

# 生成式人工智能型网络谣言的法律综合治理

柳若仪

苏州大学王健法学院，江苏 苏州

收稿日期：2025年2月17日；录用日期：2026年1月13日；发布日期：2026年2月10日

---

## 摘要

人工智能技术的发展正在重塑网络谣言的产生与传播。与传统网络谣言相比，AI网络谣言具有生成的高效性、内容的迷惑性、传播的精准性等特征。现有治理体系面对AI谣言存在平台责任缺位、刑事手段治理无力、行政监管不力等多重困境，无法有效回应对于AI网络谣言的治理需求。应当推动网络平台阶段化治理，落实主体责任；完善刑事程序中的自诉措施，明确主观要素的认定；完善行政处罚要件，规定行政处罚梯级制度，健全行刑衔接机制；同时积极探索网络谣言民事与行政公益诉讼治理，从而构建行之有效的AI型网络谣言的综合法律治理体系。

---

## 关键词

网络谣言，虚假信息，生成式人工智能，行刑衔接

---

# Legal Comprehensive Governance of Generative Artificial Intelligence-Based Online Rumors

Ruoyi Liu

Kenneth Wang School of Law, Soochow University, Suzhou Jiangsu

Received: February 17, 2025; accepted: January 13, 2026; published: February 10, 2026

---

## Abstract

The development of artificial intelligence (AI) technology is reshaping the production and dissemination of online rumors. Compared to traditional online rumors, AI-generated rumors are characterized by high efficiency in creation, deceptive content, and targeted dissemination. The existing governance

framework faces multiple challenges in addressing AI rumors, including the lack of platform accountability, the ineffectiveness of criminal sanctions, and insufficient administrative supervision, which fail to adequately meet the governance demands for AI-generated online rumors. It is essential to promote phased governance by online platforms to enforce primary responsibility, refine self-initiated prosecution measures in criminal procedures to clarify the determination of subjective elements, improve the prerequisites for administrative penalties, establish a tiered administrative penalty system, and enhance the coordination between administrative and criminal enforcement. Concurrently, active exploration of civil and administrative public interest litigation for managing online rumors should be undertaken to construct a comprehensive and effective legal governance framework for AI-generated online rumors.

## Keywords

Online Rumors, Misinformation, Generative Artificial Intelligence, Administrative-Criminal Enforcement Coordination

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

伴随着流量经济的繁荣，网络等谣言不断涌现，消费着社会公众信任，破坏社会公共秩序。更令人担忧的是，利用 AI 生成网络谣言的现象日渐增多，人工智能的快速发展正在重塑网络谣言的产生与传播。根据新闻可信度评估机构 News Guard 的测试，ChatGPT 为 100 条虚假新闻中的 80 条制作了更具细节、知识、说服力的延伸报道。2024 年 4 月、5 月，公安部公布打击整治网络谣言违法犯罪典型案例共 20 起，其中共有 6 起案例涉及利用 AI 工具生成网络谣言。

有关部门已经开始针对 AI 网络谣言展开探索治理。2023 年 9 月，最高人民法院、最高人民检察院、公安部联合发布的《关于依法惩治网络暴力违法犯罪的指导意见》指出，“对实施网络暴力违法犯罪者利用‘深度合成’等生成式人工智能技术发布违法信息的，将依法从重处罚”。针对平台，中央网信办发布的《关于开展“清朗·整治自媒体无底线博流量专项行动”的通知》，提出平台应要求自媒体使用 AI 等技术生成信息的，必须明确标注系技术生成。然而，现行规范体系并未提出有效防控治理 AI 型网络谣言的传播的针对性的解决方案，AI 网络谣言仍面临多重治理困境。

