



西南政法大学
Southwest University of
Political Science & Law

中国金融黑灰产治理 研究报告（2025）

——非法代理维权的识别标准与溯源治理

马上消费研究院

马上消费研究院

西南政法大学

中国消费金融法治研究中心

2025年

编委会

主任：孙磊

副主任：王春来 白恒斌 赫建营

编委：代辉 卞朝永 钱辉 钟坤玲

王梦汐 曹磊 易世洪 陈兴俊

杨学燕 李云峰 刘文君 张燕雨

学术顾问：岳彩申 蒋宁 曹景泉

马上消费研究院

摘要

“十五五”时期，推动金融高质量发展是推进中国式现代化的关键引擎，维护金融安全更是关乎经济社会稳定发展的核心战略命题。然而，当前金融黑灰产的猖獗态势正严重侵蚀金融根基，危害金融安全。一方面，其对人民群众财产安全和金融机构稳健运营构成直接威胁；另一方面，对金融市场秩序与金融安全防线形成系统性冲击。数据显示，2025年一季度金融黑灰产市场规模已突破2800亿元，其中，以非法代理维权为典型的违法活动，正借助流媒体平台的传播优势加速扩张，呈现出隐蔽性、针对性、暗示性与复杂性交织的特征。这些黑灰产组织通过引流获客、社群运营、伪造证据、恶意投诉、舆论操控、隐匿事实等系统化运作手段实施欺诈，导致金融市场秩序严重紊乱。以2024年为例，全国消协组织受理的金融保险类投诉量同比激增13.4%，黑猫投诉平台上金融类投诉量突破70万笔，其中因非法代理维权引发的恶意投诉占比居高不下，凸显问题严峻性。

面对这一困局，马上消费构建起多维治理体系。在基础建设层面，创新性地建立了网络金融黑灰产识别标准，将其细分为三大维度、八个类别，为切断信息传播链条、实现精准风险管理和协同治理筑牢制度根基；在技术应用层面，马上消费搭建起智能化技术治理体系，通过自研黑灰产关键词

提取与泛化模型、举证洞察模型、套路识别模型，显著提升了对黑灰产的认识、溯源和处置效能。以黑灰产套路识别模型为例，在非法代理维权场景下，其识别准确率超 90%，较传统模型识别效率提升 40%，展现出强大的技术治理优势；在治理策略层面，形成“对内固本、对外协同”的综合治理框架，对内实施全流程风险管控，对外开展溯源打击与生态净化行动，实现治理效能最大化。

尽管治理工作已取得阶段性成效，但仍需自上至下统一管理，多方协同联防联控。当前构建系统全面的治理体系面临多项挑战，包括但不限于法律法规存在滞后性与空白点，监管资源与力量配置不足，公众金融知识储备欠缺，治理手段相对单一，新型攻防技术博弈被动等。针对这些问题，建议从以下方面深化治理，一是加快完善法律法规体系，填补法律空白；二是强化跨区域、跨平台协同监管，提升监管效能；三是加强金融知识普及教育，增强公众风险防范意识；四是加大技术研发投入，推动技术创新联盟建设；五是健全行业自律机制，构建长效治理体系。以多措并举形成打击金融黑灰产的强大合力，保障消费者与机构合法权益，切实维护金融安全秩序与稳定。

目 录

摘要.....	1
导语.....	2
一、2025 金融黑灰产演变趋势.....	3
二、非法代理维权的网络传播特征.....	6
(一) 非法传播运作模式.....	6
(二) 非法传播四大特征.....	9
三、非法代理维权的网络识别标准.....	11
(一) 构建识别标准的必要性.....	12
(二) 网络黑产识别标准分类.....	13
(三) 识别标准应用实践总结.....	15
四、非法代理维权的技术治理体系.....	15
(一) 技术治理体系概要.....	16
(二) 黑灰产关键词提取&泛化模型.....	18
(三) 黑灰产举证洞察模型.....	21
(四) 黑灰产套路识别模型.....	23
五、非法代理维权的综合治理逻辑.....	27
(一) 对内：风险全生命周期管理.....	27

(二) 对外：溯源治理与生态净化	28
六、非法代理维权的溯源治理挑战	32
(一) 全面系统监管难度较大	32
(二) 复杂虚假信息辨别困难	33
(三) 溯源治理效果难以持久	34
(四) 新型攻击技术对抗滞后	36
七、非法代理维权的溯源治理建议	37
(一) 健全法律法规建设，强化执行宣传力度	37
(二) 提升纵横监管质效，强化网络平台责任	38
(三) 提高公众防范意识，构建全民监督机制	39
(四) 加大技术研发强度，强化配套资源系统	40
(五) 建立长效治理机制，灵活完善治理策略	42

导 语

“十五五”时期，在我国全面建设社会主义现代化国家新征程中占据着关键地位。这一时期，我国金融领域需在前期发展成果的基础上实现新的重大突破，以此加速推进中国式现代化的进程。而实现这一目标的关键在于，必须妥善统筹好金融发展与金融安全这两大核心要素，从而稳健地迈向金融高质量发展的全新征程。

依据《中华人民共和国国家安全法》第二条的明确规定，国家安全范畴广泛，涵盖政治、经济、金融、数据等 16 个重点领域。其中，金融安全因其直接关联国家经济命脉，在整个国家安全体系中具有举足轻重的地位。习近平总书记更是深刻且精准地指出，金融安全是国家安全至关重要的组成部分，同时也是经济能够平稳健康发展的关键基础。维护金融安全，已然成为关系我国经济社会发展全局，具有战略性与根本性意义的重大事项。正所谓“金融活，经济活；金融稳，经济稳”，金融领域的稳定与繁荣，对国家整体经济的稳健运行起着决定性作用。

然而，当前金融领域面临着一个严峻挑战——金融黑灰产的肆虐。金融黑灰产犹如隐藏在金融体系中的毒瘤，严重侵害着人民群众与金融机构的合法权益。它们通过各种欺诈手段，如非法代理维权等，使无数民众辛苦积攒的财富瞬间化为乌有，

给社会民生带来极大的负面影响。从金融市场管理秩序的角度来看，金融黑灰产的存在严重破坏了市场的正常运行规则。它们扰乱了金融市场的资源配置，干扰了正规金融机构的业务开展，阻碍了金融市场的健康发展。更为严重的是，金融黑灰产对国家经济与金融安全构成了直接威胁。其无序扩张可能导致金融风险的积累与爆发，影响国家金融体系的稳定性，进而危及国家经济的平稳运行。

据汇正财经的调查统计数据显示，在 2025 年一季度，金融黑灰产市场规模惊人地突破了 2800 亿元，相较于 2023 年呈现出约 40% 的大幅增长态势。金融黑灰产的持续扩张及其带来的严重危害，引发了国家层面的高度关注与重视。2025 年 3 月 14 日，公安部经济犯罪侦查局与国家金融监督管理总局稽查局果断联合行动，部署开展了一场为期 6 个月的金融领域“黑灰产”违法犯罪集群打击专项工作。此次行动旨在依法严厉查处金融领域“黑灰产”突出违法犯罪行为，彰显了国家打击金融黑灰产、维护金融秩序的坚定决心。

为了更为深入且有效地推进金融黑灰产治理工作，马上消费研究院与中国消费金融法治研究中心携手合作，共同发布了《中国金融黑灰产治理研究报告 2025——网络金融黑灰产识别标准与溯源治理》。该报告通过运用专业的研究方法和丰富的实践经验，深入剖析了金融黑灰产的识别标准。这有助于金融机构、监管部门以及广大民众更加准确地辨别了以非法代理

维权为代表的金融黑灰产的各种伪装形式，及时发现潜在的风险。同时，报告在溯源治理方面也进行了全面且深入的研究，从源头上探寻金融黑灰产滋生的原因，提出了一系列针对性强且切实可行的治理措施。这份报告对于切实保护金融消费者合法权益具有重要的指导意义，能够帮助消费者更好地识别和防范金融黑灰产的侵害，保障自身财产安全。在维护金融管理秩序方面，报告所提出的建议和措施能够为监管部门提供有力的决策参考，助力监管部门加强对金融市场的监管力度，净化金融市场环境。从宏观层面来看，该报告以高水平安全护航金融高质量发展，为我国金融领域在“十五五”时期实现稳健发展提供了极具价值的借鉴和参考。

一、2025 金融黑灰产演变趋势

根据《中国金融黑灰产治理研究报告 2024》，金融黑灰产，全称金融黑色及灰色产业，是指与金融领域相关的，利用非法手段谋取利益，行走在法律边缘或者明显违反法律法规的一整套搅乱金融市场秩序的“产业链”。非法代理维权是当前金融黑灰产的重要形式之一，主要围绕异常维权、反催收、征信修复、债务指导等开展业务，对个人与机构的信息和财产都带来极大危害。据威胁猎人数据显示，2024 年上半年我国黑灰产从业人数已超过 427 万人，假设全年保持增速一致，粗略测算 2024

