巴中市哲学社会科学规划项目

结项申请书

立 项 编 号 BZ25YB120

项 目 类 别 一 般 项 目

项 目 名 称 生成式人工智能技术嵌入巴中市网络舆情决策的价值、风险及防控

项目负责人 李 欣 晨

所 在 单 位 巴 中 市 委 党 校

填 表 日 期 2025.10.16

巴中市社会科学界联合会 制

2025 年 10 月

声　　明

本研究成果不存在知识产权争议；巴中市社会科学界联合会享有推广应用本成果的权利，但保留作者的署名权。特此声明。

成果是否涉及敏感问题或其他不宜公开出版的内容：是☐ 否☑

成果是否涉密： 是☑ 否☐

项目负责人（签字）

年　　月　　日

填 表 说 明

一、本表适用于巴中市社科年度规划项目、专项项目等结项申请。

二、认真如实填写表内栏目，凡选择性栏目请在选项上打“√”。课题申报信息无变更情况的可不填写《项目变更情况数据表》。

三、本《结项申请书》报送 2 份（A3 纸双面印制，中缝装订），并附最终成果打印稿（正文格式要求：主标题 2 号方正小标宋简体，其中一级标题 3 号方正黑体-GBK，二级标题 3 号方正楷体-GBK，三级标题 3 号方正仿宋-GBK 加粗，正文 3 号方正仿宋-GBK）。

四、所有结项材料须经所在单位审核并签署意见。县（区）申报者报送所在县（区）社科联审核后统一报送至市社科联，其他申报者可直接报送市社科联。

一、项目变更情况数据表

<table>
<tr><td colspan="2">立项项目名称</td><td colspan="7"></td></tr>
<tr><td colspan="2">结项成果名称</td><td colspan="7"></td></tr>
<tr><td colspan="2">是否变更</td><td colspan="2">A、是　　B、否</td><td colspan="2">变更的内容</td><td colspan="3"></td></tr>
<tr><td colspan="2">原计划成果形式</td><td colspan="2"></td><td colspan="2">现成果形式</td><td colspan="3"></td></tr>
<tr><td colspan="2">原计划完成时间</td><td colspan="2">年 月 日</td><td colspan="2">实际完成时间</td><td colspan="3">年 月 日</td></tr>
<tr><td colspan="9">项目负责人及参与人员变更情况</td></tr>
<tr><td rowspan="3">原负责人</td><td>姓　　名</td><td></td><td>性别</td><td></td><td>民族</td><td></td><td>出生日期</td><td>年　月</td></tr>
<tr><td>所在单位</td><td colspan="2"></td><td>行政职务</td><td></td><td>专业职务</td><td colspan="2"></td></tr>
<tr><td>通讯地址</td><td colspan="3"></td><td>联系电话</td><td colspan="3"></td></tr>
<tr><td rowspan="3">现负责人</td><td>姓　　名</td><td></td><td>性别</td><td></td><td>民族</td><td></td><td>出生日期</td><td>年　月</td></tr>
<tr><td>所在单位</td><td colspan="2"></td><td>行政职务</td><td></td><td>专业职务</td><td colspan="2"></td></tr>
<tr><td>通讯地址</td><td colspan="3"></td><td>联系电话</td><td colspan="3"></td></tr>
<tr><td rowspan="5">原参与人员</td><td>姓　　名</td><td colspan="3">单　　位</td><td>职　称</td><td colspan="3">联系电话</td></tr>
<tr><td></td><td colspan="3"></td><td></td><td colspan="3"></td></tr>
<tr><td></td><td colspan="3"></td><td></td><td colspan="3"></td></tr>
<tr><td></td><td colspan="3"></td><td></td><td colspan="3"></td></tr>
<tr><td></td><td colspan="3"></td><td></td><td colspan="3"></td></tr>
</table>

现参与人员	姓　　名	单　　位	职　称	联系电话

二、申请人所在单位审核意见

（审核事项:1.成果有无政治导向问题或其他不宜公开出版的内容；2.最终结果的内容质量是否符合预期研究目标。）

该成果无政治导向问题或其他不宜公开出版的内容，最终结果的内容质量符合预期研究目标。

签　章

年　月　日

三、县（区）社科联意见

（审核事项:1.成果有无意识形态问题；2.是否同意结项。）

单位（公章）:

负责人签字：

年　月　日

四、专家鉴定意见

（请在对应意见栏划“√”）

1.成果有无意识形态方面问题：　有 □　　否 □

2.是否同意结项：是 □　　否 □

3.鉴定等级：优秀 □　　良好 □　　合格 □

主审专家签字：

年　月　日

五、市社科联审核意见

单位（公章）: 年　月　日

生成式人工智能技术嵌入巴中市网络舆情决策的价值、风险及防控

摘要：生成式人工智能作为引领新一轮科技革命和产业变革的战略性技术，正深刻重构着社会治理格局。生成式人工智能技术嵌入网络舆情决策是大势所趋，二者的有机结合有助于提升网络舆情决策能力和水平。当前，巴中市数字化转型进程面临网络舆情决策风险的重重挑战，传统治理模式在风险识别精度、研判响应速度、系统适应弹性等维度呈现“监测盲区—认知断层—响应迟滞”的系统性短板，需要从技术把关、制度保障

和数字素养等方面构建风险防范机制，不断提升生成式人工智能技术嵌入网络舆情决策的安全性、可靠性、可控性和公平性。

关键词：生成式人工智能；网络舆情；数字化治理

在全球数字化浪潮与智能技术迭代的双重驱动下，生成式人工智能作为人工智能领域的革命性突破，正以其强大的自然语言处理、多模态内容生成与海量数据解析能力，重塑社会治理的各个维度。网络舆情作为社会心态的“晴雨表”与政策执行的“反馈镜”，其治理效能直接关系到基层稳定与发展大局。巴中市正处于“一市四区三地”现代化建设的关键阶段，网络舆情治理工作已被明确为维护意识形态安全、走好网上群众路线、服务经济社会高质量发展的迫切需要。在此背景下，引入生成式人工智能技术优化舆情决策流程，成为巴中市提升网络综合治理能力的潜在选择。

一、生成式人工智能技术在网络舆情决策中的核心价值

（一）有助于确认网络舆情决策问题

网络舆情政策问题的精准界定是网络舆情决策的逻辑起点。政策制定主体若能及时识别、明确界定舆情议题，并制定靶向方案与差异化措施，可有效构建舆情风险防控与危机化解路径。但在“互联网 +”数字化传播生态中，有社会影响力的“议题性事件”或“公共性话题”，会在网络场域触发网民深度讨论与观点博弈。不仅不同网民对同一对象的利益诉求存在异质性，且随舆情演进、热度发酵，同一网民的认知与诉求也可能动态

变化。网民诉求的潜在性、多元性、动态性与复杂性，成为政府舆情决策的核心制约因素。而如何依托多源舆情数据，实现网民合理诉求的系统识别、关联整合与有效筛选，并据此出台兼具回应性与操作性的政策，是政府舆情治理的重点与难点。

生成式人工智能时代的到来及其技术在网络舆情决策场景中的深度应用，为政府在复杂交织、动态演变的网络舆情环境中，精准高效识别并系统整合网民诉求提供了关键支撑。**一方面**，生成式人工智能依托强大的记忆存储与逻辑推理能力，为政府舆情管理部门构建了全周期、多维度的网民诉求追踪体系。其不仅能实现对网民诉求的实时记录与动态归档，还可通过数据分析量化信息节点的度数、介数与接近度等核心指标，精准捕捉网民对“议题性事件”或“公共性话题”的认知倾向，智能完成网民意愿与诉求的收集、分析及归纳，更能对潜在诉求演变趋势进行预判，为决策者提供具有共识性的“最大公约数”诉求参考，有效加速网络舆情问题的确认进程。**另一方面**，生成式人工智能凭借先进的机器学习与自然语言处理技术，可挖掘网络空间中人类感知阈值以下的隐性信息，深度解析复杂语言结构，实现对海量非结构化数据的高精度处理。这一能力能辅助政府决策部门全面汇聚民意、精准识别民意核心，为厘清网络舆情成因提供全方位、高可靠性的信息支撑。