## 2. AI 网络谣言的规制现状

### 2.1. 刑事法规范

针对网络谣言犯罪，我国《刑法》已初步建构起相应的罪名体系。从刑事立法来看，直接将谣言作为构成要件要素的罪名一共有 15 个，涉及刑法分则的 6 个章节。其中对网络谣言的治理最为全面的当属编造、故意传播虚假信息罪和编造、故意传播虚假恐怖信息罪两个罪名。此外，如果网络平台不履行有关的网络谣言治理义务，可能会构成“拒不履行信息网络安全管理义务罪”。

除刑法条文规定外，司法解释也针对网络谣言的传播形成了治理方案。2013 年《关于办理利用信息网络实施诽谤等刑事案件适用法律若干问题的解释》(下文简称《诽谤解释》)针对网络谣言侵害对象的差异以及行为模式的差别，分别以“诽谤罪”“寻衅滋事罪”“敲诈勒索罪”“非法经营罪”等罪名治理网络谣言及其附属行为。

## 2.2. 行政法规范

刑事入罪的高门槛决定了在实践中行政法是规制网络谣言最为普遍的手段。针对个人，网络谣言治理最主要的行政法依据是《治安管理处罚法》第 25 条第 1 项，“有下列行为之一的，处五日以上十日以下拘留，可以并处五百元以下罚款；情节较轻的，处五日以下拘留或者五百元以下罚款：（一）散布谣言，谎报险情、疫情、警情或者以其他方法故意扰乱公共秩序的；”该条确定了网络谣言行政违法的构成要件：存在网络散布谣言的客观行为、扰乱公共秩序的危害后果和主观故意。针对网络平台的责任，《网络安全法》第六十八条作出了相关规定。

民法规范从私权保障的角度出发，规定了制造传播谣言者的责任和网络平台的有关义务。在网络谣言传播者的责任方面，《民法典》1028 条、1037 条规定了在出现网络谣言等情形时，受害人有权采取更正、删除等措施，从而可以遏制网络谣言的大规模扩散。在网络《民法典》第 1194 条规定了网络用户、网络服务提供者利用网络侵害民事权益的，应当承担民事责任，从而确立了调整网络侵权的基本规则。

在网络谣言的平台责任方面，《民法典》规定了网络服务提供者的相关义务。《民法典》第 1195 条规定了“通知 - 删除”规则，如果有受害人发现有关网络谣言正在平台上传播并通知了网络服务提供者，服务提供者接到通知后，应当及时将该通知转送相关网络用户，根据构成侵权的初步证据和服务类型采取必要措施。《民法典》1197 条规定了网络平台的连带责任，即网络平台知道或者应当知道网络用户利用平台编造、传播网络谣言侵害他人民事权益却并未采取必要措施的，应当承担连带责任。

## 2.3. 民事法规范

民法规范从私权保障的角度出发，规定了制造传播谣言者的责任和网络平台的有关义务。在网络谣言传播者的责任方面，《民法典》1028 条、1037 条规定了在出现网络谣言等情形时，受害人有权采取更正、删除等措施，从而可以遏制网络谣言的大规模扩散。在网络《民法典》第 1194 条规定了网络用户、网络服务提供者利用网络侵害民事权益的，应当承担民事责任，从而确立了调整网络侵权的基本规则。

在网络谣言的平台责任方面，《民法典》规定了网络服务提供者的相关义务。《民法典》第 1195 条规定了“通知 - 删除”规则，如果有受害人发现有关网络谣言正在平台上传播并通知了网络服务提供者，服务提供者接到通知后，应当及时将该通知转送相关网络用户，根据构成侵权的初步证据和服务类型采取必要措施。《民法典》1197 条规定了网络平台的连带责任，即网络平台知道或者应当知道网络用户利用平台编造、传播网络谣言侵害他人民事权益却并未采取必要措施的，应当承担连带责任。