年从业总人数将超过 800 万人，年复合增长率（CAGR）约为 87%。

在互联网经济浪潮下，非法代理维权广泛依托流媒体平台进行市场营销与规模扩张，产生了有组织、跨地域、隐蔽性强的违法行为。《中国消费者报》根据舆情监测系统搜索显示，2024 年 1 月 1 日至 12 月 31 日，在抖音、快手、微博、小红书、微信公众号、淘宝等多个平台上，与“债权优化”“征信修复”“反催收”“投诉代理”等相关的信息达 15 万条。据中国工商银行《2023 网络金融黑产研究报告》统计，2023 年对涉及银行业的营销活动攻击情报数日均超过 700 条，年累计 52.8 万条。而根据巨量引擎统计，有关征信修复、停息挂账、反催收等关键词的生产和搜索显著增加。例如，近四年来，停息挂账的内容分增长超过 417 倍，搜索分增长超过 209 倍，意味着黑灰产生内容在平台引流的行为和客户主动搜索相关需求的行为均呈现倍数级增长。

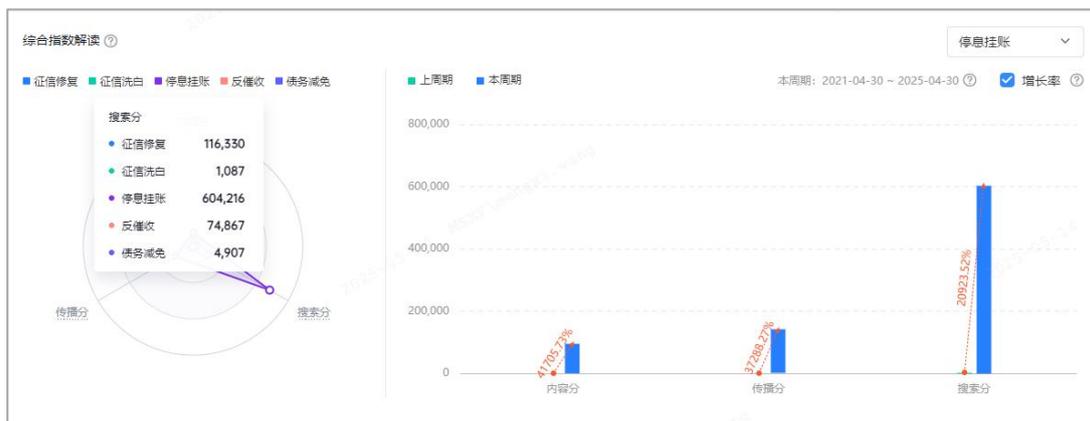


图 1 巨量引擎关于“代客维权”关键词统计

这些黑灰产行径的最终导向通常是向监管投诉，并给金融机构施压，以减免息费，同时收取高昂的咨询费。根据《中国金融消费者投诉指数年度报告 2024》显示，2024 年全国消协组织共受理金融保险类投诉 11078 件，同比增长 13.4%。黑猫投诉平台上，2024 年金融类投诉量超过 70 万笔，与上一年相比，增长数倍，同时强调，非法代理维权经常利用投诉平台进行不当投诉，据统计，重复投诉占比约占 30%，叠加恶意投诉、虚假投诉等，数量可能超过 30 万笔。

此外，以“卖征信换钱”，“背债”为主题的广告热度也在持续上涨。2024 年，威胁猎人捕获“背债”相关攻击情报呈上升趋势，背债热度上升，并在 9 月达到峰值，下半年比上半年增长 56%。黑产在选择背债人选的时候，逐渐更偏向于选择征信条件好、资质好、外貌条件好的客户。很多金融机构现有的反欺诈模型、评分卡等很难识别这类优质的高风险客户，自动化审批通过的背债人占比也越来越高。



属性	2024年新变化	变化前	变化后
属地性	属地匹配要求高	本地户籍目标客户少， 以全国户籍为主	优先选择本地户籍（当地户籍或省内 户籍人员），全国户籍为辅
年龄	年龄呈现年轻化	以中老年为主，年龄区 间在 40-60 岁	以青年、中年为主，年龄区间在 25- 45 岁更受欢迎
资质	个体资质条件好	三无人员，无资质要求	偏向自带资质条件的客户，最好名下 有房或有车或自身为企业主的群体
形象	注重个体形象气质	对形象外貌无要求，活 人即可	外貌、气质与精神形象成为新标准

图 2 2024 年背债风险情报及背债人属性，来源威胁猎人

二、非法代理维权的网络传播特征

（一）非法传播运作模式

非法组织利用互联网的传播范围与效率优势，形成了有组织有流程的运作模式，通过虚假宣传、社群操控、伪造材料、舆论炒作等手段实施欺诈。一是**引流获客**。以虚假宣传与误导性广告，非法组织在微信、微博、抖音、快手等平台发布“百分百解决债务”“征信修复”“包胜诉”等虚假广告，利用用户焦虑心理诱导咨询。同时，他们在搜索引擎、淘宝、闲鱼等平台购买“征信修复”“债务协商”等关键词，将受害者引流至私域 QQ 群、微信群。更隐蔽的，则仿冒正规律所或金融咨询机构，伪

造资质开设虚假网站、公众号，以专业形象骗取信任。

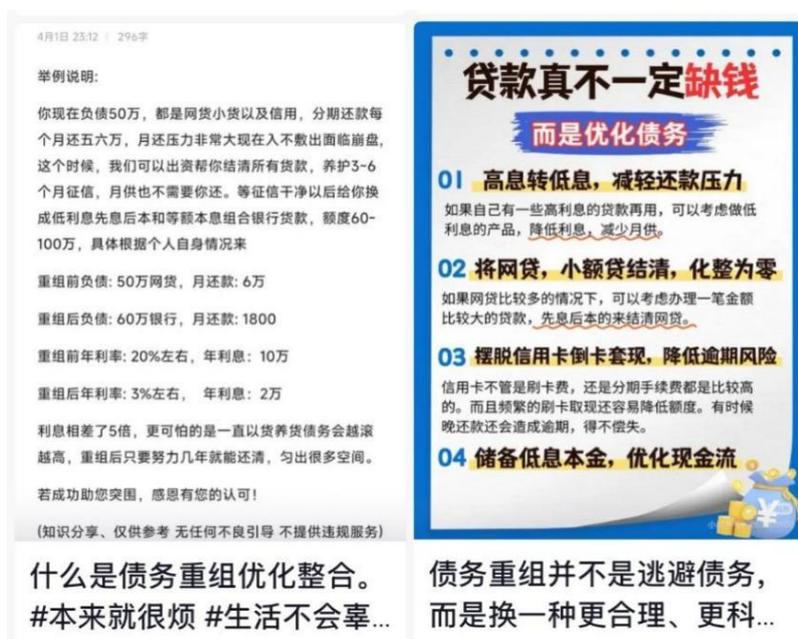


图 3 小红书非法代理维权图文广告

二是社群运营。一旦用户上钩，便将其拉入微信群、QQ群，以“维权交流”“技巧分享”“经验传授”“法务咨询”为名进行洗脑，以“专业法务团队”“内部关系”为噱头，承诺“不成功不收费”，要求用户支付定金或购买“维权教程”或“代理服务”。部分团伙甚至通过直播平台传授“反催收技巧”，教唆伪造贫困证明、病历等材料，并虚构“成功案例”截图，制造虚假可信度。



图 4 非法机构在微信等私域群发送广告

三是伪造材料。非法组织在宣传中展示伪造的“投诉回

执”“协商协议”等文件，以证明“维权”效果。部分团伙直接代开虚假病历、失业证明，甚至 PS 篡改合同、录音，捏造“暴力催收”证据。

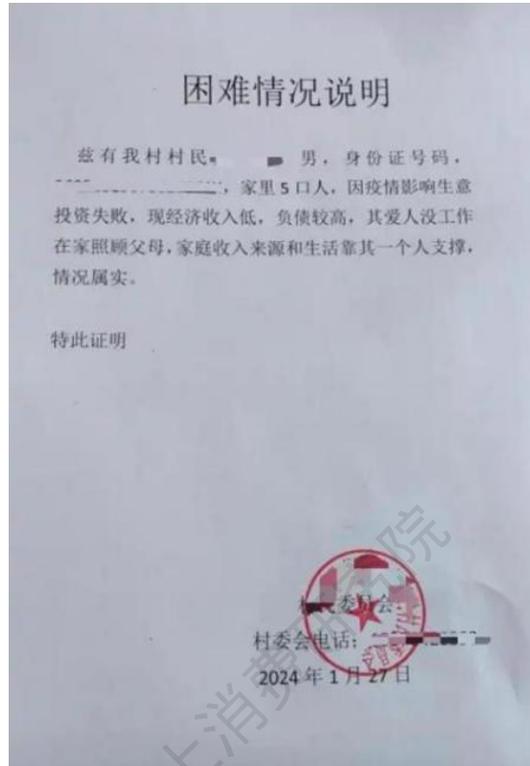


图 5 非法金融中介孙某团伙伪造的困难证明，上海市公安局供图

四是恶意投诉。非法组织会主动提供标准化投诉模板，怂恿用户以教授的投诉技巧或虚假材料向银保监会、央行或金融机构提交不实投诉。并教唆消费者反复拨打监管部门投诉热线或重复提交投诉，干扰正常处理流程。更恶劣的，则煽动群体性投诉，组织多人围攻同一机构，制造监管压力。**五是舆论操控。**为施压金融机构，非法组织在贴吧、知乎、黑猫投诉等平台散布不实信息，雇佣水军刷评转发，扩大负面影响。同时煽动消费者采取聚众闹事、网络曝光等方式制造舆论压力。此外，

