	文心 X1 Turbo 相较文心 X1，具备更长的思维链和更强的深度思考能力，同时进一步增强了多模态和工具调用能力，在性能提升的同时价格再降 50%，每百万 token 输入价格 1 元，输出价格 4 元。

了行业在技术普惠、商业可持续、法律合规三方面的前沿实践。以下表2呈现了文心一言的迭代情况。

3 AI写作平台生态实证：现状、逻辑与协议解构

图4具体描述了法律合规性基础下的商业模式探索与比较，包括平台服务协议、行业监管政策、典型平台案例、盈利模型及用户需求分析，并进一步结合公司状况与战略定位，聚焦变现模式与用户分层，突出文心一言在商业逻辑与服务协议方面的深度解析。

3.1 法律合规性基础

3.1.1 平台服务协议关键条款对比分析

通过对文心一言、DeepSeek、Kimi、豆包四大主流平台的用户协议进行文本解构（详见表2），可观察到平台方在应对法律风险时形成了一套趋同的条款设计逻辑。这种设计在免费服务模式下具有一定合理性，但在收费场景中暴露出权责配置的结构性失衡[2]。

（1）平台普遍采用“权利让渡+风险转嫁”的双轨策略确定版权归属

以文心一言协议第5.3-5.4条为例，其承认用户对生成内容享有权利，但要求用户承诺输入内容不侵犯第三方知识产权，否则“承担全部侵权责任”。类似地，豆包协议第8.3条要求用户确保上传内容合法，却未对输出内容的版权瑕疵提供救济机制。这种条款实质上将《著作权法》第11条规定的“作品创作主体责任”转移至用户，而平台作为技术提供方规避了内容创作过程中的版权审查义务。当用户付费购买“论文生成”“商业文案撰写”等专项服务时，权责不对等矛盾尤为突出——平台收取服务费用，却未相应提升版权保障水平。

（2）责任豁免条款的扩张趋势构成潜在法律冲突

各平台均以加粗、下划线等醒目格式声明输出内容“仅供参考”“不构成专业建议”（如DeepSeek第4.4条、Kimi第1.8条）。此类条款在免费场景中符合《消费者权益保护法》第26条对格式条款的提示要求，但在付费模式下可能因显失公平而无效。例如豆包向企业用户提供收费的“合同生成”服务时，仍条款第2.2.1条坚持其输出“不可作为法律文件依据”，这与《民法典》第498条“格式条款解释不利于提供方”的原则相抵触。司法实践中已有类似判例，2024年杭州互联网法院“AI文案服务合同纠纷案”中法院认定收费服务中的全面免责条款无效。

（3）内容合规机制存在执行缝隙

尽管所有平台承诺建立审核机制，如文心一言第4.3条的人工+算法审查，但协议未明确收费服务的审核标准是否升级。例如对会员用户提交的“学术论文代写”指令，平台仅泛泛禁止“违反学术诚信”，如文心一言第4.4.1条规定，却未配备针对性的内容过滤器。这显然不符合《生成式人工智能服务管理暂行办法》第12条“提供者应对生成内容依法承担责任”的监管要求。更值得关注的是，平台投诉机制均为事后救济，如Kimi要求侵权者“自行提交不侵权证明”，无法防范付费用户即时获取违规内容的风险。

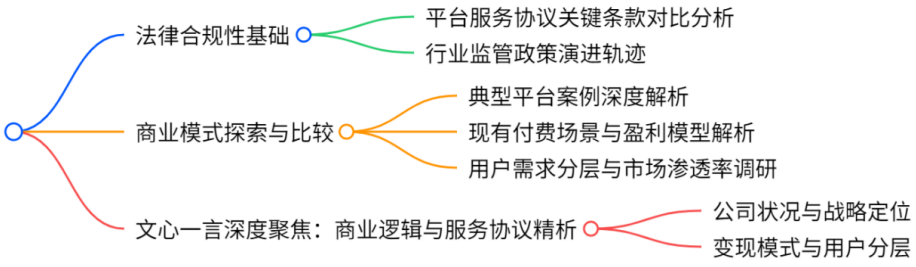


图4. “AI写作平台生态实证：现状、逻辑与协议解构”思维导图

（4）商业化条款揭示责任真空

表3梳理了目前主流AI写作平台用户的协议关键条款对比，可见当前协议普遍回避收费服务的特殊责任。典型如文心一言第7.1条保留“免费服务未来收费权利”，豆包通过API接口向企业收取调用费用，但两者均未约定付费后的内容质量担保义务。这种商业实践与法律义务的错位，折射出平台在商业模式进化中的策略矛盾——既试图通过收费扩张盈利渠道，又延续免费时代的责任规避立场。

3.1.2 行业监管政策演进轨迹

人工智能写作服务的监管框架在全球范围内呈现显著分化，中国与欧盟分别形成“渐进式治理”与“强监管合规”两条典型路径[3]。这一演进过程不仅折射出技术伦理认知的差异，更深刻影响着平台责任的法商边界界定。

（1）中国：从“包容审慎”到“精准溯源”的渐进式治理

中国监管政策历经从伦理倡导到技术溯源的阶梯式发展。2021年《新一代人工智能伦理规范》首次提出“可控发展”原则，但仅作为非强制性指导文件存在，其监管焦点停留在研发侧风险预防。真正的转折点出现在2023年《生成式人工智能服务管理暂行办法》，这部全球首部专门规章标志着监管重心从“风险防御”向“创新促进”的战略转向。该办法在明确平台对生成内容承担主体责任的同时，创造性豁免了研发机构责任，为技术迭代保留弹性空间。值得注意的是，其提出的“分类分级监管”原则因缺乏实施细则，一度依赖行业自律填补——如百度文心一言自主建立4000条语料动态抽检机制，使内容合规合格率提升至96%以上。至2025年《人工智能生成合成内容标识办法》出台，监管进入技术溯源新阶段，强制要求添加显性与隐形双标识，推动平台责任从被动响应向主动防控转型。北京、上海等地同步试点的“监管沙盒”机制，更在金融合同、医疗文书等付费写作场景探索责任豁免路径，彰显出“发展与安全均衡”的治理智慧。

（2）欧盟：从“风险禁令”到“全周期合规”的强监管框架

欧盟则构建了以风险控制为核心的全周期监管

表3. 主流AI写作平台用户协议关键条款对比

分析维度	文心一言	DeepSeek	Kimi	豆包
版权归属	•用户享有生成内容权利（§5.4） • 用户承担输入侵权赔偿（§5.3）	•输出内容权利归属用户（§4.2） • 不排除内容雷同可能（§4.5）	•用户对输出内容负责（§五.3） • 雷同内容不触发权利让渡（§五.3）	•用户拥有上传内容版权（§8.1） • 输出不视为平台作品（§8.1）
责任限制	•禁用专业领域决策（§10.4） • 不保证内容合法性（§10.3）	•否认服务连续性承诺（§7.2） • 商业用途风险由用户自担（§4.4）	•禁止用于医疗/法律决策（§一.8） • 第三方内容责任豁免（§七.1）	•需二次验证专业内容（§2.2.1） • 使用后果完全自负（§2.2.3）
合规管理	•禁止诱导违法输出（§4.4.1） • 人工复审机制（§4.3）	•强制标注AI生成标识（§4.6） • 违法内容事后删除（§6）	•输入内容脱敏要求（§四.2） • 禁止篡改AI标识（§四.4.(3)）	•动态内容监控（§5.1.7） • 用户举报响应机制（§5.4）
用户约束	•输入内容合法性担保（§5.3） • 禁止技术反编译（§2).4.4.4)）	•授权关闭训练数据使用（§4.3） • 禁止删除水印（§4.6）	•禁止越狱攻击（§四.4） • 商业秘密保护义务（§四.2）	•AI云盘内容自主负责（§2.4） • 禁用自动化爬虫（§5.1.2）
收费条款	• 保留任意收费权利（§7.1） •未区分付费责任	•未明示收费方案 • 单方变更服务权限（§9.1）	•未约定收费条款	• 会员分级收费（实务操作） •API调用计费
核心法律争议	收费服务中的版权担保缺位	数据训练默示授权的合规性质疑	雷同内容版权冲突的归责困境	付费场景下“仅供参考”条款的效力矛盾

体系。2021年《人工智能法案》提案首创“四阶风险模型”，将AI写作划归“有限风险”类别，要求强制透明披露。随着ChatGPT引发社会冲击，2023年修订案新增针对通用人工智能模型（GPAI）的严格条款，对参数规模超10^{23}FLOP或多模态模型施加特殊义务。2025年生效的最终法案确立三重刚性约束：训练数据需完成版权清算并公示来源、生成内容必须标注AI属性、算法决策逻辑框架应向社会公开。尤为严厉的是其惩罚机制——违规企业面临全球营业额7%或3500万欧元的罚款，倒逼平台重构责任链条。这种“权利保障优先”的治理范式，在2024年引入监管沙盒豁免中小企业测试责任后，逐渐转向“风险控制与技术激励并行”，反映出欧盟在维护伦理底线与促进产业创新间的艰难平衡。

（3）监管范式转变的核心特征

中欧监管路径虽呈现理念差异，却在关键领域形成治理共识。一方面，生成内容标识成为全球监管基线，中国《标识办法》与欧盟GPAI条款均要求通过数字水印、区块链存证等技术实现全链路溯源；另一方面，责任主体从单一平台向生态协同扩展，中国要求网络服务提供者、研发者与用户共担标识义务，欧盟则建立覆盖开发者、部署者与进口商的链式责任体系。这种趋同背后是共有的治理困境：当AI写作从免费工具转向付费服务，传统“技术中立”原则难以化解版权归属模糊、内容质量失控等新型风险。

政策演进正深刻重塑平台的法商边界。中国“包容审慎”框架下，免费服务仍适用过错责任原则，平台可通过事后删除履行义务；但对于会员订阅、API商用等收费场景，《生成式AI服务管理暂行办法》第12条隐含的“结果责任”导向，要求平台建立分级审核机制。欧盟更通过GPAI条款将训练数据版权审查设定为收费服务的前置条件，直接冲击文心一言、Kimi等平台的语料供应链。监管差异还催生企业合规困境，百度需同时满足欧盟算法透明度披露与中国语料抽检双重要求，DeepSeek则因数据训练“默示授权”条款面临欧洲用户集体诉讼。这些冲突揭示出监管演进的核心命题：当AI写作跨越免费阈值进入商业化深水区，平台责任必须从风险规避转向责任分级，方能在技术创新与法律合规间构建可持续的法商生态。

3.2 商业模式探索与比较

3.2.1 典型平台案例深度解析

目前市场上主流的“AI代写”平台在商业模式上呈现出多样化的特点。从图5可得，不同参与者如“教育科技机构、教育信息化厂商、AI技术提供商、跨界互联网公司”等，在“生成式写作/AI代写”赛道的技术路径、盈利方式与合规风险差异巨大。

来源：2025年AI赋能教育行业发展趋势报告

图5. AI代写市场格局分类

以文心一言为例，其依托百度强大的技术实力和海量数据资源，在文本生成技术上具有显著优势。根据百度官方公布的数据，文心一言技术路径为“通用大模型+搜索增强”。其“作文生成+知识检索联动”可关联历年范文及时政热点，在训练过程中使用了超过万亿字节的文本数据，涵盖了多个领域和行业，这使得其生成的文本质量较高，能够满足不同用户的需求。