## 3. AI 网络谣言治理的现实困境

### 3.1. 平台责任的缺位

网络谣言治理中的平台可分为三类平台：基础模型提供者、服务提供者、信息传播平台。作为守门人的数字内容平台在很大程度上决定了信息内容的生产与传播，塑造了用户的网络信息空间和互动方式。平台为网络谣言提供了生成平台和传播媒介，其既是风险制造的参与方，又是相应领域所涉各主体中主要的获益方，预防能力相比于其他主体也处于优势地位，故而让其承担一定的风险防范义务有其合理性[1]。

在网络谣言治理中，第一，包括基础模型提供者、服务提供者在内的人工智能服务提供平台的责任不明确。根据《生成式人工智能服务管理暂行办法》第 12 条规定，生成式人工智能服务提供者应当按照《互联网信息服务深度合成管理规定》对图片、视频等生成内容进行标识，但对利用 AI 工具生成的文字没有相关要求，且对平台未落实相关标准的责任和处罚都未予明确。第二，信息传播平台的责任缺失。

《关于加强“自媒体”管理的通知》中只规定了信息内容传播网站平台对含有虚构情节、剧情演绎的内

容进行显著标注，及时标记谣言标签、呈现辟谣信息的义务，但没有规定网站平台不履行相关义务的责任后果。传播平台内容推荐所依赖的算法机制蕴含着平台对舆论导向的隐形干预，其价值理念也必然会被通过呈现的内容进入用户视野——此时传播平台倘若继续恪守平台传统的中立角色，无需承担相应责任，恐怕难以适应智能化网络谣言井喷式增长的实际。

### 3.2. 刑罚治理的无力

现有刑法规范面对传统网络谣言治理已面临规制困境，在生成式人工智能显著扩大谣言传播风险、改变谣言生成方式的背景下，刑法面对 AI 网络谣言治理的无力进一步凸显。

#### 3.2.1. 罪名界定模糊

规制网络谣言在刑法实践中主要依赖两个核心罪名：编造、故意传播虚假信息罪和寻衅滋事罪。二者在行为特征和法律适用上存在部分重叠，尤其随着人工智能技术被用于生成和传播虚假信息，这种重叠导致的定罪难题愈发突出。具体而言，如果行为人利用 AI 工具编造、传播涉及疫情、险情、灾情、警情等特定社会公共事件的虚假信息，其行为可能同时符合两罪的构成要件。一方面，编造、故意传播虚假信息罪明确规制虚假的险情、疫情、灾情、警情等四类信息，利用 AI 生成的信息如属上述范围，自然落入该罪名范畴。另一方面，若该信息在网络空间传播，严重扰乱社会秩序，符合“在信息网络上编造、散布虚假信息，造成公共秩序严重混乱”的情形，则也可能构成寻衅滋事罪。这种重叠在行为对象与危害后果上形成竞合，导致司法实践中出现罪名认定上的分歧。目前，两罪在刑罚配置上存在明显差异。寻衅滋事罪的法定刑更高，最高可处五年以上十年以下有期徒刑，而编造、故意传播虚假信息罪的刑罚相对较轻。这种差异使得在行为同时触犯两罪时，如何选择适用罪名成为关键难题。若一律以重罪(寻衅滋事罪)论处，可能不当扩大重罪的适用范围，违背罪刑相适应原则；若仅以轻罪(编造、故意传播虚假信息罪)处理，又可能在行为造成特别严重后果时难以实现刑罚的惩戒与威慑功能。

