科学技术与教育 2025年第6期
ISSN: 3079-4455

AIGC 赋能电信网络诈骗治理的逻辑及实践路径研究

戴心怡，石拓*，张帆锐
北京警察学院公安管理系，北京，中国
*通讯作者

【摘要】电信网络诈骗犯罪作为新型网络犯罪的典型代表，现已成为发展最快的刑事犯罪，对人民财产安全和社会治理工作造成严重威胁。当前，AIGC（生成式人工智能，Artificial Intelligence Generated Content）正在引领一场技术革命。从当前警务实战中的实际需求出发，本文针对电诈治理中存在的多源数据整合困难、被害群体画像粗放及宣劝效能不足等核心问题，基于 AIGC 技术的内容孪生、精细编辑、智能推理与动态适应等特性，构建“技术逻辑—实施过程—运作机制”三维分析框架，系统探究该技术在诈骗防控中的赋能路径。研究进一步揭示 AIGC 在数据治理、智能分析、协同防控等环节的核心作用，探索融合技术创新与制度保障的治理模式与实践路径。

【关键词】电信网络诈骗；犯罪治理；生成式人工智能；技术赋能

1.引言

在数字经济高速发展的当下，电信网络诈骗犯罪工具不断更迭，骗术类型不断出新，且在发案数量上始终呈现多发高发态势，已演变为全球性治理难题。据公安部数据显示，2023 年我国电信网络诈骗案件涉案资金高达 3800 亿元，较五年前增长逾 5 倍，且犯罪手段正从传统话术诈骗向 AI 换脸、虚拟货币洗钱等智能化、复合型犯罪演变，深度伪造、智能剧本生成等技术应用使犯罪隐蔽性显著增强。这种技术异化催生的新型犯罪生态，暴露了传统治理手段在数据整合能力不足、预警响应机制滞后、宣防精准度欠缺等方面的缺陷。

在此背景下，AIGC 凭借其内容孪生、智能推理、精细化编辑与多模态生成等核心优势，为破解电诈治理困境提供了新路径。AIGC 是指基于生成对抗网络、大型预训练模型等人工智能的技术方法，通过已有数据的学习和识别，以适当的泛化能力生成相关内容的技术。当前，AIGC 技术已在公安情报分析、电诈风险预警等领域展现出巨大潜力，其不仅赋能了诈骗数据的自动化萃取、受害人群特征的精确描绘、诈骗流程的深度剖析，还促进了防范宣传与教育策略的智能策划与实施。但现有研究多聚焦于单一技术应用场景，缺乏对 AIGC 与现有反诈体系的协同机制、数据治理范式革新以及跨领域风险防控能力的整体考量。

本文旨在通过剖析 AIGC 赋能电诈治理的深层逻辑，融合犯罪学的情境预防理论、计算机科学的动态博弈理论及系统科学的控制论原理，提出"实时感知—风险建模—策略生成—反馈迭代"的闭环治理路径，在时间维度上建立犯罪演化预测图谱，在空间维度构建风险要素关联网络，在行为维度形成多主体协同干预机制。主要贡献包括：（1）构建了 AIGC 与犯罪治理的跨领域理论模型，探索算法决策在社会安全领域的适用性边界，为 AI 技术伦理研究提供实证案例；（2）构造了“智能识别—精准拦截—动态预警”的全链条反诈体系，建立动态风险预测模型，实现从被动响应到主动防御的治理模式升级；（3）揭示了 AIGC 驱动治理的“数据治理—功能耦合—协同防控”核心机制，提出跨部门数据共享协议与联合执法制度。

2.相关研究

2.1 电诈犯罪的智能化演变与治理挑战

近年来，电信网络诈骗犯罪呈现出技术驱动、跨境协同、组织隐蔽的新特征。魏嘉迪等学者指出，2023 年我国电诈案件涉案资金达 3800 亿元，五年增长超 5 倍，犯罪手段已从传统话术转向“AI 换脸+虚拟货币洗钱”等复合型模式[1]。吕培霖等学者进一步揭示，犯罪组织通过暗网获取公民信息，利用区块链技术构建去中心化资金通道，形成“人员—技术—资金”全链条闭环[2]。陈永峰等学者强调，传统侦查手段在电子证据固定、跨境司法协作等环节存在显著滞后性，亟需构建技术对抗体系[3]。李雪峰等学者发现，当