在收费模式上，我们通过研究百度宣布全面放开使用前的商业数据发现，文心一言收入结构=API调用45%+智能体分成30%+政企项目15%+硬件捆绑10%，此外，其采用了订阅制和按次收费相结合的方式。订阅制为用户提供了一定期限内的无限次使用权限，价格根据订阅时长不同而有所差异，例如月度订阅价格为50元，年度订阅价格为500元，享受16.7%的折扣优惠。按次收费则根据用户生成文本的字数或复杂程度进行计费，每千字收费5元。这种多元的收入组合可以有效平衡初期市场发展阶段可能面临的风险。

豆包作为字节推出的另一款知名的“AI代写”工具，其商业模式侧重于与教育机构的合作。据多鲸资本市场调研的数据显示，豆包在教育领域的市场份额达到了20%。它与众多学校和教育培训机构建立合作关系，为教师和学生提供定制化的写作辅助服务。在收费方面，主要采用按学生人数收费的模式，某学校或机构根据使用学生数量向豆包支付费用，每个学生每年的费用约为80元。

星火平台则借助科大讯飞深耕多年的资源，着重打造开放式的创作生态，鼓励开发者基于其平台开发各种写作应用和插件。根据星火官方发布的开发者报告，目前已有超过5000个开发者在其平台上注册，开发了2000多个写作相关的应用和插件。星火通过收取开发者平台使用费和交易分成来获取收益，开发者使用平台开发工具和资源的费用为每月100元，同时星火会从开发者应用的交易收入中抽取20%作为分成。

Kimi平台以其简洁易用的界面和高效的文本生成速度受到用户青睐。在市场推广方面，Kimi采用了社交媒体营销和口碑传播相结合的方式。根据社交媒体监测工具的数据，Kimi在微博、微信等平台上的相关话题讨论量每月超过10万次。其收费模式较为灵活，除了常见的订阅制和按次收费外，还推出了积分制度，用户可以通过完成任务或邀请好友获得积分，积分可用于兑换写作服务，这种模式吸引了大量年轻用户，目前其用户中年龄在18-30岁的占比达到了60%。

以下表4为四类参与者在生成式写作/AI代写领域的商业模式对比。

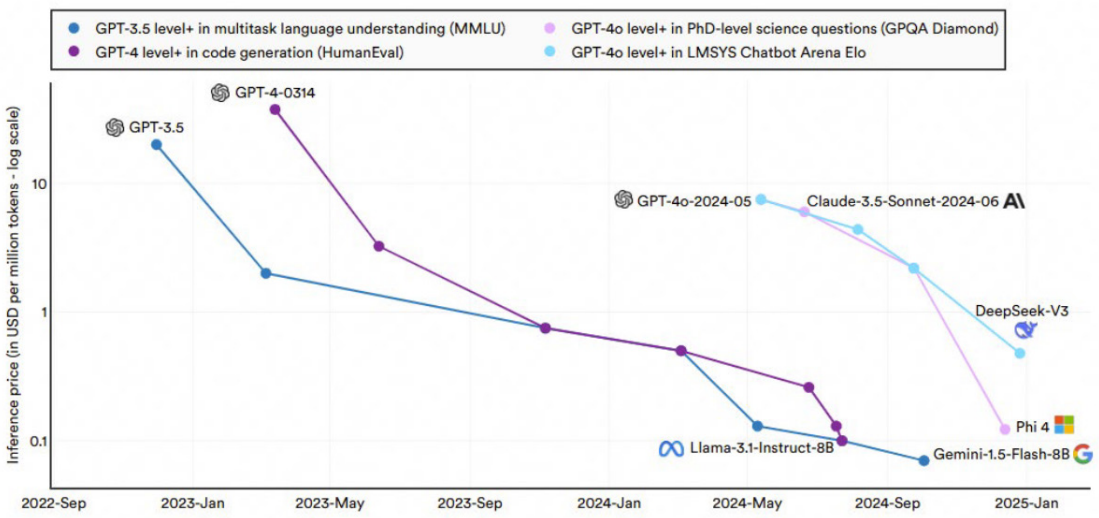
表4. 四类参与者在生成式写作/AI代写领域的商业模式对比

参与者类型	代表企业	技术路径特点	商业模式	优势	局限
教育科技机构	作业帮 猿辅导 网易有道	基于垂直领域数据（题库、作文语料）训练专用模型，聚焦学科场景（如作文批改）	C端订阅（会员功能，年费198-398元） 硬件捆绑（学习机、词典笔预装功能） B端校合作（按学校规模收授权费）	教育场景理解深，数据精准，用户付费习惯成熟	技术通用性弱，跨学科覆盖有限
教育信息化厂商	科大讯飞 希沃 鸿合	结合教育硬件（交互平板）开发场景化工具（如AI备课、智能阅卷），对接公立校需求	政府采购（省市校三级招标，单项目数千万） 硬件+软件捆绑（平板预装功能，单价2万元/台） 运维服务（按年收费）	进校渠道稳固，政策支持度高，数据合规性强	决策周期长，依赖政府预算，C端触达弱
AI技术提供商	OpenAI Anthropic HuggingFace	提供通用大模型API接口，支持多场景适配，注重伦理约束（如防学术不端）	技术授权（按token用量收费，如GPT-4约0.03美元/千token） 企业定制（高校年服务费超10万美元） 开源社区服务（模型微调收费）	技术通用性强，生态整合能力突出	教育场景数据不足，依赖合作伙伴
跨界互联网公司	百度 字节跳动 腾讯	整合流量数据（搜索、短视频、社交），开发多模态工具（如写作+知识检索联动）	广告变现（工具引流至平台，广告收入占比超60%） 硬件销售（学习平板、智能设备预装，售价2000-3000元） 云服务（SaaS平台年费50万元/校）	流量入口大，生态协同强，商业化路径多元	教育场景专业性不足，需外部合作补充

现如今，随着开源模型的大幅接入，高性能、低成本的大模型已大大提升了AI的商业化进程。

一方面，训练和推理工程持续优化，降低模型成本。如DeepSeekV3单次训练的硬件成本约600万美元，仅为GPT-4单次训练成本的1/10。在MMLU基准测试中达到GPT-3.5水平（MMLU准确率64.8%）的AI模型调用成本，已从2022年11月的20美元/每百万token，骤降至2024年10月的0.07美元/每百万token，18个月内AI成本下降99.6%。

另一方面，模型压缩与知识蒸馏技术的协同发展有效提升了轻量化模型的性能，显著增强了边缘计算场景下的端侧部署能力。如图6所示，以在MMLU上评测得分超过60%的模型参数为例，2022-2024两年内尺寸减少为之前的1/142。主要人工智能开发公司都发布了紧凑且性能高的模型，如GPT-40mini、o1-mini、Gemini2.0Flash、Llama3.18B和



数据来源：2025年AI指数报告

图6. 特定基准测试水平的推理价格变化（2022-2024年）

数据来源：wind数据库，5401家2024年A股上市公司年度报告，阿里云研究院分析整理

图7. 按TF-IDF词频分析的A股上市公司人工智能关注度梯队分布

qwen3-8b等。

再加上文化传媒、电子信息、教育等上市公司对AI的关注度更高，AI应用成效更为突出。通过对A股上市公司2024年年报中包含“人工智能”、“AI”、“大模型”等关键词出现频率的TF-IDF分析，分析不同行业上市公司对AI的关注和投入程度。如图7所示，经过行业汇总可以看出，当前文化传媒行业对人工智能的关注热点最高，其次是电子信息、教育、通信等行业，位于人工智能关注度高的第一梯队，家用电器、金融、社会服务、交通运输、装备制造等处于第二梯队。由此可以看出，上市公司对人工智能的投入力度，更多依赖于其能够为业务带来的直接价值和应用效果。

因此，百度以智能云为基座、GenAI为引擎、AI云业务为出口，在“AI代写”方面构建从技术到商业的闭环更具优势。其核心竞争力在于全栈自研带来的成本与效能优势，以及深耕行业的场景理解力，正在推动AI从“技术试点”迈向“产业落地”阶段。

百度智能云是百度面向企业级市场推出的云计算服务品牌，定位为“云智一体”的技术平台，整合了云计算、大数据和人工智能能力，为企业提供全栈智能化解决方案。GenAI是百度智能云的战略重心，以文心大模型为核心，推动生成式AI在企业场景的落地。百度文库AI功能推动订阅收入增长21%。如图8所示，AI云业务基于云平台的AI服务商业化。百度通过智能云将GenAI技术转化为企业可用的产品，实现商业闭环。

3.2.2 现有付费场景与盈利模型解析

“AI代写”的现有付费场景主要包括商业文案撰写、学术写作辅助、个人创作支持等公域和私域场景。在商业文案撰写方面，企业需要大量的产品介绍、广告文案、营销策划等文本内容，根据市场调研机构的数据，企业在商业文案撰写上的年度支出平均达到100万元。一些大型企业为了确保文案质量和品牌一致性，愿意支付较高的费用使用“AI代写”服务，每篇商业文案的收费在500-2000元不等。

学术写作辅助主要面向学生和科研人员，帮助他们完成论文、报告等学术作品的撰写。据统计，国内高校学生每年在学术写作辅助服务上的花费约为5亿元。一些“AI代写”平台针对学术写作提供了专业的模板和参考文献管理功能，收费根据论文的篇幅和难度而定，每千字收费在30-100元。

个人创作支持包括小说、诗歌、散文等文学作品的创作辅助。随着人们对精神文化需求的不断提高，个人创作市场逐渐兴起。根据相关行业报告，个人创作辅助服务的市场规模每年以20%的速度增长。目前，一些平台提供按创作时长收费的模式，每小时收费20-50元。

从盈利模型来看，“AI代写”平台的主要收入来源为付费服务收入。从表5可得，文心一言年

来源：2025年AI赋能教育行业发展趋势报告

图8. AI代写市场从技术到商业闭环

度总收入的智能云板块，依托AI+服务实现逐年上升。其中，商业文案撰写收入占比40%，学术写作辅助收入占比30%，个人创作支持收入占比30%。同时，平台还需要承担技术研发、数据存储、市场推广等成本。根据该平台的财务报告，其年度总成本中，技术研发成本占比35%，数据存储成本占比20%，市场推广成本占比25%，其他成本占比20%。

表5. 百度旗下智能云板块年收入情况

财务数据			
年报	2024财年	2023财年	2022财年
收入	22.5亿美元	14.9亿美元	10.2亿美元
收入增长	50.98%	46.15%	-76.78%

来源：百度集团各年度报告

3.2.3 用户需求分层与市场渗透率调研

用户对于“AI代写”服务的需求可以根据使用目的、付费能力和专业程度等因素进行分层。第一层为基础需求用户，主要是学生和普通个人用户，他们使用“AI代写”服务主要是为了完成日常作业、个人博客写作等简单任务，对价格较为敏感，付费能力相对较弱。第二层为专业需求用户，包括企业营销人员、科研人员等，他们需要高质量、专业性的文案和学术作品，对服务的准确性和专业性要求较高，愿意支付较高的费用。第三层为高端定制需求用户，主要是大型企业和知名作家等，他们需要个性化的、具有独特风格的写作服务，对服务的质量和创意要求极高，付费能力很强。