还剪辑片面“维权”视频，通过自媒体炒作，误导公众舆论，逼迫机构妥协。**六是隐匿事实。**为逃避打击治理，非法组织使用虚拟账号、加密货币收款，并频繁更换平台。以“泄露隐私”“起诉追责”等威胁消费者持续配合投诉，若中途退出则骚扰其亲友或曝光个人信息。部分高阶团伙甚至通过暗网交易“维权教程”或贩卖公民个人信息，形成完整黑色产业链。



图6 非法代理维权运作流程

（二）非法传播四大特征

在传播过程中，黑灰产乱象呈现出多种特征。一是隐蔽性强，非法组织跨地域、跨领域、有组织，擅于利用网络进行自我伪装，隐匿真实身份，以虚假信息 and 话术暗语混淆视听，钻法律空子，避免被第一时间反诈监测，使机构和公安难以顺藤摸瓜寻到线索。二是针对性强，非法组织通常瞄准特定目标人

群进行攻击，尤其是弱势、长尾群体。迫于个人生存发展和家庭养育压力，这部分人群存在强烈的资金诉求用于日常应急周转，但金融素养不强，就业收入不稳定，容易剑走偏锋，易受非法组织蛊惑陷入诈骗困局。三是暗示性。非法组织运用“诈骗心理学”，利用受害人的贪心、好奇、恐惧和轻信等心理，通过具有刺激性、煽动性的语言制造强烈的心理效应，同时潜移默化地影响人们的思维和判断，并进一步发挥登门槛效应，持续坑害消费者。四是复杂性。非法组织为赚取暴利，作案技巧会随着信息更新、技术迭代、管制压力、宏观形势等实时变化，表现狡猾多变。仅在网络营销阶段，暗语就有成百上千种形式，比如征信、z信、zx，以此来规避算法的关键词监控，确保信息能够覆盖更广人群，令人防不胜防。

案例 以债务优化为名伪造证明材料，诈骗金融机构资金

2025年3月，上海市公安局经侦总队“蓝鲸”护企工作站民警在走访金融机构时了解到，有一批借款人在借款逾期、面对催收时，以自身罹患重疾、家庭陷入困境等为由，提出延长还款期限、减免利息乃至本金的诉求。如遭拒绝，或金融机构提出的偿还方案未达预期时，借款人便频繁、恶意进行不实投诉，以逼迫金融机构最大限度满足其减免本息的要求。然而，金融机构发现，这些毫无关联的借款人提供的证明材料存在高度一致的模板化特征，疑似所谓“债务优化”职业黑中介的统一操作。

经深入调查，警方发现这些借款人相关材料很可能系伪造，且幕后都指向了同一个人——上海市某法律咨询公司负责人孙某。据此，经侦总队指导上海市公安局宝山分局开展立案侦查，很快锁定了一个以帮助借款人优化债务为名，通过伪造证明材料诈骗金融机构资金的犯罪团伙，并于 2025 年 4 月将孙某等 5 名犯罪嫌疑人抓获归案。

经查，犯罪嫌疑人孙某并无律师从业资格，却为了牟取非法利益，开设法律咨询公司，对外谎称是专业法律工作者，为了混淆视听、误导客户，还刻意租借与律师事务所同一楼层的办公场所。其间，孙某通过社交媒体、互联网平台及线下渠道推广引流，雇佣倪某等 4 人利用“专业债务优化、停息挂账、减免罚息、轻松分期”等广告话术对外公开招揽客户，按照逾期金额的 5%至 10%收取高额手续费，并签订“停息挂账”服务合同，通过为逾期人员伪造虚假证明、反复恶意投诉金融机构等手段，逼迫金融机构重新协商还款协议、骗取金融机构减免贷款本息。

来源：上海市公安局

三、非法代理维权的网络识别标准

非法组织通过人工方式或技术手段操纵网络信息内容，以“代理维权”为名骗取受害消费者信任，进而寻机创造违法收益。因此，对网络信息内容的甄别至关重要，而构建识别标准是甄

别的前提条件。

（一）构建识别标准的必要性

一是斩草除根：追根溯源、源头治理。当前对网络金融黑灰产的甄别缺少法律界定和统一标准，因此相关机构难以开展治理行动。构建识别标准是启动源头治理的首要凭证。一方面，机构或平台借助识别标准中的关键字段，能够像“精准滤网”一般，快速筛选出由黑灰产组织牵头发布的各类广告、帖子等有害信息。这些被筛选出的信息，将报备给网信办等相关部委进行确认与删除，从传播源头切断黑灰产的信息扩散渠道。另一方面，通过对筛选得到的广告、帖子进行IP信息的二次识别，如同“数字追踪器”，能够精准定位到非法组织的物理方位。这不仅为打击行动提供了精确的黑名单线索，还能帮助公安机关大幅缩小打击治理范围，让执法行动更具针对性和高效性。

二是提效降本：精准获客、风险管理。识别标准的构建有助于机构或平台精准掌握黑灰产锚定客群的画像和行为，方便金融机构在导流时，依据识别标准，能够对有明确扫码等异常行为的客户进行有效规避，避免将资源浪费在风险客户身上，同时精准导入资质合规的优质客户，提升获客效率与质量。此外，对于搜索或浏览相关黑灰产广告等界面的客户，机构会将其列入观察名单。在贷前审批环节，系统会将这些客户信息与黑名单库进行二次比对，同时围绕客户收入、信用记录等多维

信息展开细致的二次排查。一旦发现存疑客户，便果断拒绝通过贷款申请，从而有效降低逾期还款甚至恶意投诉的风险，为金融机构的稳健运营筑牢防线。

三是清朗生态：高效对碰、协力共治。基于唯一识别标准构建的黑名单库是跨机构黑灰产线索对碰的核心基础，更是有效共治黑灰产乱象的前提条件，还是监管判定对机构投诉有效性的有力抓手。依据识别标准对黑灰产行为进行科学的类型划分后，监管部门、公安部门以及不同机构之间能够打破信息壁垒，共用黑名单库。通过搭建共享平台，各部门、机构可以高效地进行线索对碰，实现信息互通与资源共享。这一协同机制在反诈治理、机构降诉和联合打击黑灰产等工作中发挥着巨大作用，能够形成强大的治理合力，共同维护网络空间的清朗秩序。

（二）网络黑产识别标准分类

马上消费金融股份有限公司（简称“马上消费”）通过对海量网络信息监测，将网络金融黑灰产的识别标准简要归纳为三大类型八个类别，其中黑产类包括但不限于伪造证据、渲染违规、捏造事实、施压威胁等，灰产类包括但不限于退费技巧、债务咨询、术语暗语等，以及第三类经常与黑灰产同时出现的关联词汇。

表 1 网络金融黑灰产识别标准

一级分类	二级分类	主要字段
黑产	伪造证据	伪造贫困证明、假病历、代开诊断书、伪造录音、假暴力催收证据、假合同、伪造聊天记录、代写投诉信、假工作证明、假收入证明、假骚扰截图、假流水、伪造还款凭证、假征信报告、伪造债务协议、假法律文书...
	渲染违规	暴力催收实锤、违规放贷、无资质放款、砍头息实锤、714 高炮、黑贷款、高息陷阱、砍头息...
	曲解法规	反催收教程、停息挂账、无本续贷、无还本续贷、强制续贷、低至本金、征信洗白、征信代偿、征信撤销、征信修复、信用报告错误、ZX 修复、消除逾期、信用修复、征信记录消除、修改征信、征信代理、征信加盟...
	施压威胁	集体投诉模板、起诉模板、银监会投诉模板、信访举报模板、批量诉讼指南、爆通讯录技术、强制停还教程、威胁恐吓话术、100%停催、停催技巧...
灰产	退费技巧	本金打折技术、停息挂账话术、减免申请书、投诉信模板、信访模板、法务援助套路、债务优化陷阱、委托 T、费用减免、退息退费、申请退息、免赔额、减免本金、减免利费、本金清零、一次性结清减免、信用卡退息费...
	债务咨询	逾期咨询、债务处理、债务指导、债务调解、告别负债、负债上岸、清债规划、债务托管、代还债务、599 调解...
	术语暗语	分琪、催收、JR 消费、投素咨寻、鞋商、利熄、小崔、崔哥、可延、退息退费、秒下的贷款、必下款的口子、摆脱催收技术、f 院暗语、无前期套路、代接崔债、寨务处理、征 x 申诉...
关联黑灰产	与黑灰产字段同时出现	沟通技巧、调解方案、多年维权经验、帮你协商、帮你申请、成功上岸、保证成功、100%成功、一定成功、新规、新政、新通知、新方法、新消息、攻略、技术、协助、咨询、接单、代理、中介、专业、指导、止损、维权、正规咨询公司、专业法律团队、专业律师团队...