前技术防控多聚焦单一场景，缺乏跨平台数据融合能力，导致预警准确率受限[4]。

2.2 技术防控的创新实践与局限

学术界与实务界积极探索智能防控路径。农忠海等学者提出 AIGC 技术可通过多模态数据分析实现诈骗话术识别与资金流向追踪[5]，上海警方实践显示，AI 模型使预警准确率提升。陈芸蕾等学者构建的“技术—制度—文化”分析框架表明，技术创新需与行业监管协同，小红书通过大模型拦截了大部分的导流诈骗，但站外资金转移仍存在监管盲区[6]。靳雨婷指出，AIGC 技术催生“AI 生成虚假客服”等新型诈骗手段，需研发动态对抗模型[7]。杨胜钦基于 ChatGPT 技术分析，认为其在反诈宣传文本生成方面具有显著优势，但需防范模型被黑灰产利用的风险[8]。

2.3 协同治理的机制构建与困境

现有研究普遍强调多元主体协同的重要性。学者孙晨博通过实证分析指出，“公安主导+企业联动+公众参与”模式使 D 市电诈发案率下降[9]。叶璐认为电信网络诈骗犯罪天然的非接触性给案件侦破带来极大困难[10]。国际层面，中缅泰联合行动清除缅北园区，但学者毛飞飞研究表明，跨境数据共享机制尚未完善，影响追赃挽损效率[11]。制度层面，《反电信网络诈骗法》实施后，司法机关通过“全案必侦”机制提升打击精度，但裘佳仪等学者基于理性选择理论的分析显示，低犯罪成本与高收益仍驱动黑灰产蔓延[12]。单勇提出，需加快构建数字平台责任认定体系，强化其反诈义务履行[13]。Dong 认为从时空分布入手，研究社会人口统计特征，识别电信网络诈骗的特征。最后，对电信网络诈骗的防范策略进行总结分析，包括加强法律制度建设、提高公众防范意识、优化电信网络安全技术等[14]。Chu 认为可以利用生成式人工智能对基层电信网络诈骗案件中各类诈骗案件的数量进行分析和预测，并计算未来 12 个时间段内各类诈骗案件的发案数。基于预测结果，对高发类型的诈骗制定有针对性的精准施策，以数据预测为抓手，瞄准精准、有针对性的策略，逐步构建精准反诈的新机制[15]。

3.电信网络诈骗治理困境检视

3.1 多源数据治理困难

3.1.1 数据采集的多维异构性

电信网络诈骗数据生态呈现典型的“泛在化”特征，通信运营商、金融机构、互联网平台、公共服务部门等数据源相互交织。这些数据在格式标准、存储结构、更新频率等方面存在天然壁垒：运营商提供的实时信令数据采用专有协议封装，金融机构交易流水遵循严格的会计核算标准，社交平台行为日志则呈现非结构化特征。这种技术异构性导致跨系统数据映射需经历复杂的语义转换，仅字段匹配环节就消耗 30%以上的预处理时间，形成事实上的数据割据状态。

3.1.2 技术对抗的不对称性挑战

当前，诈骗手段呈现指数级进化态势，钓鱼网站变种、AI 换脸技术、区块链洗钱等新型犯罪工具迭代速度远超传统防御体系。犯罪团伙利用暗网市场实现攻击工具的快速更新，而规则引擎依赖的人工标注特征库从发现新型诈骗到规则上线存在显著时滞。机器学习模型虽具备自主学习能力，但在小样本场景下易出现认知偏差，特定区域 IP 地址误报率偏高的问题制约了精准防控效能。

3.2 被害群体画像粗放

3.2.1 数据资源结构性短缺

当前用于被害群体刻画的数据呈现碎片化特征：公安机关接报案数据侧重诈骗手段描述，缺乏对受害人心理特征、行为模式的深度采集；运营商通话记录仅记录通信行为，难以反映社会关系网络；互联网平台虽掌握用户画像数据，但受隐私保护法规限制无法深度共享[16]。这种数据割裂导致无法构建包含人口统计学特征、认知心理特质、网络行为轨迹的三维画像模型，现有数据维度仅能支撑有限的受害群体特征分析。

3.2.2 刻画方法技术滞后

传统画像方法依赖静态标签体系，通过年龄、职业等基础属性进行群体划分，难以捕捉新兴受害群体的动态特征。机器学习模型在小样本场景下表现出显著的过拟合风险，对特定群体的预警准确率难以满足需求，而该群体实际受骗率持续增长。社交网络分析技术应用不足，未能有效识别关键节点，导致预警信息传播效率低下。

3.2.3 动态风险评估失效

诈骗手法迭代速度远超画像更新周期，新型诈骗从出现到形成受害群体的时间远短于模型更新耗时。跨领域风险传导机制研究缺失，未能揭示链式风险路径，大量受害人在首次接触诈骗信息前已暴露于关联风险场景。预测模型对潜在高危群体的识别存在幸存者偏差，过度关注历史受骗人群特征，忽