根据艾媒咨询的调研数据，目前“AI代写”服务在整体写作市场中的渗透率约为15%。其中，在学生群体中的渗透率较高，达到了30%，主要集中在大中专院校；在企业市场中的渗透率为10%，主要集中在互联网、金融等行业；在个人创作领域的渗透率为5%，但随着文化创意产业的发展，这一比例有望逐步提高。

3.3 文心一言深度聚焦：商业逻辑与服务协议精析

3.3.1 公司状况与战略定位

文心一言是百度公司推出的一款人工智能写作工具，百度作为国内领先的互联网科技企业，拥有强大的技术研发实力和丰富的数据资源。截至目前，百度在人工智能领域的研发投入累计超过1000亿元，拥有超过10000名研发人员，在自然语言处理、深度学习等核心技术方面取得了众多突破。

在GenAI层面，文心一言的战略定位是成为全球领先的智能写作平台，为用户提供高效、准确、个性化的写作服务。其目标用户群体广泛，包括学生、企业员工、作家等各类有写作需求的人群。为了实现这一战略定位，文心一言不断优化算法模型，提升文本生成质量，同时加强与各行业的合作，拓展应用场景。例如，与新闻媒体合作，提供新闻稿件撰写辅助；与出版社合作，助力图书创作等。

文心一言的商业化进展呈现出免费策略驱动用户增长、政企合作构筑壁垒、技术输出与生态协同并重的特点，其核心路径可概括为“以免费模式扩大用户基数，通过广告、云服务、硬件销售及政企合作实现规模化变现”。2025年4月1日起，文心一言宣布全面免费开放，放弃原有的年费398元会员收费模式，开放超长文档处理、专业检索增强等高级功能。这一策略直接推动用户规模激增，根据一季度报告数据：截至2025年7月，文心一言用户达4.3亿，日均调用量超15亿次，较2023年增长超30倍。

免费模式虽牺牲了会员收入，但通过占教育业务营收60%的广告变现和与微播易合作开发创作者工具云服务拓展实现间接盈利。例如，百度通过重构广告系统，利用文心大模型生成创意素材和精准定向，使广告商转化率平均提升高个位数，达内教育等案例中转化率提升23.3%，ROI提升22.7%。

3.3.2 变现模式与用户分层

我们关注在百度宣布全面放开使用前的商业数据可以发现，文心一言的变现模式主要包括订阅收费、按次收费和增值服务收费。订阅收费是其主要收入来源之一，根据不同的订阅时长和功能权限，设置了多种套餐供用户选择。如前面所述，月度订阅价格为50元，年度订阅价格为500元。按次收费则根据用户生成文本的具体需求，按照字数或复杂

程度计费，每千字收费5元。在免费开放之前，增值服务收费是文心一言为了满足用户个性化需求而推出的收费项目，例如提供专业的文案润色、风格转换、行业报告定制等服务，收费根据服务内容和难度而定，价格在50-500元不等。

在用户分层方面，文心一言根据用户的付费能力、使用频率和需求复杂程度将用户分为普通用户、高级用户和企业用户。普通用户主要使用基础的写作功能，付费意愿较低；高级用户对写作质量和服务体验有较高要求，愿意支付一定的费用享受更多高级功能；企业用户则有大规模的写作需求，需要定制化的解决方案，是文心一言的重要客户群体。在百度宣布全面放开使用前的数据显示，普通用户占比60%，高级用户占比30%，企业用户占比10%，但企业用户的付费金额占总付费金额的40%。

4 法商边界探析：收费场景下的AI代写法律风险与责任边界

图9系统梳理了AI生成内容领域面临的内容合规风险，核心聚焦于版权归属困境、学术伦理、虚假与违法信息、深度伪造滥用等关键议题。同时，图示也深入探讨了训练数据版权、生成内容确权及第三方侵权等法律合规性困境，并延伸至服务资质、隐私保护、合同责任、跨境法律及技术滥用社会影响等多个维度的风险与挑战，全面揭示了该领域复杂的法律与伦理图景。

生成式人工智能（AIGC）技术的商业化浪潮正以前所未有的速度重塑产业生态，其收费应用场景的广泛拓展在催生巨大经济价值的同时，也深刻冲击着现行法律框架的边界。正如欧盟《人工智能法案》所警示的，当AI从技术实验走向市场交易，其引发的“责任鸿沟”与“监管滞后性”已成为全球性治理难题。在我国，随着《生成式人工智能服务管理暂行办法》的施行，平台运营者被明确赋予内容安全主体责任，标志着监管从“技术中立”转向“应用问责”的关键跃迁。

文心一言作为百度推出的生成式人工智能服务，在2024年经历了从收费到免费的商业模式转变。在2025年4月1日全面免费之前，平台实施了多层次的收费策略，以满足不同用户群体的需求。这种收费策略，无疑也增加了法律风险，对责任边界的认定产生了影响。

4.1 内容合规风险

4.1.1 学术伦理红线：代写引发的诚信危机

在需要付费使用的情况下，人工智能工具可

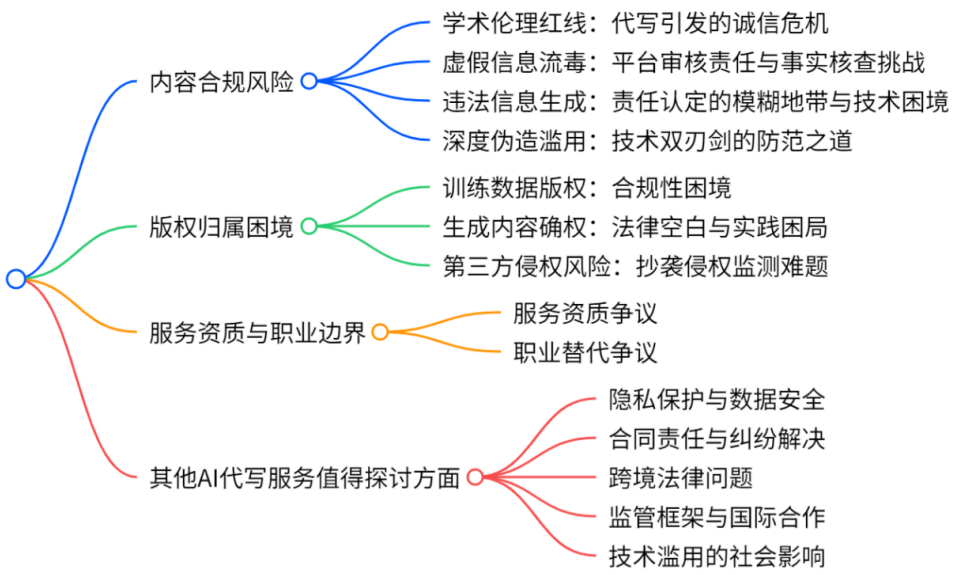


图9.“法商边界探析：收费场景下的AI代写法律风险与责任边界”思维导图

能引发学术道德问题[4]。以文心一言为例，这类工具存在被用于代替学生完成论文写作或作业的风险。该平台在2024年推出的会员收费制度，特别是每月49.9元的专业版服务，提供了不限次数的内容生成功能，这在客观上增加了学术不规范行为的可能性。根据调查发现，有商家通过技术接口提供服务，用户每月支付不足10元就能无限量获取生成内容。虽然商家声称这些内容是原创的，但实际情况可能包含抄袭他人学术材料的现象。曾有研究生在毕业论文开题报告中，80%内容依赖人工智能生成，这严重影响了学生独立完成学术任务的能力。值得注意的是，平台的使用条款存在某些矛盾情况：虽然用户协议明确禁止违规操作，但对于付费用户缺乏特别的管理措施，相关审核标准也没有相应提高。

我国2024年实施的《学位法》对学术造假问题制定了严格规定。其中第37条明确指出，如果学位论文被发现由人工智能代写，学校有权撤销已颁发的学位证书。这项规定不仅适用于毕业论文，还包括日常课程作业、实验数据报告等各类学习成果。例如北京信息科技大学在2025年出台了专门文件，禁止学生使用人工智能完成作业或报告，并对违规行为制定了具体的处罚措施。

其他国家在管理人工智能使用方面也采取了不同方法。美国有大学要求学生必须在教师监督下完成论文初稿，并且详细记录修改过程。韩国通过修改法律条文，将使用深度伪造技术制作不良影像的行为列为违法，同时在校园内开展相关法制教育。这些做法显示，通过记录写作过程和保存修改痕迹等方式，可以有效提升学术规范的管理效果。

当前教育领域对人工智能的合理使用范围尚未形成共识。很多高校没有明确规定学术成果中人工智能生成内容的允许比例，检测手段也主要依靠教师的个人经验。这种状况导致部分学生利用技术漏洞逃避检查，形成检测与反检测的技术对抗。例如有学生先用人工智能生成内容，再通过特殊软件进行二次修改，从而降低被查出的概率。

为解决这些问题，部分高校开始建立管理制度。复旦大学和福州大学都出台了相关规定，对毕业论文中的人工智能使用比例进行检测，并明确限定可使用的环节。学界目前提出的判断标准是：允许利用人工智能进行资料收集或草稿构思，但核心观点和最终定稿必须由学生独立完成。如果论文主体内容由机器生成，且学生没有进行深入思考，就会被认定为学术不端行为。这种区分方法有助于划分学术规范的具体界限，也为相关问题的处理提供了参考依据[5,6]。

4.1.2 虚假信息流毒：平台审核责任与事实核查挑战

提供AI代写服务的平台需要认真做好内容检查。按照当前相关规定，像文心这类智能软件在法律上被看作网络信息制作单位，必须做好信息安全管理工作。具体来说，平台应当给所有AI生成的内容打上特殊标记，同时及时处理违规信息。例如，文心基础版每月收费几十元就能制作商业广告文案，但运营方必须按照相关管理办法对这些文案做好标记和检查。

特别需要注意的是，如果平台没有对用户输入的指令进行必要审核，导致AI编造出虚假药品广告等情况，这样的平台不仅要接受行政处罚，还要和用户一起承担法律责任。不过在现实中，做好内容核实其实面临很多实际困难。以文心为代表的人工智能生成的内容经常变化多端，意思表达复杂，传统检查方法很难完全识别清楚。比如系统可能编造出看似真实的数字资料或参考文献，实际上这些内容都没有真实来源[7]。再加上全面检查需要投入大量人员和设备，很多中小型平台很难承担得起这些开支。

为解决这些问题，有些平台开始尝试”先用技术筛查，再人工检查，最后请专家评估”的多步骤检查方法[8]。比如有个版权保护系统，它通过智能扫描功能检查不同平台的内容，利用多媒体特征识别技术自动生成侵权关系图，同时使用区块链保存电子证据，方便后续走法律流程。国家《互联网信息服务管理规定》也明确要求对AI生成的图片视频打上标识，方便追查来源。但有个问题值得注意，文心的用户协议第5.3条把所有”内容合法性问题”

的责任都推给用户，这样完全让平台避开检查义务的做法确实存在争议。

从法律角度看，如果平台明知用户用AI制造虚假信息却不采取措施，就可能被认定为共同违法。比如网购平台如果发现商家用AI工具刷好评却不制止，就可能违反《消费者权益保护法》被处罚。再比如AI生成的内容损害他人名誉时，平台还有义务帮助查找来源并提供维权帮助。这些规定提醒我们，平台不能只收服务费却忽视自身应该承担的社会责任。