（三）识别标准应用实践总结

当前，马上消费以金融消保一体化服务的框架思路，将识别标准应用于日常舆情监测、处置及协同打击中。一方面通过日复一日的监测，马上消费梳理全网疑似黑灰产的网络信息内容，针对金融业务的非法代理维权、电信欺诈、虚假信息及已知行为信息，形成网络金融黑灰产舆情库。然后，利用黑灰产关键词提取&泛化模型对舆情数据进行预处理清洗、分类识别、特征抽象与归因分析，通过信息蒸馏提纯构建网络黑灰产动态关键词库。初期训练从百万级规模数据源解析抽取成训练形成黑灰产举证洞察模型。目前，已经积累六大类 35 小类黑灰产套路结构体系，开始在内部应用于反诉举证依据、分级客诉处置等业务。

另一方面，马上消费在内部原有数十种代客维权识别手段基础上，加入沉淀的黑产特征库数据，以此增强对黑灰产套路识别的泛化性，通过模型微调和对抗实测，初步在提高智能客服领域的客诉洞察精准性上有显著提效。反过来更精准地定向在全网筛选黑灰产信息内容，进而通过自主、协同、打击等多种方式进行网络环境治理。

四、非法代理维权的技术治理体系

马上消费综合应用大模型、定制化小模型与精细化微调，实现对交互行为套路的精准识别；同时，结合思维链推理提升

模型的溯源推断能力，并借助声纹识别、多模态数据等增强技术，探索建设黑灰产识别模型体系和溯源治理机制。

（一）技术治理体系概要

该体系以数据整合为根基，通过多源数据融合构建初始黑产关键词库。首先，依托 Deepseek 自动生成技术，结合行业机构脱敏数据与资深专家先验经验，搭建基础词库框架；其次，借助舆情监测平台对全网黑产相关数据实时爬取，利用大语言模型的自然语言处理能力，对海量非结构化数据开展智能清洗，有效过滤无效、重复及干扰信息，为后续分析筑牢数据基石。马上消费还构建了关键词提取&泛化模型，采用“模型+算法”双轮驱动策略，融合 Deepseek 与 TF-IDF 算法的自然语义理解和文本特征提取优势，敏锐捕捉具有黑产特征的隐性关键词。同时，建立数据反馈闭环，将新增黑产特征词汇、变体表述持续反哺词库，确保其始终保持对黑产演变的动态适应性。黑产套路洞察模型是体系的核心研判模块。该模型基于深度学习算法，对投诉数据及关键词进行多维度解构分析。通过关联分析、模式识别等技术，完成黑灰产类型识别与套路归因，同步将分析结果结构化存储，形成具有行业参考价值的舆情洞察数据库。该体系形成了四大实战场景，在举证应用场景中，搭建跨平台协作系统，依据分析结果快速锁定黑产违规内容，实现精准删帖、账号封禁，并基于处置反馈持续优化平台打击规则；自动化处置场景运用机器人流程自动化（RPA）技术，实现黑猫投

诉平台智能应答、黑产帖文人机协作举报及云测小号异常投诉检测，大幅提升处置效率；在监管协同场景下，定期编制包含黑产规模、手法演变、风险趋势等内容的态势报告，为监管部门提供决策依据；而在行业联防场景中，积极推动金融同业建立联合打击机制，通过情报共享、策略协同，形成跨机构治理合力。

与此同时，马上消费还将技术治理融入内部客服作业（CRM）场景中，针对与黑灰产的会话数据，借助黑产套路识别模型开展套路识别与归因分析；同时，将新增会话数据与历史行为、会话、账务等数据相结合，通过算法模型打分和黑产特征分类识别，构建起涵盖声纹、号码、设备、身份、行为和材料的黑产证据链。该证据链及相关分析结果主要应用于三个方面：其一，在客诉处置场景中，围绕黑产特征、套路和话术等风险类型，叠加催收、征信、会员等业务场景开展客诉诊断，先洞察分析客诉处置难点以判断谈判方向，再推荐协商策略并提供共情话术，推动和解并记录结果，而和解结果将反馈至黑产套路识别模型用于特征增强；其二，在安全处置领域，依据证据链的挖掘与生成，对代客维权行为进行研判和结果认定，进而实现立案打击；其三，在支持监管申诉工作时，历经材料准备、报告撰写、申诉汇报三个阶段后，申诉结果无论成功或失败，都将反哺给黑产套路识别模型，用于模型的强化与训练。

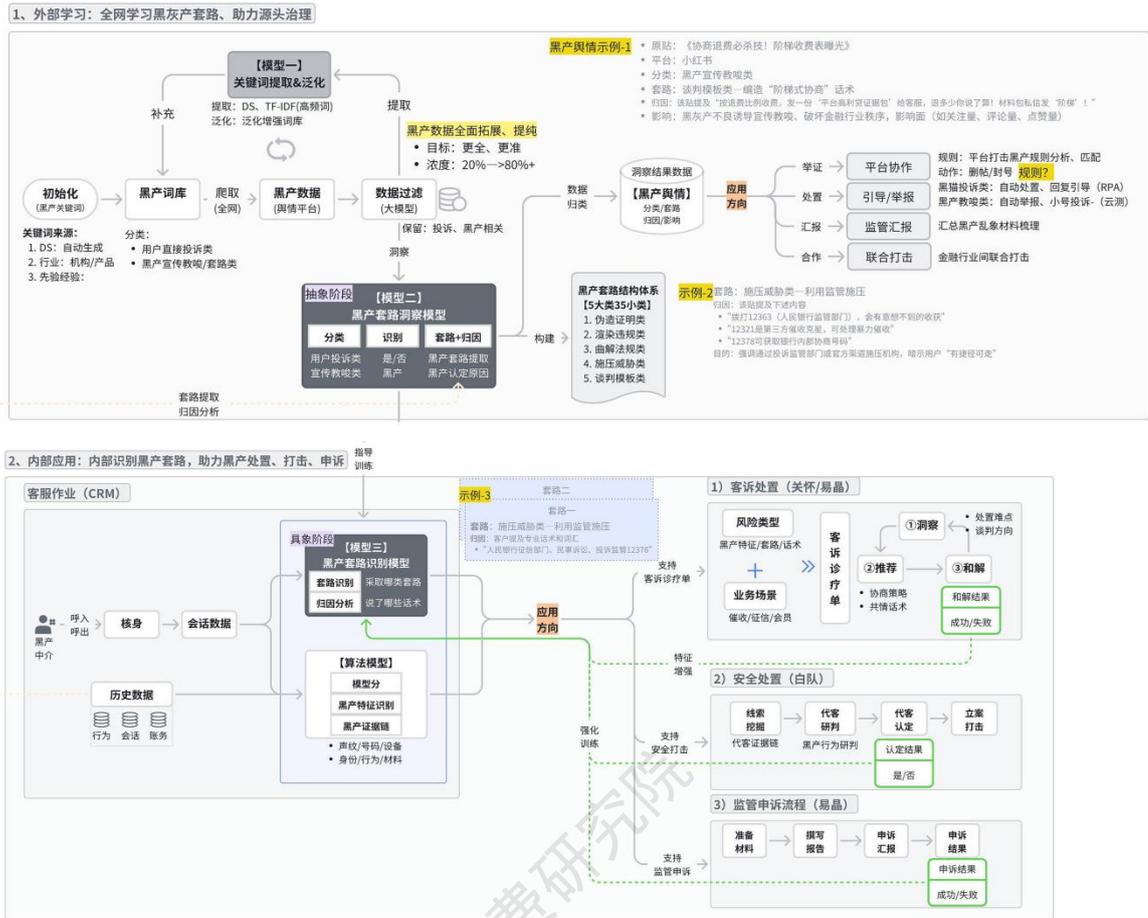


图7 马上消费技术治理体系

通过内外协同的“数据采集—智能分析—策略生成—多端处置—生态共治”闭环机制运行，马上消费实现了从被动应对到主动防控、从单点打击到生态治理的全面升级，为净化网络环境、维护行业健康发展提供了坚实保障。

具体来讲，马上消费主要围绕三大技术治理工具来应对层出不穷、复杂多变的网络金融黑灰产，即黑产关键词提取&泛化模型、举证洞察模型、套路识别模型。

（二）黑灰产关键词提取&泛化模型

聚焦及时发现与实时预警，马上消费融合前沿技术与创新策略率先研发了黑灰产关键词提取&泛化模型，对黑灰产话术字段进行精准关联与全链追踪，与此同时持续提升识别效率与泛化精准度，为洞察与深度识别奠定基础。



图 8 黑灰产关键词提取&泛化模型

在模型层面，马上消费以创新架构打造强大的分析引擎，基于 Deepseek 模型的卓越性能，创新性地采用两个任务微量模型嵌套的设计思路，构建起黑产关键词提取与泛化的核心模型。此模型如同敏锐的“数字猎手”，对黑灰产相关的网络数据、监管文件、权威行业报告等多元信息源展开实时追踪与深度解析。通过对海量信息的挖掘、提取和分类，已初步形成包含 500 个黑产“标准”关键词的基础库，这些关键词如同精准的坐标，帮助从源头精准把握黑产攻击模式。随着模型推理能力的持续迭代，其在关键词泛化能力和识别精度上取得重大突破。在实