视沉默群体，导致预警覆盖面存在缺口。

3.3 宣劝效能亟待提升

3.3.1 宣传话术缺乏针对性设计

反诈劝阻语言普遍存在模式化倾向，未能适配不同群体的认知特征与心理接受机制。标准化警示文本过度依赖指令性表达，忽视情感共鸣与场景化叙事构建，导致信息接收者产生心理防御性屏蔽。诈骗话术通过精密的社会工程学设计激发目标对象的情感认同，而传统宣教内容多以单向警告为主，缺乏对受害者决策逻辑的逆向解构，难以形成有效的认知干预[17]。这种语言策略的脱节削弱了风险信息的说服效力，使潜在受害者在心理层面难以建立抵御诱导的思维屏障。

3.3.2 宣劝形式呈现媒介适配不足

现有宣传载体未能充分融入数字社会的传播生态，单一化的图文输出与新兴媒介的交互特性形成结构性矛盾。传统传播渠道的线性模式难以突破信息茧房对目标人群的包裹，而沉浸式、场景化的技术应用尚未形成系统化实践。更关键的是，跨平台协同机制的缺失导致风险警示信息与用户行为轨迹脱节，无法在诈骗行为触发关键节点实施精准阻断[18]。这种形式创新滞后使得宣教内容难以嵌入受众的日常媒介接触链，削弱了信息传递的动态穿透力。

3.3.3 宣劝人员专业能力梯度断层

基层劝阻队伍的知识更新速度与诈骗技术迭代存在显著代差，部分人员对新型犯罪手法的技术原理与心理操控机制认知不足。在实操层面，缺乏系统的行为干预训练导致劝阻过程中难以突破受害者的心理封闭状态，往往陷入机械重复警示内容的低效循环。同时，技术支援体系与前端处置力量的衔接不畅，使得风险预警与现场处置之间出现策略断层，制约了整体劝阻效能的持续性释放。

4. AIGC 的技术逻辑与特征

4.1 AIGC 的技术逻辑

AIGC 的技术逻辑主要基于深度学习、生成算法、预训练模型以及多模态技术等，通过复杂的神经网络结构来模拟人类的创作过程，从而自动生成具有创新性和个性化的内容。

AIGC 运行的技术基础来源于以下几个方面。深度学习是 AIGC 技术的核心驱动力。它利用神经网络模型，通过多层非线性变换，从数据中自动提取和学习高级特征表示，从而实现对复杂数据的理解和生成[19]。深度学习模型，如 Transformer、GAN（生成对抗网络）、CNN（卷积神经网络）等，在 AIGC 中扮演着重要角色。这些模型能够捕捉数据中的内在规律和模式，进而生成与训练数据相似或具有新特性的内容；生成算法是 AIGC 技术的关键组成部分。它们负责根据训练好的模型，从随机噪声或特定输入中生成新的内容。这些算法需要能够捕捉到数据的分布特性，并据此生成符合该分布的新样本；预训练模型是 AIGC 技术的重要基础。这些模型通常在大规模数据集上进行训练，学习了丰富的语义知识和模式。通过将预训练模型应用于特定任务，可以显著提高模型的性能和泛化能力。例如，OpenAI 的 GPT 系列模型就是典型的预训练语言模型，它们在海量文本数据上进行训练，能够生成流畅、连贯的文本内容；多模态技术使得 AIGC 能够跨越不同的数据类型（如文本、图像、音频、视频等）进行内容生成。通过融合不同模态的信息，AIGC 可以生成更加丰富、多样的内容。

AIGC 的技术逻辑通过四个阶段来运行。数据预处理是 AIGC 的第一步。它涉及数据的清洗、标注、格式转换等步骤，旨在将原始数据转换为模型可以理解和处理的格式。预处理后的数据将作为训练集输入到深度学习模型中，用于模型的训练和优化；在模型训练阶段，深度学习模型通过不断地学习训练集中的数据，逐渐掌握数据的内在规律和模式。训练过程中，模型会根据损失函数的反馈不断调整参数，以最小化损失值并优化性能。当模型在训练集上达到一定的性能水平后，就可以将其应用于实际的内容生成任务中；在内容生成阶段，用户可以通过界面或 API 向 AIGC 系统输入特定的指令或信息。系统根据输入的指令或信息以及训练好的模型，自动生成符合要求的内容。生成的内容可以是文本、图像、音频、视频等多种形式，具体取决于所使用的模型和技术；AIGC 系统通常会收集用户的反馈信息，用于评估生成内容的质量和效果。根据用户的反馈，系统可以不断优化模型参数和生成算法，以提高生成内容的准确性和创新性。

