

TC260-PG-202611A

网络安全标准实践指南

——人工智能应用安全指引 广播电视和网络视听

(v1.0-202609)

全国网络安全标准化技术委员会秘书处

国家广播电视总局科技司

2026 年 09 月

本文档可从以下网址获得：

www.tc260.org.cn/



全国网络安全标准化技术委员会
National Technical Committee 260 on Cybersecurity of SAC



国家广播电视总局
National Radio and Television Administration

广电总局科技司



前 言

《网络安全标准实践指南》（以下简称《实践指南》）是全国网络安全标准化技术委员会（以下简称“网安标委”）秘书处组织制定和发布的标准相关技术文件，旨在围绕网络安全法律法规政策、标准、网络安全热点和事件等主题，宣传网络安全相关标准及知识，提供标准化实践指引。

本文件起草单位：国家广播电视总局广播电视科学研究院、中央广播电视总台、中国电子技术标准化研究院、中国科学院自动化研究所、北京邮电大学、北京爱奇艺科技有限公司、北京快手科技有限公司、中央网信办数据与技术保障中心、北京信工博特智能科技有限公司、中国传媒大学、海思技术有限公司、中电科网络安全科技股份有限公司等。

本文件主要起草人：王磊、郭晓霞、郑蕤荻、郭庭昊、张妍婷、吕飞霄、宋昊、王伟、刘杰、卢明樊、落红卫、王昕、郭建领、樊庆君、邵玮、司俊俊、申东洋、薛子育、刘庆同、陈维玮、梁志坚、刘媛媛、谢学兵、饶婷霜、童永祥、王超、顾翠。



声 明

本《实践指南》版权属于网安标委秘书处，未经秘书处书面授权，不得以任何方式抄袭、翻译《实践指南》的任何部分。凡转载或引用本《实践指南》的观点、数据，请注明“来源：全国网络安全标准化技术委员会秘书处”。



摘 要

为全面提升各行业人工智能应用安全水平，以高水平安全保障人工智能高质量发展，依据《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》，以及《生成式人工智能服务管理暂行办法》《人工智能生成合成内容标识办法》《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》等政策法规，制定人工智能应用安全指引系列文件。

人工智能应用安全指引系列文件包含通用文件以及行业领域文件两类。通用文件提供通用性实践安全指导，适用于各行业领域开展人工智能应用活动。行业领域文件给出适用于特定行业领域的实践指引。

本《实践指南》属于行业领域文件，用于指导广播电视和网络视听领域企事业单位开展人工智能应用活动。本文件与《网络安全标准实践指南——人工智能应用安全指引 总则》等通用文件配合使用。



目 录

1 范围	1
2 规范性引用文件.....	1
3 术语定义.....	2
4 缩略语.....	2
5 概述	3
6 基本安全指引.....	3
7 业务安全指引.....	3
8 应用场景安全指引.....	5
8.1 策划创作场景.....	5
8.2 制作播出场景.....	6
8.3 传输覆盖场景.....	8
8.4 集成分发场景.....	8
附录 A（资料性）广播电视和网络视听领域人工智能应用场景与主要风险	11
参考文献.....	15





1 范围

本文件规定了广播电视和网络视听领域人工智能应用的基本安全指引，业务安全指引，以及策划创作、制作播出、传输覆盖、集成分发等应用场景的安全指引。

本文件适用于广播电视和网络视听机构开展人工智能应用时的安全风险防范与管理。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB 45438 网络安全技术 人工智能生成合成内容标识方法

GB/T 25069 信息安全技术 术语

GB/T 41867 信息技术 人工智能 术语

GB/T 45652 网络安全技术 生成式人工智能预训练和优化训练数据安全规范

GB/T 45654 网络安全技术 生成式人工智能服务安全基本要求

GB/T 45674 网络安全技术 生成式人工智能数据标注安全规范

GY/T 411 数字虚拟人技术要求

TC260-PG-20268A 网络安全标准实践指南——人工智能应用安



全指引 总则

3 术语定义

TC260-PG-20268A《网络安全标准实践指南——人工智能应用安全指引 总则》附录 A 中界定的以及下列术语和定义适用于本文件。

3.1 广播电视和网络视听机构 Radio, TV and network audio-visual institution

经广播电视主管部门许可，从事广播电视和网络视听业务的机构。

3.2 广播电视和网络视听人工智能应用 Artificial intelligence application in radio, TV and network audiovisual