测数据中，目标浓度从过往版本的 20%大幅提升至 80%，较上一版本实现 4 倍的性能飞跃，能够更准确、全面地识别各类黑灰产关键词。

在数据层面，依托马上消费构建的数据治理机制。秉持持续优化的理念，采用广覆盖扩展策略，以历史黑产溯源数据、先验经验和安全专家建议为词库基础，搭建起动态更新的网络金融黑灰产舆情库。这一舆情库如同不断进化的“数字大脑”，实时感知网络空间中黑灰产的动态变化，将最新的黑灰产信息及时纳入监测范围。为进一步提升识别精准度与关联度，该体系定期新增数据构建训练集。每季度都会增补至少 100 条黑灰产标准关键词，持续丰富和更新模型的知识储备。同时，考虑到黑灰产发展迅速、行为复杂多变的特点，在内部建设的黑产特征数据中，通过定期过滤机制，及时剔除低危关键词，下线失效关键词，确保数据的时效性和精准性。相比业内其他同类体系，该模型体系在数据覆盖的广度和深度上更具优势，为模型的精准分析提供了坚实的数据支撑。

在应用层面，黑灰产标准关键词的动态优化成为提升外部风险态势感知能力的关键。通过常态化监测与深度洞察，该体系能够精准捕捉潜在的风险信息，如同敏锐的“安全哨兵”，在风险萌芽阶段便发出预警。系统会将捕捉到的风险信息生成标准化的预警信息，为金融机构的安全管理提供清晰、明确的决策依据。这些基于舆情监测的成果，为金融机构内部安全管理

筑牢运营基础。面对非法中介层出不穷的攻击手法，该模型可以及时发现，显著提高了金融机构的精准识别与快速反制能力，从模型的创新架构到数据的动态治理，再到应用的实际落地，这套黑灰产关键词追踪分析模型体系形成了完整的闭环，展现出强大的技术实力和应用价值。

（三）黑灰产举证洞察模型

为了持续提升对全网黑灰产舆情洞察能力，黑灰产举证洞察模型的运行机理是从外部黑产攻击视角切入，完善对非法代理维权识别标准的探索实践。



图 9 黑灰产举证洞察模型

在模型层面，黑灰产举证洞察模型构建了精密的智能识别体系。以内部安全业务的证据链为坚实基准，如同经验丰富的侦探梳理案件线索，持续高频地与模型蒸馏出的六大类黑灰产套路样本进行比对验证。这六大类样本涵盖了黑灰产最常见的作案手法，从虚假信息传播到恶意投诉设计，模型通过不断学习和迭代，精准捕捉各类黑产行为的细微特征。该模型输出的

黑产归类数据，在客服初筛、安全分级和申述举证等多个关键场景中，成为衡量识别模型准确性与鲁棒性的量化指标。在客服初筛环节，模型能快速对海量信息进行初步甄别，减少人工筛选的工作量；在安全分级过程中，依据黑产行为的危害程度进行精准分级，为后续处置提供明确方向；申述举证时，其提供的准确数据让证据更具说服力。通过持续的反馈优化，模型的推理能力与套路归因能力不断增强，成功实现从被动应对黑产攻击的“被动防御”模式，向主动识别风险、提前布局防范的“主动防御”模式转变，极大提升了安全防护的主动性和有效性。

在数据层面，黑灰产举证洞察模型构建起覆盖全网的舆情监测网络。通过接入微博、贴吧、论坛等各类网络平台的舆情渠道源，以动态更新的黑产关键词库为“过滤网”，对不同的训练数据源进行精准归类。这些关键词如同敏锐的嗅觉探测器，能够及时捕捉黑产行为在网络上留下的蛛丝马迹。基于对海量黑产数据的深度特征分析，该模型基于图神经网络和深度学习算法，自动提取黑产群体的标签特征。同时，充分结合安全管理专家多年积累的深入分析经验，将人工智能与机器智能有机融合。目前，模型已成功提取 35 个具有代表性的套路特征，这些特征涵盖黑产的作案时间规律、常用话术、传播渠道偏好等多个维度。依托这些特征，内部构建起从最新黑灰产舆情洞察、判定黑灰产套路识别，到黑产认定归因的完整链路闭环。

在应用层面，由于各平台取证标准不统一，金融机构在界定相关问题时面临着成本高、效率低的困境。同时，生成式 AI 技术的发展，使得黑灰产能够以低成本批量生成“维权”伪证，进一步加剧了举证难度。黑灰产举证洞察模型的试点部署，为解决这些难题提供了有效方案。模型通过自动化分析和智能识别，降低了业务流程对人工判定及侧面佐证的依赖，所输出的证据具有更高的客观性和可信度。特别是在针对非法代理维权这类尚无统一标准的黑灰产行为时，模型展现出强大的能力，检测精度提升了约 80%。这一突破有效丰富了证据链，显著提升了金融机构的申诉成功率。在实际应用中，某金融机构引入该模型后，案件调解效率大幅提升，原本需要数周处理的投诉案件，如今平均处理周期缩短至一周以内，极大地提高了业务运转效率。

综上所述，针对非法代客维权等网络信息，马上消费通过黑灰产举证洞察模型，能够准确识别出金融黑灰产相关违法信息，并常态化开展投诉举报工作，减少曝光量和获客效果。据统计，2025 年 1—5 月，涉及代理维权类的金融黑灰产信息，主要发布于快手、抖音、小红书、微信四大平台，合计约占识别总量的 91.54%。

（四）黑灰产套路识别模型

非法代理维权行为因动态多变及 AIGC 深度伪造等技术特征，导致常规身份验证与风控机制难以拦截，对金融机构构成

显著压力。当前，黑产对抗周期不断缩短，攻击手段快速迭代，同时金融非法代理维权场景往往样本稀缺。马上消费自研黑灰产套路识别模型，凭借对海量数据的学习能力，可为挖掘导流表现与隐蔽套路提供深度洞察与识别，有效补强现有风控手段。

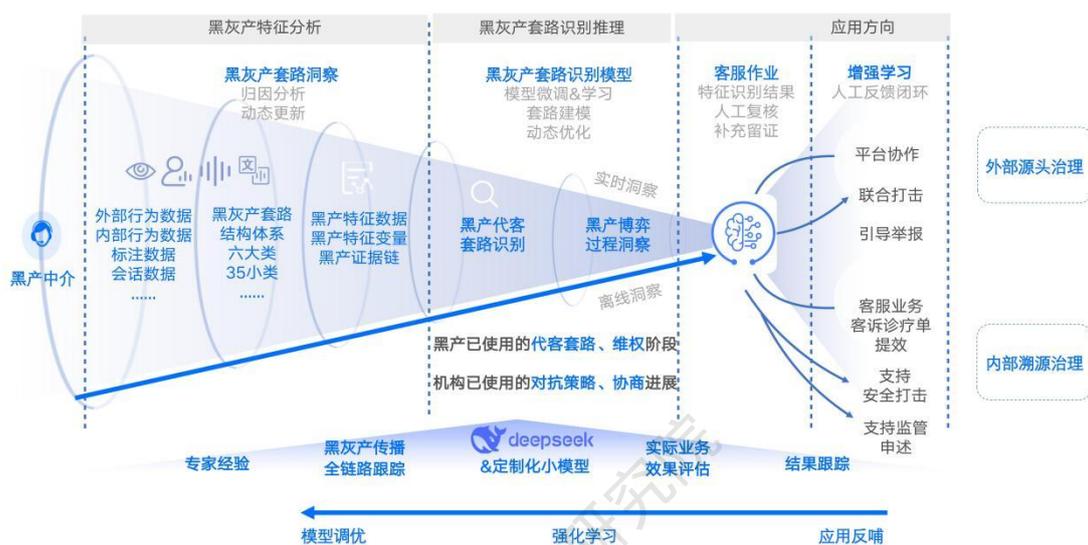


图 10 黑灰产套路识别模型

在模型层面，马上消费基于多年黑产攻防实战经验，创新性地构建黑灰产套路识别模型。该模型以自主研发的黑灰产举证洞察模型为指引，深度整合包含百万维度的黑产数据库和 35 个套路特征库、超百万级关系图谱及动态行为判定规则。借鉴医学领域“全科诊断+专科精研”的互补逻辑，将大语言模型的泛化理解能力与小模型的精准识别优势进行嵌套协同。在实际运行中，大模型如同“全科医生”，通过 Transformer 架构对海量文本、语音数据进行语义解析，能快速识别出诸如“代理全额退保”“信用卡逾期减免”等敏感关键词，并挖掘出潜在的风险语义关联。例如，当出现“有内部渠道处理债务”“提供全套

维权话术”等隐晦表述时，大模型可迅速捕捉异常信号。而小模型则像“专科医生”，针对特定套路，如“伪造征信异议申诉”“虚假消费欺诈索赔”等，通过卷积神经网络（CNN）对图像、语音特征进行深度提取，精准识别伪造痕迹。两者结合，再辅以因果推理、关联分析及模式对比三重验证机制，构建起“风险初筛—特征验证—结论确认”的三级识别链路。经实测，该模型在非法代理维权场景下的精确率稳定在 90%以上，较传统模型识别效率提升 40%。