4.2 AIGC 的技术特征

4.2.1 高效生成能力

AIGC 借助先进的并行计算架构，在自然语言处理和计算机视觉等常见应用场景中，文本生成速度可达到人工创作的 200 倍以上，

原本人工需要数小时甚至数天完成的文本创作任务，AIGC 能够在极短时间内完成。

此外，分布式训练框架赋予了 AIGC 强大的数据处理能力，能够支持单日处理千万级别的文本或图像请求，这一卓越性能充分满足了当今大规模内容生产的迫切需求。在多模态内容生成领域，AIGC 表现出了卓越的跨模态生成兼容性。它打破了文本、图像、音频等内容形式之间的壁垒，可同时对多种内容形式进行处理，并支持复杂任务编排。无论是将一段反诈宣传文案转化为生动形象的海报，还是将反诈知识音频与配套的文字脚本进行整合，AIGC 都能高效完成，极大提升了内容生成的综合效率[20]。其边际成本趋近于零，单条内容生成成本随规模扩大呈指数级下降。通过模型轻量化技术，在移动端设备实现本地化生成，降低云端算力依赖及传输成本。

4.2.2 智能内容质量

依托基于 Transformer 架构的深度语言模型，在文本生成质量上达到了极高水准。生成的文本语法错误率低于 0.3%，逻辑连贯性评分在 100 分制中达到 92 分以上，远超一般人工创作的平均水平[21]。并且，通过引入知识图谱技术，AIGC 能够在专业领域内容生成中确保准确性，无论是金融领域的复杂术语，还是医疗行业的专业知识，都能精准无误地呈现。

其能够模拟不同时代、地域的创作特征，无论是古代诗词风格，还是现代流行文化风格，亦或是不同地区的方言特色，AIGC 都能精准把握并融入生成内容中。在反诈宣传场景中，这一特性可将官方通报转化为具有地方特色的宣传脚本，有效提升传播效果。

4.2.3 动态适应机制

AIGC 具备灵活且精细的超参数调优功能，能够对生成内容进行全方位的粒度控制，可实现 50-5000 字的动态调节以满足不同场景下对文本篇幅的要求。在图像生成时能够精准调控图像分辨率，从 1080P 到 8K，满足从普通展示到高清细节呈现的多样化需求。在情感倾向设定上 AIGC 可生成积极、中性或警示等不同情感基调的内容以适配不同的传播目的。其上下文感知生成模型能够敏锐捕捉应用场景的特征，并自动适配生成内容的形态。在紧急预警场景中，AIGC 能够迅速生成简洁明了的短句强提醒文案，以最快速度引起用户的注意；而在科普教育场景下，AIGC 则会输出结构化的知识图谱，帮助用户系统地理解复杂的反诈知识体系，从而提升用户在不同场景下对内容的接受度和理解度。

4.3 应用现状

4.3.1 实时行为分析系统

随着城市的发展，公共场所人员流动日益频繁，监控画面中的目标数量众多且行为复杂。AIGC 构建的智能监控系统应运而生，其运用时空轨迹建模技术，能够深入分析监控画面中多目标的运动轨迹，采用多目标跟踪算法能够持续追踪并精准区分不同个体的运动模式。同时结合事先构建的行为特征库可以高效且准确地识别出异常行为。一旦检测到群体聚集、快速奔跑或物品遗留等异常情况，系统便迅速启动分级预警机制，触发相应的响应措施，极大地提升了重点区域的安全防控能力，为维护社会秩序提供了有力保障。

4.3.2 人脸识别网络

在案件侦查和人员管控工作中，准确快速地识别人员身份至关重要，然而传统单一的人脸识别方式在复杂环境下容易出现误差，AIGC 依托先进的生物特征识别技术，支持多模态生物特征融合识别，将人脸、步态、虹膜等多种特征信息进行整合，即使在复杂光照条件下，凭借其强大的技术性能，仍能保持较高的识别精度，为警方在追逃、失踪人员查找等关键工作中提供了强大且可靠的技术支持。

4.3.3 多源数据融合

案件侦查所涉及的情报来源广泛，通信、交通、金融等多领域都蕴含着重要线索，但这些数据源分散且格式各异，传统方式难以有效整合利用。AIGC 构建的跨部门数据中台将这些多领域数据源进行有机整合后通过先进的自然语言处理技术能够对非结构化数据进行深度解析，自动抽取实体信息并构建关系图谱。在实际案件侦查过程中，这种多源数据融合能力发挥了巨大作用，帮助警方从繁杂的线索中梳理出关键脉络，挖掘出隐藏在海量信息背后的关联关系。