（二）有助于提升网络舆情决策效率

决策时效即决策行为的时间效力，其核心特征体现为决策方案需在特定时点、时段或期限内形成方可具备有效性，要求决策者在识别决策问题后，通过及时的信息采集、综合研判与

方案确定，实现决策与问题演化的动态适配；若决策过程存在“议而不决”“决而不断”的滞后性，不仅会导致决策效益衰减，更可能因问题性质发生根本性转变使决策失效。在“互联网 +”时代，网络舆情因传播主体的多元分散性、传播渠道的多样畅通性、传播方式的实时互动性及传播内容的真伪交织性，呈现出快速蔓延扩散与风险无限放大的特征，因此需通过及时制定网络舆情决策及应对措施防控风险，但传统信息采集存储与加工处理技术的局限性、行政管理条块分割与部门行政藩篱的制约，以及人力资源配置的适配性不足，均在一定程度上影响了网络舆情决策时效。

生成式人工智能技术与网络舆情决策的结合，有助于解决上述问题，提升网络舆情决策的效率。**一是全维度信息处理，强化舆情决策的信息支撑效率。**生成式人工智能具备强大的多模态信息采集与数据加工能力，在信息采集环节，可实现对网络舆情涉及的文字、图片、音频、视频等全类型信息的全域覆盖式采集与存储，解决传统采集方式的碎片化问题；在数据加工环节，能依据决策目标对非结构化、半结构化数据进行智能化处理，将原始数据转化为决策可用的结构化信息，显著提升舆情信息的处理效率与利用价值，为决策研判提供及时、全面的信息支撑。**二是跨域信息协同，打破舆情决策的信息流通壁垒。**生成式人工智能通过两大核心能力破解信息协同难题，一方面，依托数据共享平台的技术架构，打破政府部门间的“信息孤岛”，重构舆情信息的纵横向流转通道，实现不同层级、不同领域部门间舆情信息的实时交互；另一方面，凭借超强的自然

语言学习能力，可自动适配不同部门的数据交换标准，消解“数字鸿沟”带来的技术障碍，推动跨部门舆情数据的深度整合与高效共享，从管理与标准层面提升舆情信息的流通效率，为协同决策奠定基础。三**是人机协同决策，重构舆情决策的主体互动模式。**生成式人工智能重塑了网络舆情决策中的人机关系，其技术机制使人工智能从“被动式决策工具”升级为“智能化决策主体”，与人形成“协作增强型”互动模式。这种模式下，人机关系从传统的“人主导 - 工具反馈”转向“即时主动交互”，人工智能可承担决策过程中的信息筛选、风险预警、方案推演等重复性、高负荷任务，将人类决策者从海量事务中解放，聚焦于战略判断、价值权衡等核心环节，最终实现“机器高效处理 + 人类精准决策”的协同效应，显著提升网络舆情决策的时效性与科学性。

（三）有助于增强网络舆情决策预测

预测是主体依托既有知识体系与技术工具，对事物未来发展状态开展前瞻性判断的认知活动。在社会风险防控语境下，有效的风险预测是实现风险前置性防范的核心前提。具体到网络舆情领域，网络舆情决策预测特指政府决策部门基于网络舆情过往演化的客观历程与内在规律，运用科学决策技术与方法对舆情未来趋势进行系统性推演，其核心目标在于精准把握舆情动态演化轨迹、及时优化决策方案，并通过适配性政策选择达成舆情风险消解的最终诉求。鉴于网络舆情普遍存在“循环反转爆发”的演化特征，网络舆情决策需同时具备应对突发舆情的“灭火”能力与防范舆情滋生扩散的“防火”能力，但传统网络舆