在数据层面，黑灰产套路识别模型建立了“双循环”动态更新机制，外部按月度增补外部黑产舆情信息并与历史脱敏数据构建训练集，内部打通客服工单、交易流水、设备指纹等多源系统，通过特征工程构建包含 300+非法代理维权专属特征的变量池。特别引入联邦学习技术，在保障数据隐私的前提下实现跨机构特征共享，进一步扩充模型训练边界。同时，监管部门反馈的申述结果、司法判例数据将自动反哺模型迭代，形成“数据采集—模型训练—效果验证—策略优化”的闭环生态，持续提升模型对新兴套路的捕捉能力。

在应用层面，依托声纹识别、数字水印检测、AI 伪造痕迹分析等核心技术，马上消费将黑灰产套路识别模型深度融入智能风控全流程。在鉴伪检测环节，模型可在 3 秒内完成对通话录音、合同文件等电子证据的真伪鉴定，识别准确率达 98%。曾有黑产团伙利用 AI 语音合成技术伪造客户与客服的通话，

试图投诉平台“暴力催收”，但模型通过分析语音中的共振峰、基频等细微特征，快速识别出伪造痕迹，成功拦截了恶意投诉。智能分案系统基于风险评分实现投诉工单的“红黄绿”分级，将恶意投诉处理时效有效压缩。对于高风险的红色工单，系统自动触发紧急处理流程，调配资深客服与法务人员介入；而低风险的绿色工单，则交由智能客服快速响应，大幅提升了整体处理效率。溯源模块通过图计算技术，可快速定位黑产团伙的组织架构与作案链路，为监管部门提供完整证据链支持。在某次专项打击行动中，模型通过分析多个恶意投诉案件的关联数据，成功梳理出一个涉及全国多地的黑产团伙网络，协助警方捣毁犯罪窝点3个。与行业内其他解决方案相比，马上消费的黑灰产套路识别模型在技术融合度与场景适配性上更具优势。多数竞品仍依赖单一规则引擎或简单的机器学习模型，难以应对复杂多变的黑产攻击。而该模型不仅实现了大小模型协同、多源数据融合，还能根据实际业务场景动态调整策略，为行业应对新型金融风险提供了创新范本。随着技术的不断迭代，这一模型有望成为金融行业抵御黑灰产威胁的核心利器，提升网络金融黑灰产识别覆盖率目标达到80%以上，助力构建“预防—拦截—溯源”一体化的黑产防控体系。今年下半年，马上消费将基于三大模型助力客服业务识别留证，实现对恶意投诉案件处置提效10%以上。

五、非法代理维权的综合治理逻辑

总结行业实践，为了实现非法代理维权的统筹治理，除了专注内部技术治理体系建设外，也要聚焦与外部多方合力打击和生态净化，坚持“对内固本”与“对外协同”两手并举，二者协同形成闭环防控。

（一）对内：风险全生命周期管理

对内治理侧重于金融机构内部建立网络信息识别、防控黑灰产攻击的长效机制，涵盖了从舆情洞察识别到闭环处置的全生命周期管理。主要包括识别套路、监测分析、场景防御和闭环处置三个环节。

一是黑灰产套路识别与预警。参考上文三模型技术治理方法，对外部非法网络信息内容及数据进行整合分析、高效识别与智能决策，一方面将外部非法网络信息通过不同途径进行申诉和清洗，以减少对公共资源的占用和对正常金融秩序的负面影响；一方面将识别结果转化为贷前审批的风险预警信号，洞察风险的本质、影响范围、关联方、作案手法等，并通过对泛用户的深度洞察建立更加全面完善的黑名单，为风控贷前审批提供精准可靠的参考。

二是深化全链路闭环监测分析。建立覆盖客户触点、交易环节、服务流程等全链路的监测体系。利用大数据分析、图计算、机器学习等技术，对业务操作数据进行实时或离线洞察分析，动态识别偏离正常模式的异常行为序列或关联黑灰产群体，

为风险预警和精准处置提供实时数据支撑。针对黑灰产高发的业务场景部署专项防控策略。在银行业，典型场景如信用卡欠款催收、贷款逾期处理、营销活动奖励发放等；在保险业则有退保理赔、核保调查等。通过场景化的精细防控，提升对各类黑灰产攻击的抵御能力。

三是构建风险事件闭环处置机制。当识别到涉嫌黑灰产行为时，应联动内部相关部门，依据预设处置流程，实现风险事件的快速响应与有效干预。与此同时，需完善客户争议与纠纷的多元化解机制，对于普通金融消费纠纷，建立独立、公正的调解渠道，保障客户合法诉求得到妥善处理，以此区分并剥离黑灰产驱动的恶意投诉与非法代理维权。将风险事件的处置结果、客户反馈等信息纳入闭环评估体系，持续分析防控策略的有效性及潜在漏洞，驱动内部风险管理体系的迭代优化与能力提升。

（二）对外：溯源治理与生态净化

对外治理聚焦于金融机构面向外部环境所采取的防控措施，包括情报获取、联合打击和宣传引导等方面，其核心依托于对黑灰产的态势感知、舆情洞察能力与协同共治机制。

一是舆情采集与深度分析。金融机构可自建多源舆情监测体系，实时抓取黑灰产活动线索并进行关键词聚类 and 传播路径溯源。同时，借助图神经网络技术回溯信息传播路径，追踪黑灰产链条中的关键节点与扩散轨迹。基于舆情洞察与深度分析，

可实时完善黑名单库，将相关主体信息纳入风险管控范围，并每月形成黑灰产活动趋势分析，为后续开展精准打击行动、优化内部风控模型提供了有力的数据支持。此外，通过持续的舆情监测与分析，银行还能提前预判黑灰产新动向，如新型诈骗话术、技术手段等，做到未雨绸缪，提升风险防控的前瞻性。

二是机构合作与协力共治。基于对互联网平台黑灰产内容的精准识别与证据固定，有效举证并推动联合治理打击。例如，浦发银行与小红书平台建设联动机制，每周将人工举报未封禁的涉嫌黑灰产账号反馈平台对口人，平台复核后进行直接封禁或设定为仅作者本人可见，大幅提升治理效率。金融机构之间也应加强协作，建立行业信息共享平台，进行高效线索对碰。通过共享黑灰产案例、风险数据、可疑交易信息等，实现风险联防联控。

三是助力监管处置与执行，将外部监测的黑灰产乱象信息进行汇总、梳理与分析，定期或不定期向监管机构进行专题汇报。为监管部门了解评估行业风险、制定监管政策、组织专项打击行动提供重要的决策参考和数据支撑。此外，针对行业影响突出的非法代理组织，监管部门可及时联合公安、网信等部门开展专项打击行动和查处，有效净化市场环境。

四是引导理性认知与正向宣传。面向广大金融消费者，为揭示黑灰产诈骗伎俩，主动进行正向宣传、风险提示、金融教育和行为引导。例如，监管部门多次发布风险提示、案例汇编，

揭露非法代理维权（假冒监管、谎言恐吓、伪造证据等）。同时，建立便捷的投诉举报通道，鼓励和引导消费者针对不良发帖和可疑信息进行投诉举报，形成群防群控的健康氛围。

案例 共筑生态，AIF 联盟推动数字金融安全协同治理

面对非法反催收的治理难题，在金监总局、公安、网信、工信等党政部门的指导和支持下，打击金融领域黑产联盟（简称 AIF 联盟）积极响应政策号召，在符合相关法律法规的框架下凝聚协同力量，强化警企协作，建立联动机制，通过打击反面典型，维护法律权威，净化行业环境。同时，强化科技建设，研发基于隐私计算的黑产线索碰撞平台“爱马”，为黑灰产的联合预警和处置提供技术抓手，全方位开展联防联控。

2024 年 7 月份，AIF 联盟通过爱马平台多方隐私安全计算技术，在平台数据库内“碰撞”出黑灰产线索，联合联盟成员单位美团协助广东警方成功捣毁一个非法代理维权窝点，抓获犯罪嫌疑人 45 名，对 6 名主要人员和 1 名上游制假嫌疑人员采取了刑事强制措施。自 2022 年 4 月以来，该团伙在百度、抖音平台发布债务优化广告，吸引了 9500 余名债务人员与其签署代理维权业务合同，涉及 20 余家金融机构。

此外，由联盟理事长单位马上消费牵头研发的“慧眼识代客”黑声纹识别系统已于 2025 年投入使用，为打击和防范金融黑产探索出新的路径。目前已接入 30 家机构，上传黑灰产数据近 13 万条，形成分布式黑产池，碰撞后挖掘黑灰产线索逾

千条，协助公安机关打击违法犯罪团伙 10 个，推动警方受理打击非法代理维权案件 66 起，推动警方采取刑事强制措施 171 人，推动警方采取行政拘留处罚 21 人。

2025 年 3 月 22 日，AIF 联盟与中国政法大学刑事司法学院举办第三期金融犯罪类案研讨会在京召开。针对非法反催收行为的刑法规制问题，AIF 联盟在系列金融犯罪研讨会上与产学研各界达成共识，形成行业治理参考意见：一是将非法反催收行为纳入刑法规制要秉持审慎原则，优先通过行政监管与民事调解化解纠纷，在以刑法评价时也需考察非法反催收行为背后的复杂法律关系；二是治理非法反催收行为需要加强信息共享与协同合作机制，形成多部门治理合力；三是推动相关行业规范、司法解释和立法的完善是长远之计，以期为司法实践提供更加明确和具体的依据。