#### 3.2.2. 诉讼程序及其配套制度缺失

利用 AI 工具编造、传播谣言涉嫌诽谤罪时，自诉、公诉的程序选择与程序衔接存在问题。

首先，自诉案件的立案标准过高。根据《最高人民法院关于适用<中华人民共和国刑事诉讼法>的解释》第 316 条、第 320 条规定，法院立案受理自诉案件的条件之一是“有证明被告人犯罪事实的证据。”因此法院在对自诉案件进行立案审查时，只有具有足够证据时才会受理，若自诉人提供的证据没有达到立案证明标准，法院通常会要求自诉人补充证据，若自诉人不予以补充或者补充后仍然没有达到立案证明标准否则在立案审查阶段会对“缺乏罪证”的案件作出不予受理裁定[2]。2023 年《关于依法惩治网络暴力违法犯罪的指导意见》(下文简称《网暴意见》)也并未纠正自诉立案门槛高的问题。立案阶段审查的实质化和对自诉案件证据标准设定较高的质量标准，导致自诉程序启动困难。在自诉案件中，自诉人需要提供有犯罪事实的证据。涉 AI 网络谣言案件中的绝大部分证据属于电子证据，电子证据在诉讼中作为证据，要具有真实性、完整性、关联性、合法性，要求取证主体知悉相应取证程序、取证技术和取证规则，自诉人提出的证据无论是从数量还是种类来说都需要达到一定程度，才有可能形成完整的证据链条并达到印证模式下的刑事证明标准[3]。AI 型谣言生成的黑产化趋势，犯罪团体内部高度职业化、有组织化、跨地域性、分工明确的特征，也加剧了被害人通过自诉方式寻求救济的难度[4]。

其次，自诉和公诉的衔接程序问题。《网暴意见》第 13 条明确了公诉、自诉的程序衔接规则，认为只要存在严重扰乱社会秩序的情形，就必须由自诉转公诉，表达了鼓励公诉程序的价值倾向。然而这一规定面临不尊重当事人的意愿的正当性质疑，此时只要国家机关认为属于严重危害社会秩序和国家利益的情形，就必须转为公诉程序，可能剥夺被害人的程序选择权。如何准确把握进入公诉程序标准也是实践中的难题。《网暴意见》第 12 条规定，“实施侮辱、诽谤犯罪，严重危害社会秩序和国家利益时，通

过公诉程序处理。”但是“严重危害社会秩序”的内涵模糊不清，网络空间秩序是否属于本条的社会秩序存在极大争议。实践中，已有存在利用 AI 造谣严重扰乱网络公共秩序因而涉嫌寻衅滋事罪的案例。

《网暴意见》指出应当综合侵害对象、动机目的、行为方式、信息传播范围、危害后果等因素作出判定，并列举了五种严重危害社会秩序的情形，但仍未明确网络空间秩序是否社会秩序。第三，协助对象比较笼统。公安协助取证的对象究竟是法院还是自诉人不明，而这涉及到了取证内容和后续审查证据的程序等问题。第四，协助取证的阶段不明确。此条规定确立了公安机关协助取证的义务适用阶段是在立案前，在立案后协助取证是否还有制度空间尚未明确[5]。

### 3.2.3. 主观要素证明困难

证明困境主要体现在对行为人的主观要件证明中。网络谣言型寻衅滋事罪和编造、故意传播虚假信息罪都要求行为人的主观心理状态系“故意”且“明知是虚假信息”。对于利用 AI 工具主动编造谣言的人，毫无疑问是“故意”且“明知”；问题在于，传播 AI 谣言者的“故意”和“明知”如何认定。公安机关对诸如“故意”“明知”等主观内容的证明，大多借助间接证据进行。对于主观故意要素的证明，公诉方需要承担证明被告人主观罪过的证明责任，且证据确实充分的程度；对于“明知”要素，不同于在毒品型犯罪中明确规定了适用推定规则证明明知要素，相关解释并未针对网络谣言犯罪中“明知”的证明困难作出特别规定，因此只能通过间接证据进行推理。AI 型谣言的内容形式相较于普通谣言的迷惑性更高，此时对“明知”要素的证明更具难度，是否能适用推定也存在疑问。

## 3.3. 行政监管不力

### 3.3.1. 行政处罚体系不健全

实务中对直接造谣的责任人主要的处罚规范依据是《治安管理处罚法》。该法第 25 条规定，针对制造、故意传播谣言者，即使情节严重，也仅能给予 500 元以下的罚款、10 天以下的行政拘留。而 AI 型谣言非法收益相较于传统的谣言显著提高，AI 谣言的惩罚力度之小和获利之间存在极大的不平衡，违法成本低、惩处力度不足，导致行政处罚手段失去威慑力。