情决策受限于技术工具与方法体系的局限性，其预测功能存在显著短板，难以满足“防火”需求。

生成式人工智能作为具备预测建模核心功能的先进技术平台，依托海量数据处理、算法推荐优化与模型动态推演三大核心能力，为突破传统决策预测瓶颈提供了系统性解决方案。**一是大数据关联分析，激活情感认知与行为的可预测性。**在大数据技术框架下，相关关系的价值得到极大释放，而情感认知与行为选择之间的内在关联，在海量数据支撑下具备了可预测的技术基础。生成式人工智能的大数据处理能力可通过算法推荐与模型推演，精准预测舆情客体的情感认知倾向，也能在回应客体情感认知的互动过程中，实现决策主体（政府）对客体情感与行为的引导，形成“预测 - 回应 - 引导”的闭环，为舆情决策预测提供底层逻辑支撑。**二是多维度舆情解析，构建趋势预测的信息提取机制。**生成式人工智能通过文本挖掘、情感分析、议题建模推演等技术手段，可从海量非结构化网络舆情数据中高效提取关键信息，其一，通过文本挖掘识别舆情议题的核心脉络与传播节点；其二，借助情感分析量化网民的情感倾向与核心诉求；其三，结合深度学习技术训练深层神经网络，从复杂舆情信息中提取具有标志性的“趋势关键词”，进而实现对网络舆情未来发展方向、热度变化、扩散路径的精准预测，为决策提供前瞻性依据。**三是自我学习与人机互动，优化预测模型的动态适配性。**一方面，能够在庞大复杂的舆情信息池中自主筛选高价值数据进行迭代学习与反复训练，持续优化算法推荐的精准度，提升预测结果的可靠性；另一方面，可将决策主体

与舆情客体纳入统一学习模型，形成“请求-回应”的人机互动范式 —— 通过捕捉人类决策反馈与网民行为数据，动态调整算法参数与模型结构，进一步修正预测偏差，实现预测准确性的持续提升，为网络舆情决策“防火”功能的落地提供技术保障。

二、巴中市网络舆情治理的现状

（一）巴中市网络舆情传播特征分析

巴中市网络舆情传播呈现技术迭代、文化破圈、民生为本的三维特征。未来需把握“AI 监测筑牢防线、文化 IP 凝聚共识、民生服务化解矛盾”的治理逻辑，尤其关注农村地区的技术渗透与年轻网民的话语转化，方能在数字化浪潮中实现舆情治理效能与文化传承的双赢。

一是传播平台以社交媒体主导，本地论坛与短视频成重要阵地。抖音、快手等平台凭借“短平快”特性，成为巴中舆情扩散的核心载体。例如，2025 年恩阳龙舟大赛视频播放量突破 1050 万次，通过“非遗 + 文旅”的内容设计，将地域文化转化为流量优势大；本地论坛深度参与，如麻辣论坛巴中板块日均发帖量超 200 条，以 25-45 岁中青年为主，形成独特政民对话生态，且经济发展类话题受高度关注；政务新媒体矩阵效能升级，“巴中发布”“通江文旅”等账号通过服务集成与内容创新提升传播力。例如，“巴中云上大学城”借助短视频矩阵，以沉浸式体验活动和高校合作推介吸引用户，半年内吸引 70 余所高校入驻。**二是传播内容以民生与公共政策主导，地域文化特色显著。**教育、医疗、交通等民生议题长期占据舆情焦点。例如，2025 年巴中市 12315 平台受理投诉举报 5596 件，同比

增13.19%，其中食品安全、预付式消费等问题占比超33%；地域文化深度渗透，光雾山红叶、四川方言等地方元素成为舆情传播的天然载体。微短剧《宝器》以方言和秘境探险为卖点，日均热度破2000万，成功将地域文化转化为流量优势。同时，川陕苏区历史讨论、红色旅游推荐等内容在论坛引发跨代际对话。**三是传播特征以裂变式扩散与情绪化表达并存。**传播速度快，扩散范围广，非理性表达与谣言交织。虚假信息或争议事件易在短时间内形成传播风暴。平昌县“青少年失踪”谣言经抖音传播后，获1.1万次转发，甚至引发群众拦截面包车的过激行为。类似地，人社局工作人员辱骂群众事件经问政平台曝光后，24小时内即登上川观新闻等主流媒体，倒逼涉事人员停职。