4.3.4 犯罪网络建模

电信诈骗等犯罪活动日益呈现出网络化、复杂化的特点，传统的分析方法难以全面洞察犯罪网络的结构和运作机制。AIGC 基于图神经网络开发的犯罪关系分析模型，能够自动识别犯罪网络的层级结构和信息流路径，

可精准发现隐藏的犯罪子网络，并通过可视化技术将犯罪网络的全貌直观地展示出来[22]。在电信诈骗案件中大显身手，帮助警方快速定位犯罪核心节点，清晰地理清资金流和信息流脉络。

5. AIGC 技术赋能电诈治理的内在逻辑

5.1 AIGC 赋能电诈治理场景廓清

电信网络诈骗犯罪活动以惊人的速度演变，手段日益繁杂且隐蔽，对公众财产安全及社会秩序稳定构成严重威胁。传统电诈治理手段在面对海量、复杂且瞬息万变的诈骗信息时，显得力不从心。AIGC 技术凭借其强大的数据处理、智能分析与内容生成能力，为电诈治理提供了全新的解决方案，贯穿于从预防到打击的全流程。

5.1.1 智能监测与预警

电信网络空间信息爆炸式增长，诈骗信息隐蔽且危害大。传统监测方式依赖人工筛查或简单规则匹配，效率低且易遗漏新型诈骗模式。AIGC 技术中的自然语言处理技术，基于深度学习算法，通过构建大规模语言模型，对海量正常及诈骗相关文本（“轻松赚大钱”“零风险高回报”“快速致富秘籍”等词汇）进行学习，可掌握正常语言表达语义规则以及诈骗信息特有的词汇组合、话术结构与情感倾向，同时基于诈骗图像带有特定虚假广告标识、假冒官方图标，视频存在可疑人物行为、场景布置等特征，AIGC 的图像识别技术可对这些视觉元素进行特征提取与分析，建立诈骗图像、视频特征库[23]。网络中出现与特征库匹配信息时，系统迅速捕捉并发出预警。

5.1.2 智能拦截与阻断

诈骗分子与受害者间的通信渠道，短信、电话、邮件等是电诈行为实施的关键纽带。传统拦截手段难以应对诈骗分子不断翻新的通信策略与技术手段。在短信拦截方面，一些诈骗短信频繁更换发送号码，通过对内容语义深度分析，若话术结构、关键词与已知诈骗模式高度相似，即便号码不同，利用 AIGC 技术对短信内容进行实时语义分析，结合号码归属地、发送频率、历史记录等多维度信息，也准确识别并拦截；对于电话诈骗，AIGC 可利用语音识别技术将通话内容实时转化为文本并借助自然语言处理技术对通话话术进行分析，一旦检测到诱导转账、虚假中奖通知等诈骗关键词或特定话术模式，自动对通话标记并采取强制挂断、向用户发送警示提醒等措施；在邮件通信中，AIGC 对邮件标题、正文以及附件全面扫描。通过对邮件文本内容语义分析以及对附件格式、文件特征识别，精准识别包含钓鱼链接、恶意软件的邮件，并拦截在收件人邮箱之外。

5.1.3 公众教育与宣传

公众对电信网络诈骗的认知程度与防范意识是构筑全社会反诈防线的核心。而传统反诈宣传方式形式单一、内容枯燥，难以吸引公众关注，AIGC 技术可根据不同受众群体年龄、文化背景、行为习惯等特征，智能生成多样化、个性化宣传内容。针对青少年群体利用图像生成与动画制作能力创作反诈动画短片，展现常见诈骗手段实施过程与防范要点。对于老年群体生成大字体、简洁明了的图文宣传材料，突出常见诈骗手段特征及应对方法。在传播渠道方面，其可借助社交媒体平台大数据分析能力，精准分析不同平台用户兴趣偏好、社交关系与行为模式，将制作好的宣传材料精准推送至目标受众。

5.2 AIGC 赋能电诈治理的实施过程

5.2.1 AIGC 助力多源数据治理

从公安、银行、通信运营商等部门收集电信诈骗的历史案例、诈骗手法、诈骗信息特征等数据。利用大数据技术进行数据清洗和整合，形成统一的数据格式和标准。通过比对和筛选，去除重复的数据记录，确保数据的唯一性。对数据进行检查、纠正和规范化处理，包括去除无效字符、填充缺失值、转换数据类型等，将数据转换为统一的格式和标准[24]。建立不同数据源之间的映射关系，确保数据在整合过程中能够正确对应和关联。