4.1.3 违法信息生成：责任认定的模糊地带与技术困境

人工智能写作工具可能被错误地用来制造违法信息，例如鼓动反对政府、传播极端主义或者散布色情内容。根据国家关于智能写作工具的管理办法（《生成式人工智能服务管理暂行办法》），开发这些工具的公司必须在程序编写和数据选择这些重要步骤里做好防范工作。比如像”文心一言”这类智能系统，如果学习资料里混入了非法内容，最后产生危害性言论的话，软件公司可能要负法律责任。

现在确定责任归属存在两个难点。首先是技术原因难以查清，因为智能系统生成的内容受到程序算法、学习材料和用户指令的共同影响。比如用户明明输入的是正常问题，但系统本身存在缺陷导致输出违法内容，这时候要追究谁的责任就成了难题。很多学者认为这时候应该主要追究软件公司的责任。其次是判断企业是否存在主观过错，要看他们有没有认真检查学习资料。如果公司没有仔细审核学习资料导致系统出现错误，就可能被认定为存在过失。

技术上的困难主要来自智能系统的不可预测性[9]。以”文心一言”为例，它的思考过程就像黑箱难以完全解释清楚。即便企业做了合规工作，程序漏洞还是可能产出问题内容。而且现在有各种隐藏信息和图像合成的技术手段，违规信息的形式越来越隐蔽，这让监管部门更难及时发现问题。

为了解决这些问题，我国开始实施分级管理制度。相关管理办法明确要求，具有公众影响力的智能服务必须通过安全审查并登记备案。学术界也提出”过程监管”概念，认为应该重点检查学习材料是否合法、系统运作是否透明等内容。

对于违反规定的企业，法律规定了警告、罚款、暂停服务等处罚措施；如果情况严重构成犯罪还要追究刑事责任。比如制作传播色情视频就可能触犯相关刑法。不过根据实际测试，”文心一言”系统目前对赌博、暴力、色情等违规内容的过滤效果还不错，这些方面暂时没有明显问题。

4.1.4 深度伪造滥用：技术双刃剑的防范之道

深度伪造技术在人工智能代写领域的滥用主要表现为身份伪造和视听资料篡改。例如，利用人脸替换技术制作虚假名人代言视频进行商业宣传，或篡改司法证据材料影响诉讼程序公正性。根据《互联网信息服务深度合成管理规定》，使用深度合成技术必须取得被编辑者的明确同意，并对生成内容进行显著标识。目前“文心一言”4.5Turbo版本已在输出内容底部标注“以上内容由文心人工智能生成”，体现了合规实践。

技术防范体系可采取主动防御与被动检测双轨并行的模式。在主动防御层面，诸如可扩展通用对抗性水印技术可将数字水印嵌入原始素材，当这些素材被用于人工智能处理时，水印将触发保护机制干扰伪造内容的生成。在被动检测方面，蚂蚁数科的ZOLOZ系统通过设备端安全检测与活体识别技术，在金融场景中实现对深度伪造攻击99.9%以上的识别准确率。

法律责任的认定需根据行为性质与危害程度进行区分。未经许可使用他人肖像进行深度伪造可能侵犯肖像权与名誉权；伪造身份认证信息则可能构成诈骗罪或侵犯公民个人信息罪。《网络数据安全管理条例》同时要求生成式人工智能服务提供者加强训练数据安全管理，防范数据泄露风险。

当前深度伪造治理仍面临诸多挑战：技术门槛持续降低、传播速度极快，且常通过加密货币交易与境外服务器规避监管[10]。对此，学术界建议构建“生成—传播—消除”全流程治理体系，包括建

设国家级检测平台、建立跨国执法协作机制等系统性解决方案。

4.2 版权归属困境

4.2.1 训练数据版权：合规性困境

文心一言等AI代写服务，其训练数据有可能包含受版权保护的作品，比如书籍、论文、影视片段等。按照《生成式人工智能服务管理暂行办法》的规定，服务提供者必须使用来源合法的数据，并且保证不侵犯他人的知识产权。如果训练数据里有未经授权的内容，就可能构成侵权。以文心一言为例，其训练数据含万亿级文本，尽管原协议第5.4条声称用户“享有生成内容权利”，但训练阶段的数据侵权风险未解决：专业版会员依赖海量语料库，若按欧盟规则公开来源（如知乎专栏、学术论文），需支付版权费，会员定价（49.9元/月）无法覆盖成本。

合规方面的困难主要体现在数据来源非常复杂，而且授权机制还不完善。一方面，训练数据可能通过网络爬虫从网上抓取，很难逐一确认版权归属；另一方面，现行法律对于数据集合的版权保护规定还不明确[11]。比如，美国版权局在2023年做出裁定，认为AI生成的内容不受版权保护，但如果训练数据包含侵权的内容，提供者仍然要承担责任。在中国，深圳南山区法院曾经认定AI生成的财经文章是受著作权法保护的作品，但是其训练数据的版权问题并没有明确。

为了解决这个问题，学术界提倡建立一个“合理使用+授权许可”的双重机制。对于公开可获取的数据，可以适用合理使用原则；对于受版权保护的作品，则需要获得授权。例如，欧盟的《人工智能法案》要求提供者公开训练数据的来源，并对侵权数据承担相应的责任。

4.2.2 生成内容确权：法律空白与实践困局

目前，各国对人工智能生成内容的著作权认定存在显著差异。美国版权局明确规定，完全由人工智能生成的内容不受著作权法保护，仅当内容体现人类智力贡献时方可获得版权。相比之下，中国在此领域的法律实践呈现较大复杂性。尽管司法实践中已出现承认人工智能生成内容可构成作品的案例，但关于权利归属仍缺乏明确法律依据[12]。以深圳南山区人民法院审理的“Dreamwriter案”为例，法院认定由人工智能系统撰写的财经报道具有作品属性，却未对著作权归属作出清晰界定。

这一法律困境的根源在于人机协作创作的特殊性。当用户通过输入指令获取“文心一言”等人工智能系统的生成内容，并在此基础上进行修改完善时，如何准确界定人类与机器各自的创作贡献比例成为关键难题。例如，律师使用人工智能生成法律文书初稿后作出实质性修改，此时著作权的归属便存在争议：是归属于进行最终完善的律师，还是归属于提供人工智能服务的平台？当前学界普遍认为，若用户对生成内容进行了充分体现独创性的修改与完善，则该用户可被视为著作权人；反之，若仅进行微小改动，则可能因缺乏独创性而无法获得著作权保护[13-14]。

为弥补这一法律空白，部分学者提出设立“人机协作作品”特殊类别，建议根据人类与人工智能在创作过程中的实际贡献度进行权利划分。

4.2.3 第三方侵权风险：抄袭侵权监测难题

人工智能代写技术的应用可能导致生成内容与既有作品构成实质性相似，从而引发抄袭或著作权侵权争议[15]。例如，学生使用“文心一言”生成的论文若在内容表述或结构框架上与已发表文献高度重合，则可能被认定为学术抄袭。根据我国《著作权法》规定，抄袭行为需承担停止侵害、赔偿损失等法律责任。

由于人工智能生成内容具有表现形式隐蔽、文本特征多变等特点，传统抄袭检测手段面临严峻挑战。早期版本的知网检测系统等工具对人工智能生成内容，特别是经过同义词替换、句式重构等后期处理的文本识别效果有限。实践中存在学生先获取人工智能生成内容，再通过文本润色手段规避检测的情况，这为教育机构的学术诚信监管带来困难。

针对这一困境，学术界建议采取技术检测与过

程管理相结合的综合治理方案。在技术层面，应研发专门针对人工智能生成内容的检测工具，通过分析文本特征（如词汇分布模式、句法结构特点等）识别机器生成痕迹。在管理层面，可建立使用披露制度，要求用户明确标注人工智能辅助内容，并保存创作过程记录（包括使用指令、修改日志等）以实现全流程溯源。目前，复旦大学已率先要求学生在毕业论文中明确标注人工智能生成内容的比例及具体使用范围。

4.3 服务资质与职业边界

4.3.1 服务资质争议

在涉及收费的情况下，AI代写服务可能会触及非法提供法律服务、教育服务等问题。在功能方面，文心一言的“AI律师”可以帮助用户进行法律文本的撰写和审核，提供法律咨询和合同解读等服务。但如果AI代写诉讼相关的法律文书或合同，就可能违反《律师法》中对于法律服务资质的规定。根据我国的《律师法》，只有获得律师执业证书的人员才有资格提供法律服务。如果AI代写服务的提供者没有相应的资质，那么就可能构成非法执业行为。

关于资质问题的争议，核心在于如何认定AI服务的法律性质。如果将AI代写视为一种工具，用户自己使用这种工具并不涉及资质问题；但是，如果服务提供者直接提供最终完成的成品（比如代写的法律文书），那么就可能被认定为提供了法律服务，因此需要具备相应的资质。例如，某个平台如果宣传说他们拥有“专业的律师团队+AI生成的法律文书”，那么可能因为没有获得律师事务所的执业许可而面临处罚。

为了明确这些行为的边界，学术界建议根据服务模式进行分类监管。对于工具类的服务，应该要求提供者进行风险提示；而对于提供成品的AI服务，则需要取得相应的资质。欧盟的《人工智能法案》就是一个很好的范例：它将AI服务分为了“不可接受风险”“高风险”“有限风险”和“低风险”四个类别，并分别采取禁止、严格监管、一般监管和自律等不同的管理措施，具有很强的借鉴意义。

4.3.2 职业替代争议

AI代写对传统职业的影响主要集中在法律、教育和写作等行业。比如，像文心一言这类AI生成的法律文书，可能会取代一些初级律师的工作，从而导致职业结构的调整。伦理方面的争议主要集中在技术替代的公平性以及它可能带来的社会影响。如果AI代写导致大量的法律助理失业，那么可能会加剧社会的不平等现象。

从法律的角度来看，职业替代本身并不属于违法行为，但是需要防止因为技术滥用而导致的垄断和不正当竞争。根据《生成式人工智能服务管理暂行办法》，提供AI服务的机构不能利用算法、数据或平台等优势，来实施垄断或者不正当竞争行为。如果一个AI代写平台通过低价策略排挤传统的法律服务机构，就可能违反《反垄断法》。

为了应对AI技术带来的职业替代挑战，学术界建议通过职业培训、教育改革以及调整社会保障体系等方式，来帮助从业者实现转型。

4.4 其他AI代写服务值得探讨方面

4.4.1 隐私保护与数据安全

AI代写服务在运营过程中可能会收集用户的敏感信息，比如学术论文的主题、法律需求的具体细节等[16]。根据《个人信息保护法》，服务提供者需要征得用户的明确同意，并且采取加密、访问控制等措施来保护数据的安全。

在技术层面上，可以采用动态加密、访问权限管理等方式来保护数据安全。例如，蚂蚁数科的ZOLOZ系统通过在设备端检测环境安全，以及在应用端进行活体特征识别，来防止数据泄露。在法律责任方面，如果服务提供者没有履行安全义务，导致用户数据泄露，就需要承担相应的行政处罚和民事赔偿责任。