截至 2025 年 5 月 16 日，AIF 联盟总计 146 家单位，包括邮储银行、民生银行、江苏银行、浦发银行、平安银行、腾讯、字节跳动、饿了么、兴业消金、中信消金等机构；依托联盟联合各方，协助各地警方受理非法代理维权案件 655 起，对当事人采取刑事强制措施和行政处罚措施 571 名，有效预警电信网络诈骗案件 151.22 万起，协助警方追逃 123 人，收到各地公安机关感谢信 48 封。

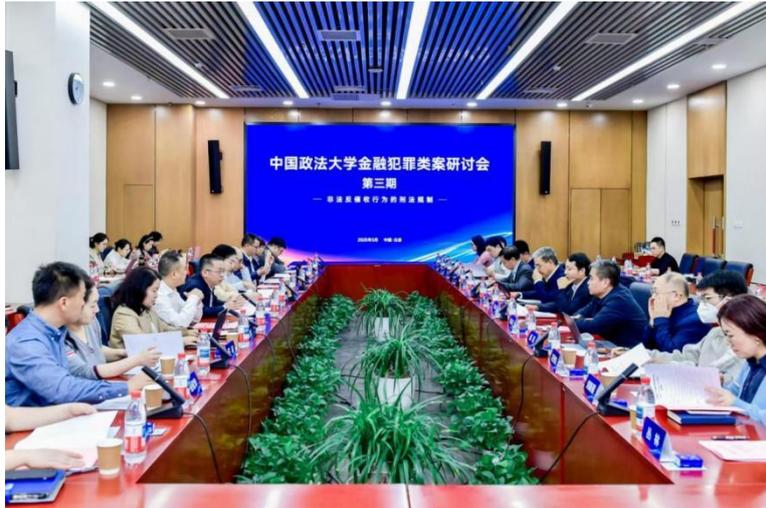


图 11 2025 年 3 月 22 日, AIF 联盟与中国政法大学刑事司法学院举办第三期金融犯罪类案研讨会在京召开

总的来说,依据黑产套路新分类,金融机构可以调整营销规则与风控阈值,堵塞漏洞;针对非法代理投诉高发环节,可以优化客服及催收策略,规范话术与操作流程,实现防御能力与合规水平双提升。通过内外协同的综合治理体系,不仅可以健全内部业务风险管理机制,以需求为导向为外部模型的底层逻辑提供参考,而且可以从外部舆情洞察反哺金融机构内部业务,同时赋能行业联防打击和监管处置执行。

六、非法代理维权的溯源治理挑战

(一) 全面系统监管难度较大

一是法律法规不完善。目前,针对非法代理维权网络信息的相关法律法规尚存在诸多空白和模糊地带。在网络信息传播的监管方面,对于非法代理维权信息的界定、发布者的责任追究、信息传播平台的义务等缺乏明确细致的规定。这使得监管

部门在实际执法过程中面临无法可依、执法尺度难以把握的困境。例如，对于一些打擦边球的非法代理维权信息，很难依据现有法律准确判断其是否违法，导致不法分子有机可乘，肆意传播此类信息。

二是监管力量不足。网络空间的信息海量且繁杂，非法代理维权信息更是分布在各个角落，监管部门有限的人力、物力难以对其进行全面、有效地监测和管控。同时，随着互联网技术的不断更新迭代，非法代理维权信息的传播手段和渠道日益多样化，监管部门在技术能力和专业知识方面也面临巨大挑战。例如，一些新兴的社交平台和小众网络社区，监管部门可能尚未建立有效的监测机制，导致非法信息在这些平台上肆意传播而未能及时发现和制止。

三是跨区域、跨平台协调困难。非法代理维权网络信息的传播往往不受地域和平台限制，一条信息可能在短时间内跨越多个地区、多个网络平台传播扩散。然而，目前我国的监管体制在跨区域、跨平台协调方面存在明显不足，不同地区的监管部门之间缺乏有效的信息共享和协同执法机制，各网络平台之间也缺乏统一的信息管理标准和协作配合。这使得在治理非法代理维权网络信息时，容易出现监管真空和相互推诿的现象，大大降低了治理效率。

（二）复杂虚假信息辨别困难

一是公众金融知识和法律意识淡薄。许多公众对金融知识

和相关法律法规了解有限，缺乏对非法代理维权网络信息的辨别能力。在面对复杂的金融问题和诱人的维权承诺时，容易被不法分子的虚假宣传所迷惑，盲目相信并寻求非法代理维权服务。一些消费者在信用卡逾期或网贷还款困难时，由于不了解正规的金融纠纷解决途径和相关法律法规，轻信了非法代理维权组织所谓“全额免息”“征信修复”的虚假承诺，不仅未能解决问题，反而遭受了更大的经济损失。

二是虚假信息与真实维权信息交织。在网络空间中，非法代理维权的虚假信息常常与真实的维权诉求和合法的维权渠道信息相互交织，增加了公众辨别真伪的难度。不法分子善于利用公众对维权的关注和需求，将非法信息巧妙地伪装成正常的维权建议或法律咨询，使得普通用户难以分辨。比如非法代理维权组织在宣传中会夹杂一些看似专业的法律条文和金融知识解读，表面上是在为用户提供帮助，实则暗藏非法目的，让用户在不知不觉中陷入陷阱。

三是信息来源复杂且难以追溯。非法代理维权网络信息的来源广泛而复杂，涉及众多个人、组织和网络平台。这些信息在传播过程中经过多次转发、改编，使得其原始来源难以追溯。同时，一些不法分子为逃避责任，往往采用虚拟身份、匿名发布等方式传播信息，进一步增加了信息溯源的难度

（三）溯源治理效果难以持久

一是非法代理维权行为屡禁不止。尽管监管部门不断加大

对非法代理维权网络信息的打击力度，但由于此类行为背后存在巨大的利益驱动，不法分子不惜铤而走险，导致非法代理维权现象屡禁不止。一旦监管稍有松懈，非法信息便会迅速反弹，重新在网络上泛滥。例如，在一些地区，监管部门开展专项整治行动期间，非法代理维权网络信息传播得到一定程度遏制，但行动结束后不久，这些信息又死灰复燃，再次充斥网络空间。

二是治理手段单一且缺乏长效机制。目前，对非法代理维权网络信息的治理主要依赖于监管部门的行政处罚和网络平台的内容审核等手段，治理手段相对单一，且缺乏系统性和长效性。行政处罚往往只能对已经发现的违法行为进行事后惩戒，难以从源头上杜绝非法信息的产生和传播；网络平台的内容审核也存在一定局限性，难以做到对所有信息的实时、精准筛查。此外，治理过程中缺乏对非法代理维权产业链的深入挖掘和全面打击，未能有效切断其利益链条，导致治理效果难以持久。

三是公众参与度不高。治理非法代理维权网络信息需要全社会的共同参与，但目前公众对这一问题的认识和重视程度不足，参与治理的积极性和主动性不高。一方面，许多公众自身缺乏辨别非法信息的能力，无法及时发现和举报此类信息；另一方面，部分公众即使发现了非法代理维权网络信息，也因担心遭受报复或认为与自己无关而选择沉默。公众参与度的低下，使得治理工作缺乏广泛的群众基础，难以形成有效的社会监督氛围，从而影响了治理效果的持久性。应对非法代理维权网络

信息治理挑战的策略

（四）新型攻击技术对抗滞后

一是技术迭代速度与监管技术能力不匹配。非法代理维权群体不断利用 AI 深度伪造、区块链匿名传播、暗网加密传输等前沿技术升级传播手段。例如，利用 AI 生成的“维权导师”虚拟形象在短视频平台发布诱导性内容，其 lip-sync 技术与真人无异，普通用户甚至监管技术系统都难以快速识别真伪，而监管部门现有技术监测模型对这类新型内容的特征抓取和算法识别存在明显滞后。

二是跨平台技术溯源机制缺失。非法信息常通过多平台跳转、动态域名、短链接组合传播，现有监管技术手段难以在短时间内完成跨平台数据关联分析和动态域名解析。非法代理维权组织通过 Telegram 发布含动态短链接的消息，链接每小时自动更换域名后缀，且跳转后的页面内容通过加密协议传输，监管部门的常规爬虫技术和数据监测系统难以实时追踪和拦截。

三是技术对抗的人才与资源短板。非法代理维权团伙往往高薪招募网络安全领域专业人才构建技术壁垒，而监管部门和网络平台在攻防对抗中面临专业技术人才短缺、研发投入不足的问题。部分中小网络平台因缺乏足够的资金和技术团队，无法自主开发针对新型加密传输协议的监测工具，只能依赖通用型内容审核系统，导致对采用自定义加密协议传播的非法信息

漏检率较高。此外，监管部门与高校、科技企业在新型网络安全技术研发上的协同机制尚未健全，难以快速将前沿技术转化为实际监管能力。

七、非法代理维权的溯源治理建议

（一）健全法律法规建设，强化执行宣传力度

一是完善相关法律法规。针对非法代理维权网络信息的特点和治理需求，加快制定和完善相关法律法规，明确非法代理维权行为的界定标准、法律责任以及信息传播各环节主体的义务。细化对非法代理维权信息发布者、传播者的行政处罚和刑事处罚规定，加大对其违法成本的威慑力度；明确网络平台在信息审核、监测、处置等方面的具体责任和义务，建立健全平台责任追究机制。