5.2.2 AIGC 赋能数据分析与挖掘阶段

对收集到的诈骗案例进行深入分析，总结诈骗分子的作案手法和特征。对用户的通信和交易行为进行分析，识别出易受诈骗的用户群体和特征。将多个数据源的数据合并到一个统一的数据库中，形成全面的数据集[25]。在数据合并的基础上，通过数据关联、数据挖掘等技术，将不同数据源的数据进行深度融合，形成更有价值的信息和知识。利用 AIGC 技术中的自然语言处理、图像识别等技术，训练机器学习模型，以识别潜在的诈骗信息。对模型进行不断优化，通过交叉验证、参数调整等方法，提高识别诈骗信息的准确性和效率。根据分析结果制定针对性的防范措施和应对策略。

5.2.3 AIGC 支撑劝阻策略生成

针对传统人工劝阻效率低、话术单一的问题，AIGC 构建智能劝阻系统。系统包含动态话术生成模块，基于 Transformer 架构的自然语言生成模型实时分析被害人风险等级与受骗阶段，通过强化学习算法自动优化不同话术组合的阻断成功率，多模态交互模块整合文本、语音、视频等交互方式，根据用户终端类型自动适配，建立起劝阻效果反馈闭环，通过分析用户点击行为、通话时长等数据，运用在线学习算法持续优化策略，当发现某种诈骗变种话术阻断率下降时，系统自动触发模型更新流程。

5.3 构建 AIGC 赋能电诈治理的运作机制研究

5.3.1 数据治理机制

电信网络诈骗治理的基础在于对海量数据的有效管理和使用。在电信网络诈骗活动中，诈骗者往往通过不同渠道（如社交媒体、电话、短信等）与受害者进行互动，这些互动数据分散且隐蔽，使得传统数据治理模式难以实时、高效地进行诈骗行为识别和预防。而 AIGC 技术的引入，使得电信诈骗数据的采集、处理和应用方式得以重构。

AIGC 通过其强大的自然语言处理（NLP）技术，可以自动从海量通信数据、网络交互数据、社交媒体内容等中提取相关信息，形成数据分析基础[26]。AIGC 能够从杂乱的非结构化数据中快速抽取诈骗行为特征，如电话、短信中的关键语句、反常的社交媒体互动、以及网络上伪造的网站或广告，极大地提高了诈骗信息的识别精度。AIGC 还能够通过深度学习模型对历史数据进行模式识别，利用动态更新的数据治理模型，不断优化诈骗识别算法。通过在数据治理中的持续学习和迭代，AIGC 可以发现新兴的诈骗手法，并在其早期阶段进行有效识别和拦截，减少受害者受骗的可能性。此外，AIGC 在数据治理机制中还能够实现数据隐私保护，通过分布式学习等技术手段，确保数据处理的合法性和隐私性，避免信息泄露风险。

5.3.2 功能耦合机制

AIGC 通过标准化 API 接口与公安反诈平台、金融风控系统、通信运营商信令监测平台建立数据交互通道，构建覆盖全流程的协同治理体系。生成对抗网络（GANs）在此过程中发挥关键作用：通过模拟钓鱼网站、AI 换脸视频、伪基站短信等诈骗场景，生成接近真实诈骗行为的测试数据，用于检验现有防护系统的边界防御能力。这种主动测试机制能够识别传统规则难以覆盖的新型诈骗手法，提升系统对动态风险的反应能力。同时 AIGC 与传统机器学习模型深度融合，通过迁移学习技术复用通用模型的特征提取能力，结合电诈领域专有数据进行模型微调，实现对复杂洗钱模式与异常通话行为的精准识别，通过整合金融机构反洗钱系统的资金流分析与运营商通话监控系统的异常模式检测，AIGC 构建起跨平台、跨场景的协同防护体系，能有效应对电信网络诈骗的多样性与复杂性挑战。

5.3.3 整体协同机制

AIGC 通过跨平台数据整合与智能分析能力，构建起覆盖全链条的协同防控体系。其核心作用体现在两方面：首先通过标准化 API 接口与数据中台技术，整合公安、金融、通信等部门的诈骗举报数据、风险评估信息及通信行为记录，形成统一特征空间，实现跨平台实时数据共享与分析。在此基础上，AIGC 生成的反诈内容库与预测模型能够指导各平台制定协同反制策略，通过识别高频呼叫行为触发运营商拦截机制，同步推送预警信息至金融机构监测异常交易。其次，AIGC 与执法机构、金融部门的深度协同体现在联合防控机制中：公安机关利用 AIGC 生成的诈骗话术库与场景模拟系统追踪新型犯罪趋势，金融机构依托 AIGC 风控模型识别多级转账与虚拟货币混币模式，运营商通过 AIGC 分析结果优化信令监测阈值[27]。这种协同体系突破传统治理的孤岛效应，实现从风险预警、资金拦截到犯罪追踪的全流程无缝衔接，有效应对电信诈骗的跨区域、跨平台特性。