4.4.2 合同责任与纠纷解决

用户与AI代写服务提供者之间的合同需要明确双方的责任划分。例如，服务协议中应该约定生

成内容的版权归属、违约责任以及争议的解决方式等等。不过，以文心一言为例，目前的用户协议当中显然对前述内容未能作出明确有效的规定。根据《民法典》，如果协议条款存在免除服务提供者责任、加重用户义务等情况，那么这些条款可能被认定为无效。

在纠纷解决机制方面，建议引入在线仲裁或者区块链存证等方式，以提高解决问题的效率。

4.4.3 跨境法律问题

AI代写服务可能会涉及到跨境数据流动以及法律适用方面的冲突。例如，一个中国用户使用境外的AI平台代写论文，那么他的数据可能存储在境外的服务器上，因此需要遵守欧盟的《通用数据保护条例》（GDPR）等相关规定。我国的《生成式人工智能服务管理暂行办法》要求境外向境内提供服务时，必须符合中国的法律，否则将会采取技术措施进行处理。

为了应对跨境合规方面的挑战，学者建议加强国际合作，推动建立统一的AI治理标准。例如，欧盟的《人工智能法案》和美国的《国家人工智能倡议》等，在生成内容责任的认定上存在差异，因此需要通过多边谈判来进行协调。

4.4.4 监管框架与国际合作

我国已经建立了以《生成式人工智能服务管理暂行办法》为核心的监管框架，其中涵盖了安全评估、算法备案以及内容标识等方面的要求。在国际层面，欧盟的《人工智能法案》也已经开始实施，重点关注风险分级、透明度以及责任追究等方面。

监管方面的挑战包括技术发展的快速性以及各国标准之间的差异。例如，我国对于生成内容标识的要求与欧盟的“高风险”分类之间存在差异，这可能会导致跨境服务的合规成本增加。学术界建议通过国际组织（比如联合国教科文组织）来推动标准的互认和执法协作。

4.4.5 技术滥用的社会影响

AI代写可能会加剧信息茧房的现象，并且降低公众的判断能力。例如，虚假新闻在通过AI生成之后，可能会快速传播并且误导舆论。此外，技术滥用还可能导致法律行业的信任危机，从而影响司法的公正性。现实中，如果AI生成的法律文书中存在错误，那么可能会损害当事人的权益。为了防范这些社会影响，学术界建议加强公众教育，提高公众对于AI生成内容的辨别能力。

4.5 小结

文心一言等AI代写在收费场景下的法律风险与责任边界问题，涉及学术伦理、虚假信息、版权归属、服务资质等多重挑战。当前监管框架虽已初步建立，但仍面临技术快速迭代与法律滞后的矛盾。明确平台责任、完善版权规则、强化审核机制，以及推动国际合作，是平衡技术创新与风险防控的关键。未来需通过动态立法、技术治理与公众教育，构建适应AI发展的法治生态，确保技术应用在合规轨道上创造价值。

5 面向收费时代的平台治理：合规体系创新设计

图10系统阐述了AI时代平台法律责任体系的演进与治理框架创新。其核心内容包括：直接责任与间接责任（避风港原则）在AI影响下的边界重塑，以及将算法可解释性（XAI）嵌入免责情形的迫切需求。图示进一步构建了覆盖事前、事中、事后的全链条治理方案：从事前筑牢合规地基，到事中建立AI驱动的实时监控，再到事后完善AI赋能的纠纷解决。同时，提出了三大创新治理工具：实施可解释AI以破解算法黑箱、建立动态分级审核机制、构建跨平台黑名单共享联盟，最终形成法律责任分级模型与全流程合规义务清单，为平台治理提供了系统化的创新方案。

在数字经济迈向深度收费化的关键转型期，平台治理面临前所未有的复杂挑战。用户为数字产品与服务付费意愿的提升，既为平台经济注入了新活力，也使得平台责任边界、数据合规义务及治理效能问题被置于放大镜下检视。传统“避风港”原则在收费模式下正遭受严峻拷问——当用户为内容

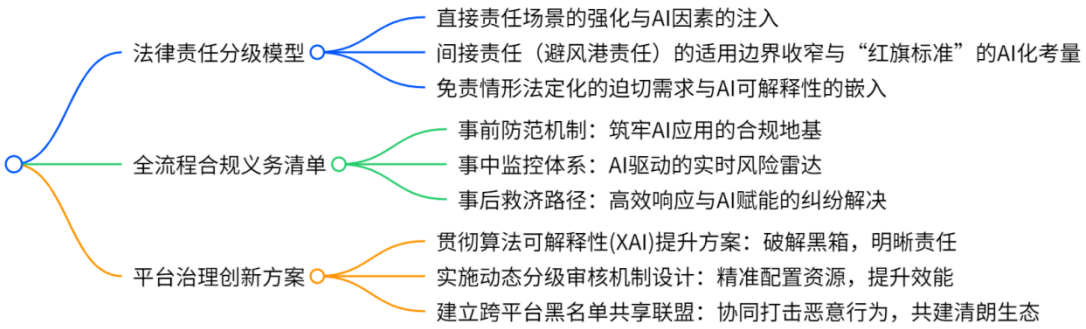


图10. “面向收费时代的平台治理：合规体系创新设计”思维导图

或服务直接付费时，平台是否还能以“技术中立”或“通道角色”推卸责任？文心一言等人工智能写作工具的爆发式渗透，更在内容生成、版权归属、责任认定等维度为平台治理体系增添了全新的复杂性变量。

5.1 法律责任分级模型：构建适应收费模式与AI特性的责任框架

收费模式的深化与AI生成内容的普及，彻底重构了平台责任认定的逻辑基础，亟需构建精细化的法律责任分级模型。

5.1.1 直接责任场景的强化与AI因素的注入

当平台从“被动通道”转变为“主动服务提供者”或“内容共同创作者”时，其责任性质发生根本性跃迁[17]。在收费场景下，若平台利用文心一言等AI工具主动生成并向用户直接提供付费内容（如AI撰写的研究报告、法律文书、营销文案等），或深度介入并实质性地控制、编辑、推荐AI生成内容，则平台应承担类似“出版者”的直接责任。例如，某付费知识平台使用文心一言批量生成行业分析报告并直接售卖，报告出现事实性错误或侵权内容，平台作为内容的生产组织者与直接获利者，显然无法援引传统免责条款，需承担直接的违约责任乃至侵权责任。欧盟《数字服务法案》(DSA)对在线平台作为“托管服务”和“在线平台”的区分，以及其对“非常大型在线平台”(VLOPs)施加的更重义务，体现了对平台角色与责任差异化的认知趋势。

5.1.2 间接责任（避风港责任）的适用边界收窄与“红旗标准”的AI化考量

对于用户利用平台提供的AI工具（如集成“文心一言”接口的写作助手）生成并上传的内容，平台通常仍可主张间接责任适用“通知-删除”规则。然而，收费模式显著提高了平台对用户内容应尽的注意义务标准。当AI生成内容具有明显的侵权或违法特征（“红旗”），或者平台从特定侵权内容中直接获得经济利益，其“避风港”抗辩空间将被极大压缩。例如，付费小说平台用户利用内置AI写作功能生成并上传明显抄袭知名作品的情节段落，平台若因该内容吸引付费订阅而获利，且抄袭特征明显（“红旗”），其未主动采取过滤措施即可能丧失免责资格。美国DMCA中的“直接经济利益”条款和“红旗标准”在此类场景下的适用性将显著增强。

5.1.3 免责情形法定化的迫切需求与AI可解释性的嵌入

为避免责任认定的模糊性，亟需通过立法或司法解释将特定情形下的平台免责条件予以法定化、明晰化。尤其是在AI参与内容生成与分发的场景下，以下免责情形应重点考量。

首先，是AI技术固有缺陷导致的不可预见错误。如平台已采用主流、合规的AI模型（如已备案的“文心一言”），并实施合理的安全训练措施（如过滤有害数据、设置安全护栏），仍不可预见地出现有害或侵权输出，且在事后及时采取有效补救措施（如

撤回、修正、通知）。

其次，是用户恶意绕过安全机制。用户通过精心设计的Prompt（提示词）故意诱导AI生成违规内容，平台的安全机制在常规监测下难以实时、完全阻断此类恶意行为。

最后，即第三方来源数据污染。若平台使用的训练数据来源于合法授权的第三方数据库，该数据库本身存在未被发现的侵权或违法内容，导致AI输出结果出现问题，平台对此无主观过错且难以事前察觉。明确此类免责情形，既能保障平台在合规投入后的稳定预期，也能鼓励技术创新，避免过度责任阻碍AI应用的健康发展。

5.2 全流程合规义务清单：贯穿业务生命周期的风险防控链

平台治理必须从静态合规转向动态风控，建立覆盖事前、事中、事后的全流程合规义务清单，并将AI治理能力深度融入其中。

5.2.1 事前防范机制：筑牢AI应用的合规地基

第一，要进行AI模型的准入与评估。应建立严格的第三方AI模型（如“文心一言”）接入审核机制，评估其安全性、可解释性、数据合规性、伦理合规性及版权处理机制。进而，对模型可能产生的偏见、歧视、幻觉（Hallucination）等风险进行充分测试与评估。

第二，须对用户协议与风险告知进行重构。在用户协议中，要清晰界定AI生成内容的标识要求、版权归属规则（用户享有？平台享有？共同享有？）、使用限制以及可能存在的风险（如事实性错误、版权不确定性）。对收费用户，根据权利义务相一致的原则，需进行更显著、更具体的风险提示。明确告知用户在使用平台提供的AI写作功能（如调用“文心一言”API）时的责任边界。

第三，实行内容安全与版权过滤的AI前置。部署先进的AI内容安全系统（可集成或借鉴“文心一言”的安全能力），在用户输入提示词（Prompt）阶段即进行初步风险扫描（如识别恶意诱导、违禁关键词），并在AI生成内容输出时，同步进行版权相似度比对（如接入版权库）、违禁内容识别（涉恐、暴、毒、虚假信息等）、敏感信息过滤等。同样地，收费内容应适用更严格的前置过滤标准。

第四，夯实数据合规基石。要严格遵循知情同意、最小必要、目的限制等原则收集和处理用户数据，特别是用于训练或优化AI模型的数据。建立用户数据的分类分级保护制度，确保数据跨境传输的合法性。对用于训练AI的语料库进行彻底的版权和合法性审查。

5.2.2 事中监控体系：AI驱动的实时风险雷达

第一，进行AI生成内容的动态监测与标识。具体操作方法为：利用技术手段（如数字水印、特定元数据），对平台内所有AI生成内容进行强制性、不可篡改的显著标识（例如，“此内容由AI生成”）。建立实时监测系统，追踪AI生成内容的传播路径、用户互动情况，利用AI技术（如自然语言处理、图像识别）对海量内容进行自动化、高效率的合规性二次筛查，尤其关注收费热点内容。文心一言在输出时左上角自动显示标识的做法，为平台提供了可借鉴的范例。

第二，实施高风险节点与主体的AI增强审查。运用大数据分析和机器学习模型，识别高风险用户（如频繁投诉对象、发布历史不良者）、高风险交易（如大额、异常支付模式）、高风险内容主题（如医疗健康、金融投资、时政新闻）。对识别出的高风险对象和内容，触发人工+AI的增强审核流程。对收费内容及其关联的营销推广信息，实施更频繁、更深入的质量与合规抽检。