人工智能技术在广播电视和网络视听业务中的集成与应用。

3.3 数字虚拟人 digital human

基于现实世界设计，通过计算机生成，再借助真人或计算驱动，在多模态输出设备呈现的虚拟人物。

[来源：GY/T 411—2024，3.1]

4 缩略语

TC260-PG-20268A《网络安全标准实践指南——人工智能应用安全指引 总则》附录 B 中的缩略语，以及下列缩略语适用于本文件。

CDN：内容分发网络(Content Delivery Network)

IPTV：交互式网络电视(Internet Protocol Television)



MIDI: 乐器数字接口 (Musical Instrument Digital Interface)

5 概述

人工智能技术在广播电视和网络视听领域的应用涉及广播电视、IPTV、互联网电视、互联网视听服务四个层次的业务，主要包括策划创作、制作播出、传输覆盖、集成分发等场景。广播电视和网络视听人工智能应用场景见附录A。

广播电视和网络视听机构在开展人工智能应用时，应符合国家相关法律法规和政策文件的要求，满足行业发展需求，按照本文件第6章至第8章提出的安全指引开展研发集成和应用，防范人工智能应用安全风险，保障人工智能应用安全可控。

6 基本安全指引

广播电视和网络视听机构在开展人工智能应用活动过程中：

- a) 应参考 TC260-PG-20268A 《网络安全标准实践指南——人工智能应用安全指引 总则》，提升人工智能应用安全水平，防范人工智能应用安全风险。
- b) 使用生成式人工智能技术面向公众提供服务的，应符合 GB/T 45654、GB 45438 相关要求。
- c) 应具备人工确认、手动优先、快速切换和回退机制。

7 业务安全指引



广播电视和网络视听机构在使用人工智能技术开展业务过程中：

- a) 使用人工智能技术生成、制作、传播的广播电视和网络视听节目，应符合内容管理、播出管理的有关要求。
- b) 使用人工智能技术生成、制作、传播的广播电视和网络视听节目，不应含有下列内容：
 - 1) 违反宪法确定的基本原则，煽动抗拒或者破坏宪法、法律法规实施；
 - 2) 危害国家统一、主权和领土完整；
 - 3) 泄露国家秘密、危害国家安全或者损害国家尊严、荣誉和利益；
 - 4) 诋毁优秀民族传统文化，煽动民族仇恨、民族歧视，侵害民族风俗习惯，歪曲民族历史或者民族历史人物，伤害民族感情，破坏民族团结；
 - 5) 违背国家宗教政策，宣扬邪教、迷信；
 - 6) 危害社会公德，扰乱社会秩序，破坏社会稳定，宣扬淫秽、赌博、吸毒，渲染暴力、恐怖，教唆犯罪或者传授犯罪方法；
 - 7) 侵害未成年人合法权益或者损害未成年人身心健康；
 - 8) 侮辱、诽谤他人或者散布他人隐私，侵害他人合法权益；
 - 9) 法律、行政法规禁止的其他内容。



- c) 不应使用人工智能技术生成时政类新闻；使用人工智能技术生成、制作、传播非时政类的新闻时，应按照 GB 45438 的相关规定，在适当位置、区域显著提示。
- d) 应采取有效措施，对使用的数字虚拟人进行注册或者登记，加强数字虚拟人等数据安全。数字虚拟人的使用范围和应用方式，应按照广播电视和网络视听主管部门有关规定执行；使用数字虚拟人形象及语音合成的，应依法取得权利人授权，并在播出时向观众显著提示。不应使用数字虚拟人形象从事时政新闻类播音主持活动。
- e) 在未经授权的情况下，使用人工智能技术生成、制作音视频节目的，不应出现仿冒他人形象和声音等侵犯人格权的行为。
- f) 开展人工智能训练数据采集、标注和管理等活动时，应符合 GB/T 45652 和 GB/T 45674 的规定。

8 应用场景安全指引

8.1 策划创作场景

- a) 应落实播前审查要求，避免出现文化歧视倾向、历史虚无主义表述、社会公序良俗违背等价值观偏差。
- b) 应保障知识产权，人工智能使用的图片、视频、声音等素材及生成内容应符合知识产权法律法规要求，具有版权或得到相应