6. AIGC 赋能电诈治理的实践路径

6.1 构建一致性价值目标

在电信网络诈骗治理中，AIGC 技术的应用不仅仅是对现有技术的补充，更是推动治理模式转型的重要动力。因此，首先需要确立 AIGC 赋能电信诈骗治理的核心价值理念，即通过技术创新构建高效、精准、全局的反诈骗防控体系。

这一理念的核心在于通过 AIGC 的智能生成与分析能力，打破现有反诈体系中数据割裂、效率低下的瓶颈。在传统反诈体系中，各平台、机构的数据通常各自为政，难以形成有效的协同，这就使得诈骗分子能够利用

信息不对称、监管漏洞开展诈骗行为。AIGC的引入，可以通过生成性模型和预测性分析，实现跨平台、跨领域的数据整合与情报共享，从而提高反诈骗体系的整体效率。AIGC的价值理念还应包括智能生成和自动化分析的双重驱动。通过AI生成技术，反诈骗体系能够实时应对新兴的诈骗手法，快速更新防控策略；通过AI分析技术，系统可以高效处理海量数据，实现从被动防御到主动预警的转变。这种技术驱动的价值理念，能够帮助电信诈骗治理体系应对复杂的诈骗环境，实现动态、智能的反诈防控。

6.2 应用AIGC技术赋能电信网络诈骗治理的具体路径

AIGC通过智能话术库构建、诈骗场景预测及多模态融合技术，形成立体化的电诈防控体系。其基于序列到序列（seq2seq）模型对历史诈骗通话记录、短信内容进行深度解析，构建动态特征库以捕捉“高回报”“安全验证”等高频关键词的上下文关联，并结合情感分析模块识别话术诱导性强度。当检测到新型话术模式时，系统触发迁移学习流程完成模型更新，解决传统规则库滞后性问题。其次，将生成对抗网络（GANs）与时空预测模型结合通过分析用户通信记录、地理位置、消费习惯等数据生成三维风险热力图，识别如“工作日晚间商业区周边小额高频转账”等风险组合，指导运营商加强特定区域信令监测，将被动响应转变为主动干预，跨模态特征对齐网络整合语音韵律、短信语义、图像视觉等多维度特征实现对诈骗电话语音波动、短信URL链接、钓鱼网站图像的立体识别，显著提升可疑交易识别准确率[28]。

6.3 形成AIGC赋能电诈治理格局的组织保障

AIGC技术在电诈治理中的规模化应用，依赖于技术、政策与人才的三维保障体系构建。在技术层面，需构建“云边端”协同的智能基础设施，通过部署高性能计算集群提升数据处理效率，搭建分布式算法训练平台支撑模型迭代更新，同时建立跨部门数据中台实现多源数据整合。政策法律层面，应制定AIGC技术应用的合规框架，明确数据采集、使用与共享的边界条件，建立诈骗信息生成的伦理审查机制，完善数据隐私保护与安全审计制度，确保技术应用符合《个人信息保护法》等相关法规要求。人才保障方面，需建立跨学科人才培养体系，通过校企联合实验室培养兼具AI算法能力与反诈业务知识的复合型人才，同时完善专家智库机制，吸纳网络安全、法学等领域专家参与技术决策，为AIGC赋能电诈治理提供可持续的智力支持。

7.总结与展望

本研究系统探讨了AIGC技术赋能电信网络诈骗治理的内在逻辑与实践路径。研究发现，AIGC通过其高效生成、智能内容和动态适应三大核心技术特征，重构了电诈治理的全流程：在数据治理层面，AIGC通过多模态数据融合与跨平台整合，破解了传统数据割据难题；在分析决策层面，其生成对抗网络与迁移学习技术实现了对新型诈骗手法的动态识别与预测；在应用场景层面，智能话术库、多模态劝阻系统及精准宣传策略显著提升了防控效能。通过构建“技术-制度-人才”三位一体的保障体系，AIGC推动电诈治理从被动响应转向主动防御，为数字时代新型犯罪治理提供了技术范式参考。

未来研究需聚焦三大方向：其一，深化AIGC伦理风险防控，建立生成内容安全审查机制与算法透明度框架，平衡技术创新与社会安全；其二，探索跨国协同治理模式，构建跨境数据共享协议与联合执法机制，应对电诈犯罪的全球化趋势；其三，推动AIGC与区块链、物联网等技术的深度融合，构建覆盖“人-机-物”三元空间的立体防控网络。此外，需关注AIGC技术迭代对犯罪形态的反向影响，建立“技术对抗-治理升级”的动态博弈模型，确保治理体系的持续适应性。研究还需强化实证分析，通过大规模应用场景验证AIGC的实际效能，为政策制定提供更具说服力的实践依据。