（二）巴中市网络舆情治理的现状

近年来，巴中市高度重视网络舆情治理工作，积极采取多种措施，致力于营造健康、有序、清朗的网络环境，取得了一定成效。

一是网络舆情监测预警体系初步成型。巴中市通过多部门协作与技术手段应用，已初步构建起覆盖多领域的网络舆情监测预警体系并有效发挥作用。其中，市场监管系统于2023年上线网络舆情监测系统，通过指定关键字对自媒体、网站、微博、微信等站点实时全景扫描，聚焦群众关心问题开展搜集、分析和统计，提升了自动转办效率；应急管理局在2024年建立完善突发事件舆情监测、快速反应、会商研判等工作机制，搭建安全生产和自然灾害等网上舆情监测平台，当年高效应对网络舆情事件9起，及时防范重大事故风险5起；2025年，

巴州区综治中心整合“智慧社区”“12345 热线”等数据,构建起“监测 — 研判 — 处置”闭环；明途科技则联合中国广电四川网络股份有限公司巴中市分公司，打造以“全方位、全过程、全覆盖、全天候”智慧化监测监管为标准的广播电视和网络视听监测平台，该平台运用大数据、人工智能等技术，对巴中市三县两区融媒体中心的网络视听内容进行统一监测监管，可快速精准识别涉黄、暴恐等敏感内容，并对监测数据及态势进行可视化展示，实现了公共舆情智能预警。**二是网络内容建设的正向引导。**打造“响网巴中”“理论大家讲”“小巴说政”等网络宣讲品牌，创新举办各类短视频大赛，如“巴中创文”短视频大赛等，鼓励创作积极向上的网络作品，弘扬主旋律。连续 5 年开展“网眼看巴中”网友线下交流活动，邀请自媒体人士参与，将巴中重点领域发展与互联网人士关注热点相结合，通过网络直播、短视频、图文等形式宣传巴中，传播网络正能量。例如，2024 年“网眼看巴中光雾赏杜鹃”新媒体集中创作活动，创作发布新媒体作品近 100 条次，全网传播量 200W+ ，提升了巴中旅游资源的知名度。同时，积极开展“网络中国节”系列主题活动，挖掘传统节日文化内涵，推出形式多样的融媒体产品，弘扬优秀传统文化。**三是网络生态治理成效显著。**针对网络直播及网络营销活动中的违法违规行为，利用网络安全宣传周等节点普及法律法规，提高网络人士法治意识，并开展专项行动查处违法行为，曝光典型案例。

三、生成式人工智能技术嵌入巴中市网络舆情决策的潜在风险

（一）算法偏见加剧网络舆情决策中的信息茧房效应风险

生成式人工智能技术嵌入巴中网络舆情决策的风险本质上是技术理性与地域社会复杂性碰撞的产物。随着生成式人工智能向纵深发展,AI 技术凭借其强大的内容生成与话语塑造能力,正在以隐蔽的方式对主流意识形态进行冲击,由于其依托海量互联网数据进行训练,表面上是在追求“价值中立”,实际上可在无形中放大边缘观点,弱化核心价值取向,让主流价值观的引导作用在不知不觉中被技术削弱。

一是算法内容选择偏向性加剧信息茧房效应。该技术基于用户画像的个性化推荐机制,通过强化学习持续推送契合用户既有认知的本地舆情信息(如特定民生议题的同质化内容),导致用户信息接收范围被算法逻辑框定,形成“认知窄化”现象,削弱公众对区域舆情生态的整体性感知。**二是算法推荐的群体分化效应助推舆论极化。**在巴中市舆情场域中,算法依据用户标签构建的信息茧房具有群体聚集特征,不同社会群体的舆情信息环境呈现显著异质性,这种信息接触的结构性差异削弱了公共话语空间的共识基础,导致群体间观点对立性增强,不利于形成建设性的舆情互动机制。**三是训练数据的样本偏差引发决策失准。**若算法训练数据过度依赖巴中市部分区域或特定群体的舆情样本,而忽略乡村地区、边缘群体的舆情表达,将导致模型对区域舆情的捕捉呈现“中心—边缘失衡状态,使舆情决策系统难以识别全域性舆情痛点,最终造成舆情研判的系统性偏差,侵蚀决策的科学性与公正性根基。

(二)监管缺失导致网络舆情决策中面临算法黑箱风险

生成式人工智能技术嵌入巴中市网络舆情决策的潜在风

险，其核心矛盾深刻根植于技术权责失衡这一底层逻辑，而这种失衡直接催生出算法黑箱与责任模糊两大突出问题。在网络舆情决策场景中，生成式人工智能技术应用的专业性与复杂性天然形成了信息壁垒，使基层决策者往往难以完全理解算法模型的训练数据来源、特征权重分配及推理逻辑，使得舆情分析过程如同被包裹在不透明的“黑箱”之中，进而引发次生舆情风险。