的授权许可。仅取得训练许可的素材，不应推定已经取得生成内容复制、改编、发布或者传播的许可。

- c) 部署基于关键词、语义识别、图像比对等内容审核工具，对人工智能生成内容实施全量初步筛查，重点识别违法违规、暴力色情及不符合广播电视和网络视听播出规范的内容。
- d) 建立涵盖权威公开信息来源作为参考的可信信源库；涉及关键事实、政策表述、统计数据和引述内容等，应进行核验，如两个可信信源交叉核验等。
- e) 设置初审、复审、终审等多层级人工审核机制，每级指定专人负责内容审查，及时发现并纠正潜在问题。
- f) 审查训练数据合法性与价值观合规性，并建立人工智能应用全生命周期日志，记录数据、模型、输入及交互，日志应防篡改、可追溯。
- g) 建立人工智能输入内容和输出内容的价值观、歧视内容、商业违法违规、他人合法权益等评估、监测和预警机制，及时发现和处置风险。
- h) 人工智能生成合成内容的文件元数据中应添加元数据隐式标识。

8.2 制作播出场景

- a) 建立全过程追溯机制，关键、敏感、自动生成内容应人工复审，节目成片由相关责任人终审，确保内容安全、责任可溯。



- b) 基于客观事实或历史知识的内容制播，应尊重历史和客观规律。
- c) 利用人工智能对历史影像资料进行修复时，应以“修旧如旧”为原则，以原始档案和可靠史料为基础，禁止缺乏史料支撑地推断性生成，确保符合史实与播出要求。
- d) 修复、上色与扩图应遵循最小干预原则，避免引入非实证信息或造成画面失真。
- e) 宜支持对外部采集素材进行人工智能生成合成内容判别和标识管理。
- f) 使用生成式人工智能技术制作播出的节目，可能导致公众混淆或者误认的，应在合理位置、区域显著提示；不会导致公众混淆或者误认的，可在片头或片尾注明含有人工智能辅助生成内容。
- g) 使用生成式人工智能技术生成、制作的微短剧，应按照《微短剧发展管理办法》的规定，在每集明显位置添加提示标识。
- h) 使用生成式人工智能技术制作播出的节目，其显式标识和元数据隐式标识在制作和播出全流程不得删除，并在媒资管理、节目传播过程中携带的同时，支持对相关信息进行完善。
- i) 网络视听机构应在用户生成内容上线前判别其是否为人工智能生成合成；判别为生成合成的，应按照 GB 45438 的规定添加显式标识和元数据隐式标识。



- j) 宜在生成合成内容中添加数字水印等形式的隐式标识,但不应影响主观观看体验。
- k) 宜通过完整性校验、传播链路检测等技术措施,提高标识在转码、剪辑、跨平台传播后的可保留性、可识别性和可追溯性。

8.3 传输覆盖场景

- a) 建立故障动态入库与模型更新机制,定期纳入新型故障知识,确保模型覆盖最新故障类型。
- b) 仅允许将验证后的解决方案录入知识库。
- c) 不应使用未经审核的外部数据源。
- d) 人工智能生成的策略应附带“知识库溯源标签”,通过标签可获取引用的知识库原文摘要。
- e) 针对故障的人工智能生成策略仅可作参考建议,不直接执行。

8.4 集成分发场景

- a) 算法透明度提升:分发服务提供者采用合理的方式和途径向用户公开算法机制,可设置“算法说明中心”并以图文或视频解释推荐规则。
- b) 信息茧房治理:
 - 1) 在用户兴趣偏好外引入多样化推荐策略或对同类题材设置推送上限;
 - 2) 提供一键切换至获取非偏好内容的功能;



3) 用户画像应适当、合理，不得包含歧视、恶意等信息。

c) 推荐内容审核:

1) 持续完善内容审核机制，内容分发前应经过自动化审核和人工复核;

2) 建立违规和不良信息特征库，并定期更新;

3) 建立内容审核责任追溯机制，明确审核责任;

4) 对分发内容进行实时监测，发现违规或不良信息应立即停止推送并采取消除、屏蔽等处置措施。

d) 特殊群体保护:

1) 根据国家法律法规，提供未成年人模式;