### 3.3.2. 行刑衔接不畅

首先，反向衔接机制虚化。AI 网络谣言往往涉及网络黑产业链条，AI 网络谣言的生成软件服务提供者、生成者、传播者都需要承担责任，但并不是所有环节都值得动用刑法手段，同一案件中也不是所有的犯罪嫌疑人都值得动用刑罚手段，在需要进行责任分流时，行刑反向衔接的价值得以体现[6]。在 AI 产业链中，涉及到上下游，如提供个人信息注册账号、AI 平台等。然而实践中反向衔接机制虚化，且存在诸多尚不明确的问题，例如同步移送案件证据材料问题，以及是否需要扣除刑事办案期限、行政处罚时效是否需要中止等问题。

其次，行刑证据转换不明确。《中华人民共和国行政处罚法》第 27 条第 2 款明确了行政执法机关和司法机关之间应当加强证据移交、接收的衔接，但由于行政执法和刑事司法的取证程序和证据标准的规范性差异，导致行刑衔接过程中证据转化和衔接机制不健全。这种情形在网络谣言治理中尤为突出。一方面，行政证据转化为刑事证据的规则模糊性带来司法适用困境。《中华人民共和国刑事诉讼法》(以下简称《刑事诉讼法》)第 54 条虽规定“行政机关在行政执法和查办案件过程中收集的物证、书证、视听资料、电子数据等证据材料，在刑事诉讼中可以作为证据使用”，但由于行政权与司法权力属性之不同，在刑事司法中直接使用行政证据认定案件仍具有理论争议。另一方面，案件中电子数据规范，认定问题亟待健全。AI 网络谣言案件中均是以电子数据方式存储、传输，行政机关移送的电子数据证据是刑事司法机关认定案件的关键证据，但行刑衔接中缺乏立法指引来规范被移送电子数据的认定程序，直接影响

相关罪名的判罚。此外，利用 AI 编造、传播谣言涉嫌诽谤罪时，自诉程序中存在如何将行政证据转为可供自诉人使用的刑事诉讼证据的问题，也存在多种争议[7]。

## 4. 网络谣言治理的路径优化

### 4.1. 落实平台责任

#### 4.1.1. 实行阶段化治理

针对基础模型提供平台、应用部署平台和网络信息应用传播平台，应当分阶段依据不同平台的功能定位落实差异化的平台责任。

首先，事前预防阶段。第一，针对基础模型提供者，深度合成服务的提供者应按照《深度合成管理规定》的规定履行备案和注销、变更手续，以便行政监管部门及时对违规平台采取不予上架、警示、暂停服务、下架等限制措施；从数据治理、算法设计等方面完善虚假信息治理工具开发创新，推动技术审核。从技术角度，人工智能提供者需采取合理的预防措施，将自我监管介入点前置，在模型训练时便植入伦理规范以确保通用模型输出内容更符合人类的价值取向。通用人工智能提供者内容生产全过程采取技术措施审查、过滤有害信息，应针对有害信息，结合内容生产场景、提供者风险预见能力差异化构建内容审查的注意义务[8]。该注意义务以技术性审查为标准，提供者采取符合技术特点的内容审查机制，便应当认定履行了相关义务。第二，针对应用部署平台，应当在交互环节中引入拒答机制。鼓励模型在缺乏充分信息支持时主动表达不确定或无法回答，通过容忍沉默而非胡编，真实性导向落实到模型行为层面，因迎合用户预期而制造错误信息。第三，针对作为内容分发平台的网络信息应用服务提供者和传播者，如果此类平台的注意义务仅停留在被动的“通知 - 删除”层面，则无法满足治理网络谣言的现实需求。平台应当针对深度伪造技术，更新构建融合深度伪造信息和关联计算的动态理论模型；针对 AI 谣言所偏好的风险议题，从叙事策略等角度构建相应的信息识别机制，有效识别各类频繁诱导网民群体传播扩散谣言的话题营销、账号发言、热搜组织、直播运营等内容业务，精准排查歪楼引战、制造对立、激化矛盾的重点账号，针对内容创作门槛较低的账户发布的信息进行特殊关注，使社交媒体平台的事前预警机制能深度干预网络谣言生产传播的环境因素与内容生态，以便迅速采取对应的处置措施。完善网络谣言人工审核制度，承担合理提示和标识义务，针对可能属于虚假信息的应当显著标识。