第三，施行算法决策透明与干预。对影响用户权益（如内容推荐、信用评估、定价策略）的核心算法，建立透明化机制（在不泄露商业秘密前提下解释基本原理和主要参数）。设置人工干预接口，允许合规人员对算法运行中发现的系统性偏差或突发风险进行紧急干预和调优，确保算法在“事中”阶段的可控性。

5.2.3 事后救济路径：高效响应与AI赋能的纠纷解决

第一，构建畅通的投诉举报机制。建立便捷、

多通道（在线表单、邮件、电话）的侵权投诉和违法举报入口。利用AI（如基于NLP的分类和摘要系统）对海量投诉进行初步分类、优先级排序和关键信息提取，大幅提升处理效率。确保收费用户的投诉享有优先响应级别。

第二，贯彻“通知-必要行动”规则的AI化执行。收到有效侵权通知后，要迅速利用AI技术快速准确定位侵权内容（或链接），并根据分级模型和预设规则（结合直接/间接责任判断、是否收费、是否AI生成等因素）采取差异化行动（如限流、屏蔽、删除、断开链接、终止服务）。行动过程需记录留痕，行动结果应及时反馈通知方和被投诉方。

第三，实施AI辅助的争议调解与裁决。探索利用AI工具辅助处理事实相对清晰、争议不大的纠纷。例如，在版权小额纠纷中，AI可快速比对涉嫌侵权内容与权利作品，提供相似度报告和初步判断建议；在交易纠纷中，AI可分析聊天记录、交易流水等证据，提出调解方案参考。复杂纠纷仍需人工介入。

第四，落实系统性风险回溯与模型迭代。建立事后分析机制，对重大合规事件、高频投诉类型、AI模型失效案例进行深度复盘。分析结果用于优化事前防范策略、事中监控规则，并反馈给AI模型提供方（如“文心一言”团队）用于模型迭代升级，形成“事前-事中-事后”的治理闭环和AI能力的持续进化。

5.3 平台治理创新方案：技术赋能与协同共治

应对收费模式与AI代写的双重挑战，亟需超越传统监管思路，拥抱技术驱动的治理创新。

5.3.1 贯彻算法可解释性(XAI)提升方案：破解黑箱，明晰责任

首先，进行技术路径深化。平台应积极研发或采用先进的XAI技术（如LIME,SHAP,Counterfactual Explanations），使其核心算法（特别是内容生成、推荐、审核、定价算法）的决策逻辑在一定程度上可被人类理解。例如，解释“文心一言”为何生成了某段特定文本（基于哪些输入特征、训练数据的哪部分权重较高），或者解释推荐系统为何将某篇AI生成的文章推送给特定用户。

其次，推动解释内容标准化与场景化。具体而言，须制定平台内部的算法解释标准，明确在不同场景下（如用户投诉、监管问询、内部审计）需要提供何种深度和形式的解释（如特征重要性排序、决策规则模拟、反事实示例）。对直接影响用户付费决策或重大权益的算法（如信用评分、个性化定价），需提供更详实、用户友好的解释。

最后，进行可解释性赋能责任认定。将算法可解释性的输出作为判断平台是否存在主观过错、是否履行合理注意义务的重要依据。清晰的解释有助于区分是技术固有缺陷、用户恶意利用还是平台自身疏忽，为前述法律责任分级模型的应用提供坚实的技术支撑。缺乏基本可解释性的“黑箱”算法，在责任纠纷中将使平台处于极其不利的地位。

5.3.2 实施动态分级审核机制设计：精准配置资源，提升效能

第一，进行多维度风险画像构建。构建涵盖“内容风险”（如涉敏词、图像违规度、版权相似度、虚假信息概率）、“主体风险”（用户信用历史、投诉记录、行为异常度）、“场景风险”（是否收费、传播热度、涉及领域如金融健康）的多维度动态风险评估模型。该模型应能实时或近实时地输出综合风险评分。

第二，实施审核资源智能调度。基于动态风险评分，自动将内容、用户或交易分配至不同审核通道。首先，是高风险，其将触发“AI预审+多人交叉人工复审”的严格流程，必要时即时冻结或限制。其次，是中风险，要采用“AI深度审核+人工抽检”模式。最后，为低风险，主要依赖AI自动化审核，辅以极小比例随机抽检[18]。

第三，进行反馈驱动的模型优化。要持续收集人工审核结果、用户投诉、处置效果等数据，将其作为新的训练数据反馈给风险评级模型和AI审核引擎，实现审核策略的持续迭代优化。例如，当发现某种新型AI生成的隐蔽违规内容（如利用“文心一

言”生成的具有误导性的伪科学文章）未被现有模型有效识别时，应及时更新模型特征和规则。收费内容默认应获得更高的基础风险评级。

5.3.3 建立跨平台黑名单共享联盟：协同打击恶意行为，共建清朗生态

首先，要进行联盟机制创新。在严格保护用户隐私和商业秘密、遵守反垄断法规的前提下，探索建立去中心化或受监管的可信第三方平台。该平台负责安全地汇总、匹配、共享经过各参与平台严格核实的高风险主体信息（如恶意侵权人、职业刷单者、黑灰产账号、利用AI大规模生成有害内容的操纵者）。共享信息应聚焦于必要的行为特征标识符（如不可逆哈希处理的设备指纹、行为模式标签）和风险类型标签，而非原始个人数据。

其次，须实施技术保障与规则约束。利用区块链技术确保共享记录的不可篡改性和可追溯性；采用联邦学习、安全多方计算等隐私增强技术，在数据不离开本地的前提下进行风险信息匹配和协同分析；制定清晰、具有法律约束力的联盟章程，明确数据使用范围、保密义务、退出机制、争议解决及法律责任。

最后，聚焦AI滥用治理。联盟应特别关注利用AI技术（如“文心一言”等大模型）进行规模化、自动化违规行为的特征识别与共享（如特定的恶意Prompt模式、AI生成的虚假评论/刷量文本的共性特征、用于绕过安全机制的对抗性攻击方法）。这种协同防御对遏制利用AI代写技术的新型黑灰产至关重要。联盟成员在接收到高风险信息后，可在各自平台内依据自身规则进行预防性风控（如加强审核、限制功能），提升整个生态的防御能力。

5.4 小结

收费模式重塑了用户与平台的关系，AI代写则深刻改变了内容生产与分发的本质。二者叠加，对平台治理提出了系统性升级的迫切要求。构建“法律责任分级模型”是厘清权责边界的法律基石；落实“全流程合规义务清单”是实现风险防控的操作手册；而推动“算法可解释性提升”、“动态分级审核”、“跨平台黑名单共享”等治理创新方案，则是运用技术手段和协同思维提升治理效能的关键引擎。

以“文心一言”为代表的AI内容生成工具，既是这场变革的催化剂，其自身的发展与治理实践也为平台合规体系创新提供了宝贵的实证参考与能力支撑。面向未来，平台治理的成功将愈发依赖于法律规则的清晰指引、技术能力的深度赋能以及多元主体的有效协同。唯有在创新与规范之间找到动态平衡点，方能构建一个既能激发数字经济活力、保障用户权益，又能有效管控风险、促进AI代写技术健康发展的可持续平台生态。这不仅是平台企业的生存之道，更是数字时代法商融合发展的必然方向[19]。

6 商业模式优化升级：构建可持续的付费写作生态

图11展示了AI写作平台实现战略定位突破的整体框架。该框架始于对行业痛点与用户需求的精

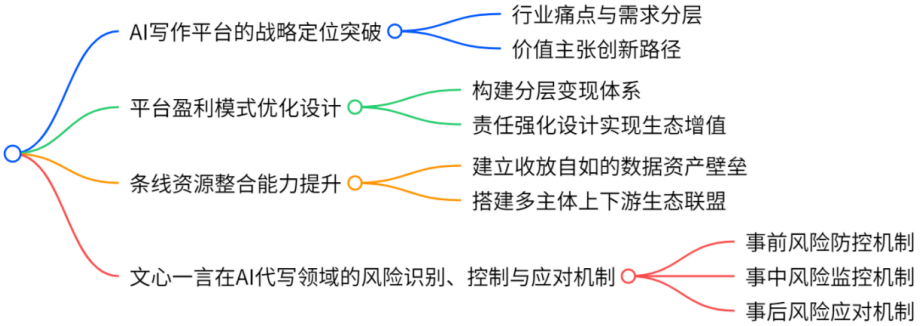


图11. “商业模式优化升级：构建可持续的付费写作生态”思维导图

准分层，进而规划价值主张的创新路径与盈利模式的优化设计，旨在构建一个多层次的分工与变现体系，并通过强化责任设计为生态系统增值。为实现此目标，平台需着力提升条线资源整合能力，建立可控的数据资产壁垒，并搭建涵盖多主体的上下游生态联盟。同时，框架还强调了覆盖业务全流程的风险管理机制，包括事前防控、事中监控与事后应对，最终形成如“文心一言”等平台在AI代写领域特有的风险识别、控制与应对体系。

6.1 AI写作平台的战略定位突破

6.1.1 行业痛点与需求分层

“AI代写”行业目前存在一些痛点，如文本质量参差不齐、缺乏个性化、版权问题等。从需求分层来看，基础需求用户更注重写作效率，希望快速生成符合要求的文本；专业需求用户则对文本的专业性、准确性和逻辑性有较高要求；高端定制需求用户追求独特的创意和个性化的表达。

为了突破战略定位，平台需要针对不同需求分层解决行业痛点。对于基础需求用户，可以通过优化算法，提高文本生成速度和质量，同时提供简单易用的操作界面；对于专业需求用户，加强与各行业专家的合作，引入专业知识和数据，提升文本的专业性；对于高端定制需求用户，建立专业的创作团队，与用户进行深度沟通，满足其个性化需求。

6.1.2 价值主张创新路径

因此，文心一言平台可以通过技术创新、服务创新和模式创新来实现价值主张的创新。在技术创新方面，不断研发新的算法模型，提高文本生成的准确性和创造性。例如，引入强化学习算法，使AI能够根据用户的反馈不断优化生成结果。在服务创新方面，提供一站式的写作服务，包括选题策划、资料收集、文本生成、润色修改等全流程服务。同时，建立用户社区，促进用户之间的交流和分享。在模式创新方面，探索与共享经济相结合的模式，让用户可以共享自己的写作经验和模板，实现资源的优化配置。

6.2 平台盈利模式优化设计

6.2.1 构建分层变现体系

根据用户需求分层和付费能力，构建更加精细化的分层变现体系。如图12所示，对于基础需求用户，提供免费基础功能和低价订阅套餐，吸引大量用户使用，通过广告收入和增值服务收费实现盈利。对于专业需求用户，推出中高端订阅套餐，提供更多专业功能和服务，如行业报告生成、专业文案润色等，收取较高的费用。对于高端定制需求用户，提供一对一的定制化服务，根据项目难度和要求收取高额费用。例如，某平台针对基础需求用户提供免费版，每月可生成10篇基础文本；付费订阅版每月收费20元，可生成50篇文本，并享受无广告体验。对于专业需求用户，推出年度订阅套餐，收费1000元，提供行业专属模板、专业数据支持等服务。对于高端定制需求用户，根据项目复杂程度，收费在5000-50000元不等。