参考文献

[1]魏嘉迪，赵晓凡，陈丽，等.电信网络诈骗犯罪防治研究综述[J].中国人民公安大学学报（自然科学版），2024，30（02）：102-108.
[2]吕培霖，张文韬.AI电信网络诈骗犯罪治理研究[J].福建警察学院学报，2024，38（01）：23-31.
[3]陈永峰，王慧.新安全格局视角下电信网络诈骗犯罪治理困境与优化路径[J].武汉公安干部学院学报，2023，37（04）：50-55.
[4]李雪峰，王铼.电信网络诈骗的特征与治理路径[J].人民论坛，2023，（20）：65-67.

[5]农忠海，蒋萍，侯文雷.人工智能生成内容AIGC大模型在公安工作中的应用探讨[J].电脑知识与技术，2023，19（13）：29-31+38.
[6]陈芸蕾，商玉玺.电信网络诈骗治理的理念调试与路径抉择：一个技术的分析框架[J].江苏警官学院学报，2024，39（03）：107-114.
[7]靳雨婷.AIGC背景下新型网络诈骗手段与对策研究[J].网络安全技术与应用，2025，（02）：143-145.
[8]杨胜钦.从ChatGPT看AI对电信网络诈骗犯罪治理的影响[J].犯罪与改造研究，2024，（05）：26-33.
[9]孙晨博.从运动到常规：电信网络诈骗犯罪治理模式的选择与调适[J].江西警察学院报，2024，1-15.
[10]叶璐.电信网络诈骗犯罪的治理对策研究——以D市为例[J].现代商贸工业，2024，45（04）：183-185.
[11]毛飞飞.跨境电信网络诈骗犯罪运作流程及治理路径研究[J].网络安全技术与应用，2023，（07）：156-158.
[12]裘佳仪，叶腾.电信网络诈骗犯罪原因分析及治理对策研究——以日常活动理论和理性选择理论为视角[J].产业与科技论坛，2024，23（09）：26-28.
[13]单勇.数字平台与犯罪治理转型[J].社会学研究，2022，37（04）：45-68+227.
[14]Anyang D. The Characteristics and Prevention Countermeasures of Telecom Network Fraud Crime under the Background of Big Data. Journal of Sociology and Ethnology, 2023, 5(5): 77-81.
[15]Chu Y. A Communication Research on Reshaping the Content Ecosystem of New Media Platforms Based on AIGC Technology. Frontier sin Computing and Intelligent Systems, 2024, 9(3): 24-27.
[16]许桂敏，符迪豪.探究电信网络诈骗犯罪治理模式——以被害预防为路径[J].山东行政学院学报，2024，（06）：121-128.
[17]刘晓萌.电信诈骗年轻化趋势——受害者心理分析与防范策略[J].西部学刊，2024，（24）：64-67.
[18]邓宁江.打击治理电信网络诈骗犯罪的困境及路径[J].辽宁警察学院学报，2025，27（01）：22-29.
[19]金龙君，翟翌.论个人信息处理中最小必要原则的审查[J].北京理工大学学报（社会科学版），2023，25（04）：140-150.
[20]刘明亮.人工智能生成内容（AIGC）技术特征及应用场景分析[J].信息记录材料，2023，24（10）：234-236.
[21]杨敏然，张新兴，陶荣湘.现状与趋势：国内人工智能生成内容（AIGC）研究透视[J].图书馆理论与实践，2024，（02）：56-65.
[22]吴照美，李子硕.电信网络诈骗犯罪的治理研究——理论依据、基本思路及主要措施[J].武汉公安干部学院学报，2024，38（04）：7-12.
[23]吴丹.人工智能背景下电信网络诈骗犯罪样态及数字化治理[J].公安研究，2024，（10）：87-97.
[24]斯彬洲，孙海春，吴越.基于大语言模型和事件融合的电信诈骗事件风险分析[J/OL].数据分析与知识发现，2025，1-19.
[25]徐娟.人工智能时代电信诈骗防控策略探讨[J].数字通信世界，2024，（11）：153-155.
[26]张夏恒，马妍.生成式人工智能技术赋能新质生产力涌现：价值意蕴、运行机理与实践路径[J].电子政务，2024，（04）：17-25.
[27]豆云飞.高校电信网络诈骗多元协同治理路径探析——基于高校反诈常态化视角[J].河南司法警官职业学院学报，2024，22（01）：123-128.
[28]刘鹏里.电信网络诈骗犯罪打击治理效能研究[J].辽宁警察学院学报，2025，27（01）：16-21.