在事中阻断环节，除了传统的删除、屏蔽等手段，网络信息应用服务提供者和传播平台应当采取技术手段放大理性言论、建设健康生态环境，扭转相关虚假信息的优势传播地位，改善劣质的社区生态。例如，可以要求发布谣言的当事人提供相应凭证，否则设置“内容存疑”标签并限制传播；应当精准定位关键舆情节点，利用自身的信息发布与传播能力及时在相关信息中嵌入关键信息，使得理性信息、权威信息拥有优势转播地位。对于短时间内上升的社会热点事件，重点审核该事件的网络评论信息，尤其是对平台自媒体、网络大 V 所发布的信息内容进行合法性审核，预防以碎片化信息替代完整信息，转发权限限制、冻结或注销账号等措施避免虚假信息在网络平台大量传播。

在事后惩治、复盘阶段，应当加强信息传播平台的算法设计。通过对生产虚假信息的主体账号的识别，比如通过对已查明的虚假信息的传播网络进行数据建库，以这些数据为基础，从中计算寻找虚假信息的传播源，并重点关注重复参与多次传播网络的账号[9]。

#### 4.1.2. 明确平台责任后果

从基础模型研发者的责任界定来看，基础模型研发者的责任边界应与其实际技术控制能力相匹配。虽然在研发阶段可通过程序预设人工智能的行为边界，但随着模型进入下游部署与使用环节，研发者对实际运行过程的控制力将显著减弱。基于鼓励技术创新的政策导向，不宜对其赋予过重的责任，而应限

定在针对可预见的生成偏差承担基础性风险防控义务的合理范围内。

其次，就内容提供者而言，其应对内容的分发负有相应的注意义务。如果提供者明知存在虚假信息，或经权利人明确告知已涉及侵权，却仍未采取合理措施，最终导致侵权结果扩大的，应当对扩大的损害部分承担相应责任。具体而言，在民事法律责任层面，社交媒体平台未履行网络信息治理义务、给他人造成损害的，应对其不作为造成的损害扩大部分承担间接侵权责任。其次，平台的行政法律责任。应当明确网络平台不履行相应义务的责任后果，对行政处罚适用的条件进一步明确，以便于网络服务提供者识别相应风险，否则网络平台将面临不明确的法律责任风险。最后，平台的刑事法律责任。应当将拒不履行网络安全管理义务罪视为威慑性手段，无需积极追求该罪的适用。拒不履行信息网络安全管理义务罪在立法技术上以逐层递进的方式，通过“行政不法认定的前置性 - 行政程序的前置性 - 网络服务提供者的拒不作为 - 造成严重后果”的层级认定，在刑法和行政法领域内实现了对网络服务提供者不法行为的分层规制和责任分配[10]。其设置目的在于督促其积极履行相关义务，通过罪名内嵌的“避风港原则”，责令平台整改，设置出罪程序，从而对平台施以合规压力，督促网络平台积极履行相应义务。