一是技术权力失衡导致的算法黑箱基础风险。地方治理主体因技术能力不足难以掌控外部供应商的通用模型，使得模型因巴中地域特色数据样本匮乏，形成对本地复杂舆情的选择性识别偏向，且模型对舆情要素的权重分配机制处于技术遮蔽状态，导致决策信息被算法隐性规则预先重构。**二是权责界定缺失引发的责任模糊传导风险。**当生成式人工智能技术生成的舆情分析存在认知偏差或引发伦理争议时，技术供应商以“模型输出仅为参考”推卸核心责任，地方政府因现行法律对 AI 辅助决策权责界定模糊而陷入追责被动，基层治理者成为责任下沉的最终承载者。**三是治理链条断裂导致的系统效能损耗风险。**上述算法黑箱与责任模糊相互叠加，形成“技术失控 — 责任空转 — 信任流失”的恶性循环，在巴中市技术监管资源有限的治理场景中，进一步偏离网络舆情决策的公共价值基准，削弱地方治理的科学性与公信力。

（三）技术滥用造成网络舆情决策中面临舆论操控风险

随着生成式人工智能技术的迅猛发展，其滥用问题正以更隐蔽、更具破坏性的方式侵蚀网络舆论生态，直接引发网络舆

情深度伪造与舆论操控的多重风险，给个人权益、社会秩序乃至国家安全都带来了前所未有的挑战。

一是深度伪造内容的感知混淆效应冲击舆情真实性基底。该技术通过生成逼真的虚假文本、图像或视频，利用视觉与语义的高度仿真性突破公众认知防线，导致巴中市舆情场域中“事实性信息”与“虚构性内容”的边界模糊化，增加舆情真伪鉴别的技术门槛与认知成本，削弱舆情决策的事实依据有效性。**二是算法驱动的舆论操控产业化加剧舆情治理复杂性。**在巴中市特定舆情场景中，生成式人工智能技术可批量生产带有定向情感倾向的舆情内容，通过算法推荐实现精准分发，形成“虚假信息—情绪放大—群体极化”的操控链条，这种低成本、高效率的舆论操控模式，可能被利益相关方利用以扭曲公众认知，干扰舆情自然演化进程，对舆情决策的客观性构成挑战。**三是深度伪造溯源机制的技术滞后性导致责任追溯困境。**当前生成式人工智能技术生成内容的溯源技术尚未形成标准化体系，在巴中市舆情事件中，一旦出现深度伪造引发的负面舆情，其内容生产者、传播者的责任界定面临技术瓶颈，不仅导致舆情处置缺乏精准靶向，还可能因追责不力助长虚假信息的扩散惯性，削弱舆情决策系统的风险预警与干预效能。

四、生成式人工智能技术在网络舆情决策中的风险防控路径

（一）技术治理：打破算法黑箱，提升透明度

在当前人工智能深度介入网络舆论治理过程这一背景下，应发挥人工智能技术在其中的积极作用，明确相关法律法规、界定人工智能应用行为的主体责任、打破“算法黑箱”，构建多

元的治理体系。

一是建立内容审核机制。研发专门针对生成式人工智能所生成内容的审核技术，利用图像识别、自然语言处理等技术手段，对生成的文字、图片、视频等内容进行实时监测和筛选，及时发现并拦截虚假信息、有害信息和不良信息。例如，对于AI生成的新闻报道，可通过事实核查算法，与权威数据库进行比对，验证信息的真实性。**二是强化算法透明度。**要做到算法透明，就要打破平台的“算法黑箱”，算法技术的开发者需要加强对算法的伦理审查，要督促算法应用的互联网企业及时、合理、有效地公开算法基本原理、优化目标、决策标准等信息，提高算法决策过程的透明化和可解释化，使得人们能够理解算法决策的原因和依据。**三是建立多元优化算法机制。**通过多算法协同、动态迭代、交叉验证等方式提升舆情决策的准确性与可靠性，形成“技术互补矩阵”，覆盖舆情分析的全流程，是防控技术滥用风险的核心路径。此外，建立对算法的公平性和偏见性评估机制，确保算法决策不受到个人特征和社会因素的影响，减少“信息茧房”效应对网民的影响，从而减轻网络舆论极化，增强网络舆论共识度。