二是强化法律执行力度。加强司法部门、公安机关、监管部门的执法能力建设，提高执法人员对相关法律法规的理解和运用水平，确保法律的严格执行。建立常态化的执法检查机制，加大对非法代理维权网络信息的监测和查处力度，做到有法必依、执法必严。同时，加强跨部门、跨区域的联合执法协作，形成执法合力，提高执法效率。建立金融监管部门、网信办、公安机关等多部门协同作战的联合执法机制，共同打击非法代理维权网络信息的传播和相关违法犯罪行为。

三是开展法律宣传。通过多种渠道和形式，广泛开展金融

法律法规和非法代理维权风险防范的宣传活动，提高公众的法律意识和风险防范能力。建议利用电视、广播、报纸、网络等媒体平台，开设专题节目和专栏，普及金融知识和法律法规；组织开展线下宣传活动，深入社区、学校、企业等场所，举办讲座、发放宣传资料，面对面为公众讲解非法代理维权的危害和防范方法，引导公众树立正确的维权观念，自觉抵制非法代理维权信息。

（二）提升横纵监管质效，强化网络平台责任

一是加强监管力量建设。加大对监管部门的人力、物力和财力投入，充实监管队伍，提高监管人员的专业素质和业务能力。同时，积极引进和运用先进的技术手段，如大数据分析、人工智能监测等，提升对非法代理维权网络信息的监测、识别和处置能力。建立基于大数据的网络信息监测平台，实时收集、分析各类网络信息，通过智能算法精准识别非法代理维权信息，并及时向监管部门发出预警，实现对非法信息的快速响应和处置。

二是建立跨区域、跨平台协调机制。加强不同地区监管部门之间以及监管部门与网络平台之间的沟通协作，健全跨区域、跨平台的信息共享和协同治理机制。通过定期召开联席会议、建立联合工作小组等方式，加强工作协调和经验交流，共同研究解决治理过程中遇到的问题。同时，推动各网络平台建立统一的信息管理标准和协作配合机制，加强对平台内信息的审核

和管理，形成全网联动的治理格局。建立全国性的非法代理维权网络信息治理信息共享平台，实现各地监管部门和网络平台之间的信息互通、资源共享，提高治理工作的协同性和整体性。

三是强化网络平台责任。明确网络平台在治理非法代理维权网络信息中的主体责任，要求平台建立健全信息审核、管理和举报处理制度，加强对平台内信息的日常监测和审核，及时发现和处置非法代理维权信息。对未履行平台责任、导致非法信息大量传播的平台，依法依规予以严肃处理。同时，鼓励网络平台积极运用技术手段，如设置关键词屏蔽、建立风险提示机制等，从源头上遏制非法代理维权信息的传播。要求网络平台对涉及“征信修复”“债务减免”等敏感关键词的信息进行严格审核，对疑似非法代理维权信息进行标注提醒，引导用户谨慎对待。

（三）提高公众防范意识，构建全民监督机制

一是增强公众风险防范意识。通过案例警示、风险提示等方式，向公众揭示非法代理维权网络信息的危害和常见套路，引导公众树立正确的风险防范意识，提高自我保护能力。收集整理非法代理维权的典型案例，制作成宣传视频、宣传手册等资料，通过媒体平台、社区宣传等渠道广泛传播，让公众深刻认识到非法代理维权可能带来的经济损失、法律风险和个人信息泄露等问题，从而自觉远离非法代理维权陷阱。

二是建立公众举报奖励机制。鼓励公众积极参与非法代理

维权网络信息的治理，建立健全公众举报奖励机制。对举报非法代理维权信息的公众给予一定的物质奖励和精神表彰，同时保护举报人的合法权益。通过公众的广泛参与，形成全社会共同监督的良好氛围，及时发现和打击非法代理维权网络信息的传播。例如，设立专门的举报热线、举报邮箱或在线举报平台，方便公众举报非法代理维权信息；对查证属实的举报，按照举报信息的重要程度和影响范围给予举报人相应的奖励，激发公众参与治理的积极性。

（四）加大技术研发强度，强化配套资源系统

一是强化技术研发与升级。一方面加大资金投入与人才培养监管部门应设立专项基金，用于支持针对非法代理维权新型技术监测与拦截的研发项目。鼓励高校、科研机构与企业合作，在计算机科学、网络安全等相关专业增设“网络信息安全治理”方向，通过定向培养、联合实验室等方式，为监管部门和网络平台输送具备实战能力的技术骨干。一方面构建技术创新联盟。由政府牵头，联合监管部门、网络平台、科研机构和技术企业，定期开展技术研讨会，建立成果共享机制。针对 AI 深度伪造、区块链匿名传播等难题，共同研发基于多模态数据融合的 AI 深度伪造检测系统等解决方案，同时探索区块链技术在监管溯源中的应用，实现对匿名传播信息的有效追踪。

二是完善跨平台技术溯源体系。建立统一数据共享平台，整合各网络平台数据接口，搭建“网络信息统一溯源平台”。

运用大数据、云计算技术对全网数据进行实时采集与分析，平台须具备数据快速检索、关联分析功能，定位非法信息传播源头。通过建立数据加密传输与安全存储机制，保障平台数据的安全性和可靠性。建议制定跨平台协同规范，出台相关政策法规，明确各网络平台在信息溯源中的责任与义务。要求平台建立应急响应机制，一旦发现非法代理维权信息，需及时将相关数据提交至统一溯源平台，并配合监管部门开展溯源工作。对未履行义务的平台施以处罚，同时将平台合规情况纳入信用评价体系，强化对平台的约束力度。

三是加强人才与资源整合。建议建立人才交流机制，由监管部门与网络平台之间定期开展人才交流活动，互派专业技术人员挂职锻炼。通过实际参与对方的工作场景，提升技术人员对不同领域需求的理解和应对能力。同时，建立专家库，邀请行业内知名专家作为顾问，为治理工作提供技术指导和决策支持，定期组织专家开展专题讲座和技术培训，提升整体队伍的专业水平。整合技术资源，搭建资源共享库，汇聚各类监测工具、数据模型、加密破解技术等资源，供监管部门和网络平台按需使用。鼓励企业将自主研发的先进技术开源共享，对作出突出贡献的企业给予税收减免、荣誉表彰等奖励，激发企业参与技术创新和资源共享的积极性，形成全社会协同治理的技术合力。

四是推动技术标准与协同机制。建议制定技术标准规范，

制定针对非法代理维权网络信息监测、溯源、拦截等环节的技术标准，明确内容识别、加密协议破解、动态域名解析等技术的操作规范和性能指标。确保技术手段的有效性和一致性，为监管工作提供科学的技术依据。实现多源情报数据的高效关联与整合。现阶段应优先采取“人机协同”方式，优化单一模型识别的精度局限，构建动态监测体系，利用人工智能算法对海量网络信息进行实时扫描与初步筛查，形成风险预警清单；监管人员则针对高风险数据进行深度研判，结合行业经验与法律准则，明确非法代理维权行为的判定边界。同时，需搭建跨部门、跨领域的多源情报数据共享平台，打通监管机构、互联网企业、通信运营商间的数据壁垒，运用图计算、知识图谱等技术实现数据的结构化关联，构建非法代理维权网络的全景画像。

（五）建立长效治理机制，灵活完善治理策略

一是加强行业自律。推动金融行业、网络信息服务行业等相关行业建立健全行业自律机制，制定行业规范和自律公约，引导行业内企业和从业人员自觉遵守法律法规，加强自我约束和管理。例如，金融行业协会可以组织会员单位开展自律检查和评估活动，对违规从事非法代理维权业务或为非法信息传播提供便利的企业和个人进行行业内通报批评，并采取相应的惩戒措施；网络信息服务行业协会可以制定行业信息管理规范，督促会员单位加强对平台内信息的审核和管理，共同维护网络信息传播秩序。

二是完善信用惩戒体系。将非法代理维权行为纳入社会信用体系建设范畴，建立健全信用惩戒机制。对从事非法代理维权活动的个人和组织，依法依规将其失信行为记录到信用档案，并通过信用信息共享平台向社会公开披露。同时，对失信主体实施联合惩戒，在信贷、市场准入、行政审批、资质认定等方面依法予以限制，使其一处失信、处处受限。例如，将非法代理维权组织和个人列入失信黑名单，限制其申请银行贷款、参与政府采购项目、从事相关行业经营活动等，通过信用惩戒手段加大对非法代理维权行为的威慑力。

三是持续跟踪和评估治理效果。建立对非法代理维权网络信息治理工作的持续跟踪和评估机制，定期对治理措施的实施效果进行监测和评估，及时发现问题并调整完善治理策略。同时，关注非法代理维权网络信息的新变化、新趋势，提前制定应对预案，不断提升治理工作的针对性和有效性。委托专业的第三方机构对非法代理维权网络信息的治理效果进行评估，通过数据分析、实地调研、公众满意度调查等方式，全面了解治理工作的成效和存在的不足，为进一步改进治理工作提供科学依据。

马上消费研究院