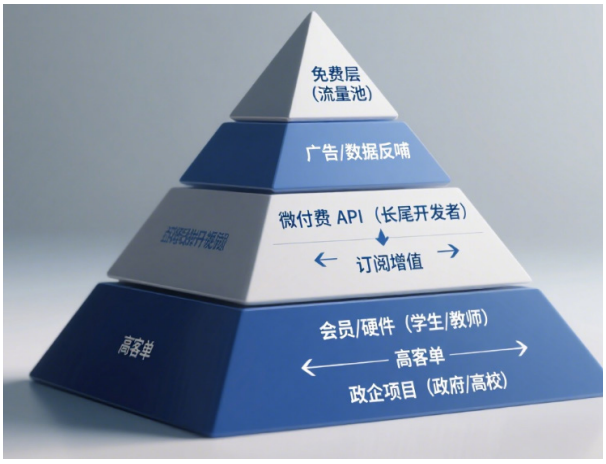


图12. 用户群体分层示意

6.2.2 责任强化设计实现生态增值

平台需要强化自身责任，保障用户权益，实现生态增值。在版权保护方面，建立严格的版权审核机制，确保生成的文本不侵犯他人知识产权。同时，为用户提供版权登记和维权服务，增强用户对平台的信任。在数据安全方面，加强数据加密和存储管理，防止用户数据泄露。建立用户反馈机制，及时处理用户投诉和建议，不断优化服务质量。通

过责任强化设计，可以提高用户满意度和忠诚度，吸引更多用户使用平台服务，从而增加平台收入。例如，某平台在加强版权保护后，用户投诉率降低了30%，用户续费率提高了20%。

6.3 条线资源整合能力提升

6.3.1 建立收放自如的数据资产壁垒

数据是“AI代写”平台的核心资产之一。平台需要建立完善的数据收集、整理和分析体系，收集多源数据，包括公开文本数据、用户生成数据、行业专业数据等。通过数据清洗和标注，提高数据质量，为算法模型训练提供优质数据支持。同时，加强数据安全管理，建立数据备份和恢复机制，防止数据丢失和泄露。

6.3.2 搭建多主体上下游生态联盟

平台可以与内容创作者、技术提供商、渠道合作伙伴等建立多主体上下游生态联盟。与内容创作者合作，获取优质原创内容，丰富平台的素材库；与技术提供商合作，引入先进的技术和算法，提升平台的技术实力；与渠道合作伙伴合作，拓展销售渠道，提高平台的市场覆盖率。

6.4 文心一言在AI代写领域的风险识别、控制与应对机制

6.4.1 事前风险防控机制

在AI代写业务开展前，文心一言需结合自身商业模式识别潜在风险。政策法规方面，其依托百度生态覆盖多类用户，若学生利用其代写论文，既违反学术诚信，也可能因平台未明确禁止而陷入合规争议，毕竟其服务协议中版权归属条款若与代写内容挂钩，易引发法律纠纷。市场竞争上，作为百度旗下产品，虽有技术积累和生态加持，但面对豆包的多语言交互优势、Kimi的垂直领域深耕，可能导致用户流向竞品。技术层面，其模型训练依赖百度海量数据，若数据中存在错误信息，会导致生成内容准确性不足；同时，用户使用代写服务时输入的私密信息，若因数据安全漏洞泄露，会严重冲击付费用户信任。用户需求方面，其针对普通个人、专业人士、企业用户的分层服务，若对企业用户的定制化代写需求如行业报告理解偏差，会影响企业版套餐的续约率。

因此，组建专业法务团队是必要之举。需结合相关法律，在服务协议中明确禁止代写学术论文等违规内容，尤其针对企业用户，要在合同中细化定制内容的版权归属与使用范围。加强市场调研，对比竞品定价与服务，优化自身分层付费体系，比如在专业版中增加行业专属模板，突出百度搜索数据实时整合的优势。技术研发上，投入资源优化模型算法，利用百度智能云的算力提升内容生成逻辑性，对企业用户的敏感数据采用端到端加密，定期开展安全测试，将数据泄露风险纳入付费服务的安全承诺。通过用户画像分析，精准把握不同层级用户需求，例如为企业用户配备专属客服，提前沟通代写内容的细节要求，确保服务贴合预期。

6.4.2 事中风险监控机制

在业务运营中，文心一言需结合其服务流程实时监控风险。内容质量上，其智能审核系统需重点监测付费用户生成的代写内容，比如企业用户付费生成的营销文案若存在虚假宣传表述，会影响企业客户声誉，进而反噬平台；普通用户用基础版生成的文案若出现抄袭，会引发版权纠纷。用户行为方面，部分用户可能利用其按次付费模式代写作业，或企业用户借助定制服务生成违规商业文件，这些行为若未及时拦截，会让平台面临监管风险。技术运行中，其服务依赖百度服务器，若因用户集中使用代写功能导致服务器负载过高宕机，会影响付费用户的使用体验，尤其是订阅用户可能因服务中断要求退款。市场反馈方面，用户对不同付费套餐的投诉需重点关注，如专业版用户投诉内容原创度不足，会影响套餐的续订率。

因此，需升级智能内容审核系统，结合人工复核，对生成内容进行实时监测。针对企业用户的定制内容，安排行业专家审核；对普通用户的代写内容，接入百度的反抄袭数据库，一旦发现问题立即触发修改提示。用户行为分析系统要重点识别高频次、相似主题的代写需求，如检测到多次生成“课

程论文”相关内容，立即弹窗提醒违规后果并拒绝服务。技术监控体系需与百度智能云联动，实时监测服务器负载，当代写服务并发量过高时，自动扩容资源，同时设置故障预警，确保服务中断不超过1小时。设立专门的用户反馈通道，对付费用户的投诉实行分级响应，企业用户投诉2小时内对接专属客服，普通用户投诉12小时内给出解决方案。

6.4.3 事后风险应对机制

当风险发生后，需结合文心一言的商业模式特点明确应对策略。若出现用户利用其代写违规内容被曝光，引发舆论危机，会直接影响企业版套餐的销售。此时应立即成立危机公关小组，联动百度公关团队发布声明，强调平台对违规行为的零容忍，公开封禁涉事用户账号的处理结果，同时说明已升级审核系统，对企业用户的定制内容增加人工核验环节，以挽回企业客户信任。技术故障导致服务中断时，技术团队需借助百度智能云的备用服务器快速恢复服务，针对订阅用户，自动延长服务期限；对按次付费用户，退还本次服务费用并赠送优惠券，减少用户流失。内容质量问题引发投诉时，如专业用户投诉生成的学术摘要存在错误，需立即为用户提供免费重新生成服务，将问题反馈至模型训练团队，利用用户反馈数据优化模型，同时将改进结果告知用户，提升专业版用户的满意度。

事后需对风险事件进行复盘，比如分析服务器宕机的原因，优化资源调度机制；总结违规用户的行为特征，升级识别算法。将复盘结果应用于服务优化，例如调整付费套餐的服务保障条款，在企业版中增加“内容合规险”，增强商业模式的抗风险能力。

7 未来图景：构建嵌入法治的可持续AI写作法商生态

图13系统阐述了如何将法律深度嵌入商业运营，使合规性成为平台的核心竞争力。其核心路径包括推动平台定位从“工具提供者”升级为“生态治理者”，并通过监管协同创新共建良性发展环境。图示进一步揭示了外部司法裁判与立法变革正驱动商业模式与责任体系的重构，促使合规工作从事后响应转向全周期管理，并最终内化为构建商业护城河、实现价值创造的关键。底层逻辑在于治理升级与三维治理框架的实践探索，并提出了“监管沙盒”试点、版权清算中心共建及法治化AI训练数据交易所等前沿治理机制的具体构想。

7.1 法律深度嵌入商业：合规即核心竞争力

人工智能写作服务的商业化进程正经历一场静默的革命——法律规范从外部约束逐步内化为商业模式的核心基因。这一深刻变革在司法裁判的规则输出、立法体系的制度创新与企业合规的主动重构中渐次显现，共同催生出“合规即竞争力”的新型法商生态。当技术创新与法律规制的边界日益模糊，平台的责任重构不再是被动防御，而成为商业价值的重要创造维度。

7.1.1 司法裁判重塑商业模式：从技术中立到场景化责任

司法系统通过个案裁判为行业划定行为边界，

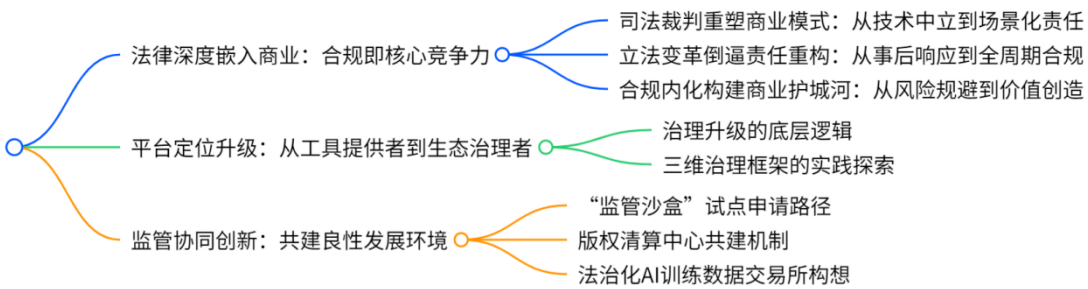


图13.“未来图景：构建嵌入法治的可持续AI写作法商生态”思维导图

将抽象法律原则转化为具体商业规则。2025年杭州中院审理的“小某书虚假种草案”即具里程碑意义。该案中，某AI写作工具因提供“小红书平台爆款笔记生成”服务被诉不正当竞争。法院创造性地提出“四要素审查法”：首先考察场景特定性，被告工具专门针对小红书平台设计文案模板；其次剖析诱导性设计，其宣传语明确鼓励用户“一键生成高赞笔记”；再次确认营利性本质，该功能需付费订阅且调用API单独计费；最终认定注意义务缺失，平台未设置虚假内容过滤机制。尤为关键的是，判决书明确区分免费与收费服务的责任标准：通用诗歌创作功能因技术中立性免于追责，但针对特定平台的付费写作服务需承担更高审查义务。这一裁判逻辑后被浙江省杭州市中级人民法院发布的《关于以高水平知识产权司法服务保障杭州建设人工智能创新高地的意见》（简称“治智16条”）司法政策吸纳，要求平台依据服务类型、应用场景、营利性质动态调整合规策略，标志着司法实践已将对法律的前瞻性内化视为商业模式可持续性的基石。

7.1.2 立法变革倒逼责任重构：从事后响应到全周期合规

立法变革则从制度层面倒逼责任体系重构。2025年新修订的《反不正当竞争法》直面AI写作商业化中的新型风险，其第13条首次赋予“合法持有数据”明确法律地位，严禁技术手段非法爬取数据。这一规定直接冲击平台的语料供应链——百度文心言被迫公示1.2亿条训练数据来源备案号，豆包协议新增用户上传内容版权担保条款。更具突破性的是第21条对平台规则滥用的规制：要求建立生成内容双标识系统，设置不正当竞争举报入口，严禁算法推荐诱导购买侵权服务。当欧盟《人工智能法案》要求通用人工智能模型公开决策逻辑，中国《生成式AI服务管理暂行办法》强制实施语料动态抽检，跨境合规成本已成为实质性的商业壁垒。法律合规由此完成从“事后灭火”到“全程防控”的范式转换，其成本结构深刻重塑着市场竞争格局。