（二）法律规制：构建多层次监管框架

生成式人工智能技术在网络舆情决策中的广泛应用，既带来了效率提升，也引发了数据滥用、虚假信息传播、算法歧视等法律风险。构建多层次监管框架，通过立法完善、执法协同、司法保障及合规指引，为生成式人工智能技术在网络舆情应用划定法律边界，是风险防控的核心制度路径。

一是专项立法明确核心规制对象。针对生成式人工智能技术在舆情领域的特殊性，需在现有法律体系基础上制定专项规则。界定“生成式人工智能舆情工具”的法律属性，明确开发者、运营者、使用者的权责划分。针对舆情场景中的高频风险，在《网络安全法》《数据安全法》《个人信息保护法》基础上，制定《生成式人工智能服务管理暂行办法》等专项规章，细化禁止性条款。此外，明确舆情数据采集的“合法、正当、必要”原则，禁止通过爬虫技术非法抓取非公开数据，要求对敏感数据的采集需单独获得同意。建立“数据来源追溯制度”，生成式人工智能训练数据中若包含舆情相关内容，需记录数据来源、授权情况及清洗流程，避免使用侵权或非法数据训练模型。**二是建立“主责 + 协同”的监管体系。**明确监管主体与职责分工，网信部门作为核心监管主体，统筹生成式人工智能舆情应用的全流程监管；公安、市场监管、行业主管部门按领域分工，监管特定行业的舆情 AI 应用。针对跨境舆情 AI 服务，由网信部门联合海关、国家安全机关建立准入审查与动态监测机制，防范数据出境与意识形态风险。**三是明确权利救济与责任追究路径。**当公民或企业因生成式人工智能舆情分析遭受权益损害，可依据《民事诉讼法》《个人信息保护法》提起诉讼，要求侵权方停止侵害、赔礼道歉、赔偿损失。建立“公益诉讼”制度，针对 AI 生成虚假舆情导致公共利益受损的情况，检察机关或社会组织可提起公益诉讼，要求责任方消除影响。

（三）伦理引导：推动“技术向善”的价值观

生成式人工智能在网络舆情决策中的应用，本质上是技术

工具对社会舆论生态的介入与干预。若缺乏伦理约束，其可能沦为放大偏见、操纵认知、侵犯隐私的工具。以“技术向善”为核心的伦理引导，并非简单的道德说教，而是通过价值观塑造、规则嵌入、责任绑定，为技术应用划定“不可逾越的红线”，从源头防控伦理风险。

一是将伦理准则嵌入技术全生命周期。生成式人工智能舆情工具在研发阶段确立伦理设计优先原则，制定伦理开发指南开展伦理影响评估，引入由多方组成的伦理审查委员会进行前置审查，如情感分析算法通过多元语料库减少偏见；应用阶段明确禁止性应用清单，推行伦理使用承诺书制度接受监督，像某城市疫情舆情监测仅用匿名化数据；迭代阶段建立伦理投诉反馈通道，将偏见指数、隐私保护得分等伦理指标纳入算法优化体系。**二是强化多元主体的伦理责任。**技术开发者需筑牢伦理底线，承担算法伦理设计责任，在代码中嵌入伦理护栏过滤不良分析结果，同时公开技术原理与局限性避免夸大效果；应用方要坚守伦理决策原则，拒绝算法迷信，将 AI 结果仅作为参考并结合人工等多方意见，且需承担结果伦理责任，对决策影响负责并整改失误；监管与社会应强化伦理监督，监管部门制定伦理监管细则明确处罚标准并定期检查，公众、媒体和学术机构则分别发挥监督哨作用、曝光滥用案例及发布白皮书评估现状并提改进建议。**三是要培育“技术向善”的文化共识。**开展伦理教育与科普，对技术开发者将“AI 伦理”纳入必修培训，借案例教学强化伦理意识，培养技术与责任并重思维；对公众通过科普内容解读工具原理与风险，提升认知以避免盲目态度。