## 4.2. 完善刑法治理

### 4.2.1. 准确把握入罪标准

《网暴意见》第 12 条规定，利用“深度合成”等生成式人工智能技术发布违法信息的，依法从重处罚，从而确定了利用 AI 发布违法信息属于加重处罚情形。但从重处罚的前提是制造、传播谣言需要判断行为是否入罪。刑法是最后的、补充性的工具，因此必须谨慎入罪，精准把握违法犯罪行为的认定标准 [11]。具体而言，应当把握“严重扰乱社会秩序”的具体内涵。网络上的不实言论如果涉及重大灾害或重大犯罪，则不实言论的发布通常会立即对财产、民众的健康安全造成直接损害，给社会造成重大恐慌，或者改变行政执法方向，那么这就是一种严重破坏社会秩序的行为。如果只是造成网络秩序的混乱，那么此时不宜入罪。此外，在定罪时应当谨慎适用寻衅滋事罪。《网络诽谤解释》第 5 条规定对网络型寻衅滋事罪有“造成公共秩序严重混乱”的要求。公共秩序属于社会秩序的子系统，从文义上是比非网络型寻衅滋事罪更高的要求，故网络型寻衅滋事罪的入罪解释应当更为严格。当相关谣言能够被解释为编造、故意传播虚假信息罪中的四类虚假信息时，在罪名适用上应当优先适用构成要件明确的编造、故意传播虚假信息罪，才符合罪刑法定中的明确性原则。

### 4.2.2. 完善涉自诉案件程序

首先，调整自诉程序立案标准。针对自诉案件的立案受理阶段，应当回归《刑事诉讼法》的立法原意，取消自诉案件立案受理的证明标准，仅要求当事人就立案时的程序性事实承担证明责任；同时强化开庭前的庭前审查义务，在开庭前过滤不符合法定条件的自诉案件，此时对是否有足够的证据予以审查，既可以过滤不符合自诉条件的相关案件，也有利于保障后期开庭审理的高效、顺利运行。

其次，细化公安协助取证义务。在涉 AI 谣言自诉案件中，激活公安机关协助调查、法院调取证据等相关制度。降低“提供证据确有困难”的标准；细化协助取证制度的启动标准。明确协助取证的性质，协助取证的时间阶段在刑事立案之前，因此不应当适用刑事侦查取证的相关规定。

最后，完善自诉和公诉的程序衔接。理顺告诉才处理制度与自诉制度的逻辑关系；在自诉转公诉程序中，充分尊重被害人的程序选择权。对于相关案件“严重危害社会秩序”的公诉条件，必须谦抑解释；在自诉向公诉程序转换的过程中，以说服当事人撤回自诉为原则，裁定不予受理、驳回起诉为例外。应明确网暴案件自诉转公诉过程中撤回自诉、上诉、和解等法律效果，赋予被害人相应的程序救济权；但是在自诉转公诉后，为避免程序冲突，不应当赋予当事人上诉权。

### 4.2.3. 明确主观要素认定

基于网络空间的特殊性和网络言论的特殊性，认定“故意传播”网络谣言的行为往往十分困难，故意与过失难以分辨。此时可以参照“编造并传播证券、期货虚假交易信息罪”的立法模式，对于传播AI谣言者进行刑法治理主观的证明程度必须达到“确实知道”的程度。一般体现为两种情形：行为人编造并自行转播谣言，或传谣者与造谣者事前同谋，共同传播谣言。例如，若用户输入的提示词具有明显诱导虚构倾向(如“请编造某事件细节”)、多次调整参数以规避真实性约束，或事后将虚假信息用于牟利、诽谤等恶意目的，则可推定其存在直接或间接故意；反之，若用户基于常规问题提问且生成内容具有隐蔽性错误(如虚构专业文献)，并在知悉后及时采取更正、删除等措施，则更可能认定为过失，不应当轻易认定行为人的主观心理为“明知”。司法实践需依托用户与AI的完整交互日志，结合其核查成本、传播场景与事后行为，构建“技术事实与主观意图相印证”的主观要素认定框架。