7.1.3 合规内化构建商业护城河：从风险规避到价值创造

领先企业正将法律要求转化为差异化竞争优势，构建起以合规为内核的商业护城河。在协议架构领域，文心言会员服务推出“侵权补偿保障”，承诺因训练数据瑕疵导致的版权纠纷由平台全额赔付；豆包为企业API用户提供“事实准确性审核”增值服务；Kimi则开发“创作过程留痕”功能，自动生成内容修改轨迹报告以满足学术透明性要求。技术融合层面，双标识系统成为行业标配——DeepSeek在生成文本中同步嵌入文末标签与元数据水印，有效提升了内容溯源效率；百度“价值观过滤引擎”允许用户自定义屏蔽敏感词库，显著降低了违规内容输出的风险。更具战略意义的是商业模式创新：百度将语料合规审核系统模块化为“文心合规云”解决方案，向中小写作平台输出技术服务；月之暗面借鉴香港监管沙盒经验，在医疗文书生成场景开展受限测试并获政策豁免，率先抢占垂直市场。可以看到，具备完备合规体系的平台在用户付费意愿和风险控制方面均表现出明显优势，印证了法律内化对商业价值的创造性释放。

随着2025年10月新《反不正当竞争法》实施，法律深度嵌入将呈现更复杂的图景。平台需对第三方插件生成内容承担穿透性责任——文心言智能体功能已开始独立审核插件输出；医疗、法律等专业写作场景实施领域合规认证，豆包在医疗输出页嵌入医师审核编号以获取用户信任；跨境服务则催生弹性合规架构，Kimi开发的双模式协议可动态切换欧盟GDPR与中国个保法条款。在此进程中，合规能力已超越风控范畴，成为生存底线与增长杠杆的复合体。正如杭州中院在判词中所揭示：“当技术创新与法律规制的平衡点从冲突区转向共生带，最敏锐的企业早已将条文转化为商业密码”——这恰是法律深度嵌入商业的核心要义：在法治框架内重构价值创造逻辑，于合规基因中培育可持续竞争力。

7.2 平台定位升级：从工具提供者到生态治理者

人工智能写作服务的商业实践正经历一场根本

性蜕变——平台角色从被动的技术工具提供者，转向主动的生态系统治理者。这一转型不仅源于监管压力与市场诉求的叠加作用，更本质地反映着技术红利消退后，行业对可持续增长路径的集体探索。当算法成为信息生产的基础设施，平台的责任疆域必然超越代码边界，向规则制定、利益平衡与价值分配等深水区拓展。

7.2.1 治理升级的底层逻辑

技术赋权与风险外溢的悖论构成转型原动力。免费时代“技术中立”的护城河在收费场景中迅速瓦解：当用户支付溢价购买“论文降重”“合同起草”服务时，平台再难用“算法黑箱”推诿内容责任；API商用接口使第三方开发者批量生成营销文案，却导致虚假广告在社交平台泛滥；Kimi等平台开放的智能体商店，更让违规内容生产呈链式扩散。这些乱象倒逼监管重拳出击——欧盟《数字服务法案》（DSA）将大型AI平台定义为“看门人”，要求承担风险防控主体责任；中国《生成式AI服务管理暂行办法》第15条则明令平台建立“生成内容管理机制”。法律合规成本的高企，促使企业重新审视定位：被动防御已不足以化解系统性风险，唯有主动构建治理生态，方能实现商业可持续性。

7.2.2 三维治理框架的实践探索

领先平台通过规则重构、技术赋能与利益再平衡，初步构建起生态治理框架：

规则治理层面，平台正从用户协议制定者升级为生态规则缔造者。各主要平台通过开发者协议和服务规则，建立对第三方开发者的管理体系。这些体系通常包含多层约束：基础层设定内容安全底线要求，进阶层明确数据使用规范，卓越层则对高价值场景设定更严格标准。行业也在探索建立内容价值分配机制，通过技术手段实现收益的合理分配和风险的有效防范。此类设计将法律义务转化为可执行的运营规则，使平台从责任承担者蜕变为规则仲裁者。

技术治理层面，审核能力从封闭功能转化为开放基础设施。部分平台向生态伙伴提供内容审核技术支持，其多模态风险识别模型可检测多种违规内容形式，帮助开发者满足合规要求。还有平台构建复合型技术防护体系[20]：生成内容实时存证，隐形标识嵌入内容，建立快速溯源机制。这些技术方案的输出大幅降低全行业合规成本，行业报告显示，采用专业审核系统的开发者侵权投诉量显著下降，平台由此获得生态治理的技术话语权。

利益治理层面，平台探索版权清算与数据价值分配新机制。面对训练数据版权争议，行业正在探索建立版权协作机制，吸纳出版集团、学术机构与内容创作者参与。平台作为中介方完成多重匹配：需求端提交语料使用申请，供给端设定授权条件，通过技术手段执行分账。这种机制在特定领域取得进展，部分版权方通过授权特定内容资源获得可观收益。更深刻的变革在于数据价值分配——有平台探索建立用户贡献激励机制，用户输入经脱敏处理后用于模型优化，其贡献可获得相应回报。这种将法律要求的”知情同意”转化为经济激励的设计，使合规真正成为生态共治的纽带。

7.3 监管协同创新：共建良性发展环境

人工智能写作服务的可持续发展，亟需打破“监管滞后-商业试错”的恶性循环。通过“监管沙盒”试点申请路径优化、版权清算中心共建机制探索、法治化数据交易所构建的三维协同，正推动形成“风险可控的创新走廊”，为AI写作产业开辟“法商共治”的新生态。

7.3.1 “监管沙盒”试点申请路径

监管沙盒机制已从金融领域延伸至AI写作场景，形成协同管理架构。企业申请进入沙盒需提交测试方案，核心要件包括风险自评估报告、消费者保障计划和熔断机制。从实践案例来看，这种机制展现出一定灵活性，允许平台根据测试反馈调整参数并优化服务。但协同机制仍存在衔接问题：不同层级的监管关注重点存在差异，导致跨区域服务企业面临协调挑战。解决方案借鉴现有经验，探索建立分层授权机制——企业在完成基础合规审查后，可获得特定场景的测试权限，这有助于提高

创新效率。

7.3.2 版权清算中心共建机制

版权清算中心的建设正在破解训练数据确权困境。行业探索的清算模式实现多个环节的创新：通过技术手段将分散的权利信息进行标准化处理，建立创作者协作机制；定价机制根据语料使用频次、商业价值和稀缺性建立弹性计价模式；分账系统实现收益的自动化分配。核心技术突破在于权利贡献度分析技术，该技术通过数字标识解析生成内容中各版权素材的权重系数，有效解决多源文本融合的权利分割难题。这种机制在部分平台应用后，通过技术手段实现内容价值的合理分配，推动创作者收益提升。

7.3.3 法治化AI训练数据交易所构想

法治化数据交易所的架构为数据要素流通提供制度保障。行业探索的数据管理系统构建多层管控体系：准入环节要求特定类型数据持相应资质文件进入，专业文本需标注质量等级；交易环节采用分阶段权益确认机制，通过技术手段保障交易安全；监管环节建立全流程信息记录，明确使用场景与保存期限。配套建立的风险防范机制按一定比例计提资金用于风险处置，实际运行数据显示，采用规范数据管理系统的机构侵权投诉量明显下降。该体系推动语料交易标准化，提升训练数据交易效率。

当前协同机制面临跨境规则适配挑战。国际监管规则要求训练数据可追溯，与国内数据出境管理规定需要协调。在特定区域试点的数据跨境管理机制正探索创新路径：在保障安全前提下允许符合条件的数据定向流动，同时建立跨境侵权联合处置通道。推动三大机制形成有机整体——沙盒降低创新合规成本、清算中心重构利益分配、交易所保障要素流通，促进AI写作产业真正构建起法商融合的可持续发展生态。

致谢

该文系第一届全国法商案例分析大赛二等奖阶段性成果，现公开发表。

参考文献

[1]吴汉东. 人工智能生成作品的著作权法之问[J]. 中外法学, 2020, 32(03): 653-673.

[2]程威. 大模型嵌入公司治理的法理省思与规范调适[J]. 东方法学, 2025, (03): 31-46.

[3]东方. 人工智能生成内容(AIGC)的治理：欧盟立法与“中国路径”——基于欧盟《人工智能法》的分析与解读[J]. 图书馆, 2025, (06): 6-12.

[4]田贤鹏, 肖智琦. 生成式AI赋能研究生科研写作的学术伦理与风险防控[J]. 现代教育技术, 2024, 34(08): 23-32.

[5]刘金波, 陈晓峰, 曾巍, 等. 人工智能与学术期刊变革”笔谈[J]. 湖北社会科学, 2025, (04): 5-26.

[6]张仟煜, 刘恺骁, 杨洁. AI辅助或代写论文拷问大学的容忍边界[N]. 中国青年报, 2024-07-08(005).

[7]马治国, 贾金润. 通用人工智能信息内容风险的伦理治理[J/OL]. 大连理工大学学报(社会科学版), 1-9[2025-08-18]. https://link.cnki.net/urlid/21.1383.C.20250811.1307.016.

[8]鞠海兵. GAI对新闻受众的影响、风险与应对[J]. 传媒, 2025, (14): 54-56.

[9]赵癸萍.生成式人工智能嵌入大学生学风教育的积极变革、潜在风险及其应对[J].思想政治教育研究,2025, 41(03): 162-168.

[10]黄静, 韩松言, 田宇航. 生成式人工智能深度伪造风险的样态特征、生成逻辑与监管策略[J]. 电子政务, 2025, (05): 31-41.

[11]张平. 生成式人工智能实现突破创新需要良法善治——以数据训练合法性为例[J]. 新经济导刊, 2023, (08): 26-28.

[12]王迁. 三论人工智能生成的内容在著作权法中的定位[J]. 法商研究, 2024, 41(03): 182-200.

[13]崔国斌. 人工智能生成物中用户的独创性贡献[J]. 中国版权, 2023, (06): 15-23.

[14]朱阁, 崔国斌, 王迁, 等. 人工智能生成的内容（AIGC）受著作权法保护吗[J]. 中国法律评论, 2024, (03): 1-28.

[15]刘云开. 人工智能生成内容的著作权侵权风险与侵权责任分配[J]. 西安交通大学学报(社会科学版), 2024, 44(06): 166-177.

[16]蔡翠红. 推动构建人工智能全球善治新范式[J]. 国家治理, 2025, (14): 57-65.

[17]昂格鲁玛. 论大型内容生成式人工智能模型设计者及提供者的侵权责任[C]//上海市法学会. 《上海法学研究》集刊2023年第6卷——2023年世界人工智能大会青年论坛论文集. 北京化工大学;, 2023: 246-252.

[18]张晨颖. 反垄断智慧监管的理据与图景[J]. 探索与争鸣, 2024, (05): 102-112+179.

[19]高富平, 张启航. 可信AI：人工智能法律治理的内在逻辑与实现路径[J/OL]. 学术探索, 1-11[2025-08-18]. https://link.cnki.net/urlid/53.1148.C.20250709.1742.004.

[20]高翔, 闫钰琪. 我国生成式人工智能服务的复合型监管治理框架及其优化——基于政策文本和深度访谈的混合研究[J]. 电子政务, 2025, (05): 2-15.