推广伦理实践标杆案例，评选应用典范总结可复制经验形成示范效应，建立激励机制对伦理技术突破团队给予支持与奖励，引导行业向善竞争。推动跨领域伦理对话，定期举办论坛邀请多方讨论伦理争议问题凝聚共识，建立动态响应机制，针对新型伦理问题及时论证并更新规范。

（四）防御体系：构建“AI+安全”协同生态

生成式人工智能在提升网络舆情分析效率的同时，也带来了深度伪造内容泛滥、舆情操纵智能化、决策偏见放大等新型风险。构建“AI + 安全”协同生态，通过技术、主体、机制的多维融合，形成覆盖“事前预防 - 事中监测 - 事后处置”的全链条防御体系，是防控生成式 AI 舆情风险的根本保障。

一是打造“AI 驱动的智能盾牌”。通过 AI 技术与安全技术的深度融合，构建抵御生成式 AI 舆情风险的技术屏障。构建生成内容溯源与鉴伪系统，开发基于大模型的多模态鉴伪算法，利用 AI 生成内容的指纹特征实时识别虚假内容，同时通过区块链技术对关键信息存证，确保可追溯、不可篡改以压缩伪造内容空间；建立舆情风险动态监测与预警系统，借助知识图谱构建风险关联分析引擎识别异常传播模式提前预警人工操纵型舆情，通过实时语义变异捕捉算法动态更新语义库避免风险漏判；打造决策安全增强系统，利用算法偏见检测与修正工具识别隐性偏见并生成修正建议，开发反制性生成技术的 AI 回应优化系统，针对谣言或负面舆情自动生成辟谣内容以降低人工回应问题。**二是构建“多元共治”的防御共同体。**明确不同主体防御职责，构建“政府主导、企业主责、社会协同、公众参与”的生态

格局。政府部门出台安全标准明确企业责任边界，建立国家级风险监测中心实现跨域协同处置；科技企业中互联网平台部署鉴伪工具前置拦截违规内容,AI 技术企业在设计中嵌入“安全基因”降低滥用风险；社会组织与专业机构通过行业自律公约、第三方审计及交叉研究提供监督支撑与技术人才保障；公众则通过科普提升辨别能力，并借助举报渠道参与监督，形成全民防御补充力量。**三是确保生态高效运转的制度支撑。**建立风险共享与联动处置机制，由政府、企业、机构共享典型风险案例形成数据库供参考，同时制定跨主体联动流程，在重大风险出现时政府启动应急、平台限流、机构提供技术支持实现“发现即处置”;设立技术迭代与标准更新机制,通过“AI 防御技术沙盒”鼓励测试新型防御技术加速落地，依据技术迭代速度定期更新安全标准与检测指标避免滞后；构建伦理约束与问责机制，制定伦理准则禁止危害社会稳定的舆情内容传播并明确法律责任，建立分级问责制度，对未落实措施的企业追责、对恶意操纵者严惩以形成制度威慑。

五、结语

生成式人工智能技术嵌入巴中网络舆情决策，既是数字时代治理现代化的必然选择，也是提升舆情响应效能的重要契机。其在舆情监测的全面性、分析的深度性与决策的前瞻性上所展现的价值，为巴中构建清朗网络空间、化解社会矛盾提供了技术赋能的新路径。然而，技术应用伴随的算法偏见、数据安全、伦理失范等风险，也警示我们在拥抱技术红利时必须保持理性克制。

防控风险并非否定技术，而是为了让技术更好地服务于治理需求。巴中在推进生成式人工智能与舆情决策融合的过程中，需始终坚持“技术向善”的导向，以制度建设筑牢风险防线，以技术创新提升防控能力，以多元协同凝聚治理合力。唯有将技术优势与本土治理经验深度结合，在价值挖掘中规避风险，在风险防控中释放价值，才能让生成式人工智能真正成为感知社情民意的“千里眼”、化解舆情危机的“智慧脑”、服务群众需求的“连心桥”，为巴中推进国家治理体系和治理能力现代化注入持久的数字动力。