## 4.3. 优化行政规制

首先，对于行政法规范中的公共秩序应当进行限缩解释。在网络谣言治理中，对于公共秩序的认定不得随意扩张，应当保持与刑事法治理中公共秩序概念的一致性，公共秩序指的应当是真切事关“公众生活的平稳与安宁”的有序社会状态，不能将其泛化为抽象的“网络公共秩序”，也不能将其仅限于“真实的社会空间秩序”。针对行政处罚体系性失衡的问题，应当坚持行政处罚中的处罚必要性原则。不应当随意使用《治安管理处罚法》第26条的寻衅滋事对其传播网络谣言的行为进行处罚，寻衅滋事治安处罚应明确定位于辅助性地位。应当在强化对事实的认定和对证据的采纳，注重收集相对人违法行为情节、主观过错、悔过程度等各方面的证据，在认真、充足地调查基础上形成事实认定，再进行治安处罚。

其次，调整处罚力度和处罚手段，根据造谣者的情节、谣言的传播范围等设置阶梯式的量罚；同时丰富行政处罚的种类，可以增加对特定机构责令停产停业、责令关闭等处罚，从而打击相关产业链。在调整量罚时应当注意，应体现宽严相济，避免过度惩罚，要既能有效地维护社会秩序，又不过度侵犯个人权利；并针对加大处罚力度的行为做好具体理由的阐释，在处罚力度与法律公正中寻求一个好的平衡点进行立法会使其更具备法律价值。出于行政处罚惩戒与预防并重的双重功能，在处罚适用中应当考虑行为人首次违法的情节，结合相对人的悔过程度，给予其一次改过自新的机会，通过强化对违法行为人的教育和指引，促使其自我纠错以避免造成危害后果。

最后，通过立法落实网络谣言治理中先行政管制、后刑事制裁的理念。可以模仿拒不履行网络安全管理义务罪的立法形式，在涉网络谣言相关罪名中增加行政不法的前置要件，以编造、故意传播虚假信息罪为例，可以在构成要件中加入“已被有关部门提醒拒不删除信息、不消除风险的”，从而实现通过保留刑事法治理的可能激活行政法治理的充分实现。通过衔接行政前置程序，实现针对网络谣言的传播进行快速反应，进行实时的甚至事前的调整，发挥针对网络谣言的预防作用。

## 参考文献

- [1] 劳东燕. 网络暴力刑法治理的基本立场[J]. 政法论坛, 2024, 42(3): 39-54.
- [2] 谢登科. 网络暴力犯罪自诉程序的立案证明标准反思[J]. 当代法学, 2024, 38(3): 84-96.
- [3] 张谦. 网络暴力自诉案件中的公安协助取证制度研究[J]. 法治研究, 2025(2): 134-145.
- [4] 谢登科. 网络暴力犯罪的公私协同治理模式[J]. 法律科学(西北政法大学学报), 2023, 41(5): 92-101.
- [5] 张培铮. 网络暴力自诉案件的公安协助取证制度形成逻辑与理性反思[J]. 四川警察学院学报, 2024, 36(1): 51-67.
- [6] 李怀胜. 网络犯罪治理的刑行衔接: 基本价值与运作模式[J]. 南京师大学报(社会科学版), 2023(3): 99-111.
- [7] 蔡元培. 网络暴力自诉案件中的证据供给与风险预防[J]. 中国应用法学, 2023(5): 86-97.

- 
- [8] 高阳. 通用人工智能提供者内容审查注意义务的证成[J]. 东方法学, 2024(1): 189-200.
  - [9] 朱笑延. 从内容监管到生态调控: 网络暴力信息治理的平台义务重塑[J]. 南京大学学报(哲学·人文科学·社会科学), 2024, 61(1): 76-91+163-164.
  - [10] 杜小丽. 社会治理视角下拒不履行信息网络安全管理义务罪再审视[J]. 中国法律评论, 2024(3): 86-95.
  - [11] 姜涛. 网络谣言的刑法治理: 从宪法的视角[J]. 中国法学, 2021(3): 208-228.