2) 宜提供敬老功能，如切换大字体。

e) 客服交易管控：不自动修改账户、订单或支付状态；涉及资金或合同终止的请求立即转接至人工客服处理队列。

f) 客服知识库管理：仅可访问公开知识库，回复应绑定标准化知识库条目，用户可查看对应知识原文。

g) 客服回复边界：未匹配知识库的提问仅可返回预设内容，知识库匹配置信度低于阈值时，立即终止回复并推送转人工弹窗。

h) 客服会话隔离：启用新会话时清空历史缓存，不继承历史对话记录。



- i) 敏感信息脱敏：身份证号与银行卡号应仅保留必要位数、其余以“*”替代，手机号脱敏为[前三位]****[末四码]，订单号仅显示末四位，地址信息仅保留市级行政区划。
- j) 推荐信息保护：基于个人信息开展内容推荐的，应加强个人敏感信息保护，维护用户权益。
- k) 终端信息保护：采集用户声音、图像时，应按照最小必要原则，对必须采集的个人敏感信息采取加密、访问控制等措施，对个人信息的采集、传输、存储、使用进行全流程保护。





附录 A

(资料性)

广播电视和网络视听领域人工智能应用场景与主要风险

A.1 应用场景

广播电视和网络视听人工智能应用场景见表A.1。

表A.1 广播电视和网络视听人工智能应用场景

序号	一级场景	二级场景	场景说明
1	策划创作	剧本/脚本创作	创意构思、情节设计、初稿撰写、对白优化。
2		剧本/脚本分镜	依据剧本文本生成含镜号、景别、运镜的结构化分镜表及画面参考。
3		歌词创作	依据主题、情感或风格生成符合韵律与结构的歌词。
4		辅助新闻生产	辅助信息采集、稿件撰写、多语言翻译、多媒体整合。
5		演播室场景设计	辅助生成演播室背景与舞台视觉。
6		人物形象设计	基于文本描述生成写实或风格化人物外观、表情及姿态。
7		服化道设计	依据角色设定与场景需求生成服装、妆容、道具。
8		音乐编曲	根据旋律、歌词或风格描述生成音频片段、MIDI 序列或乐谱。
9		关键帧生成	根据剧本或分镜描述生成动画关键帧或故事板图像。
10		节目片段生成	根据文本指令或场景描述生成短时长节目内容片段。



序号	一级场景	二级场景	场景说明
11	制作播出	新闻剪辑	自动识别素材中的关键画面与同期声，依据预设规则制作新闻成片。
12		影视剧剪辑	依据剧本及演员情绪、动作完成粗剪，并按预设艺术风格精剪。
13		视频编辑	自动选择剪切点、添加转场，实现多轨道音视频同步。
14		字幕生成	利用自动语音识别生成直播字幕，或为录播节目匹配时间轴并翻译至多语种。
15		智能配音	基于自然语言处理与语音合成技术，将文本快速准确地转化为高质量人声配音。
16		台词翻译	将影视剧、新闻等作品中的原始台词自动转换为目标语言，保持语义准确、语境一致并支持口型同步。
17		视频修复	去除老旧影像的划痕、噪点与抖动，提升画质。
18		视频增强	将低分辨率、低帧率的视频素材提升至4K/8K超高清画质标准。
19		特效制作	自动完成场景扩展、虚拟元素合成、动力学模拟及风格化渲染，实现影视级视觉特效的快速生成与实时预览。
20	传输覆盖	知识问答	基于专业知识库，自动响应系统架构、设备配置、链路规划、频率指配、故障排查等咨询，并提供可追溯依据。
21		运行维护策略建议	依据实时运行数据、历史维护记录及网络拓扑，自动生成预防性维护计划、故障预警报告与优化调整策略。
22	集成分发	内容审核	依据行业内容管理、播出管理相关法规，结合行业负面样本数据，采用文本、图像、音视频特征识别技术，对拟



序号	一级场景	二级场景	场景说明
			分发内容实施全量筛查，发现并处置违规内容。
23		宣发素材生成	依据节目元数据、受众画像及平台规范，生成并定制海报、片花、预告片、短视频、文案及互动话题等多平台宣发素材。
24		内容推荐	依据用户行为数据、内容标签及上下文信息，生成面向不同终端用户的个性化内容推荐列表，并提供推荐理由及可解释依据。
25		广告投放	依据受众特征、投放目标及实时效果反馈，生成广告创意、投放计划、受众包及动态广告素材，实现精准投放与效果预测。
26		客户服务	根据用户咨询内容、知识库及业务规则，生成智能客服应答、故障排查指引及个性化业务办理方案。
27		系统运维	依据实时运行数据、历史故障记录及系统拓扑，生成运行状态报告、故障预警、性能优化策略及运维脚本；对关键指标支持实时告警与处置建议。
28		画质增强	终端实时或离线生成高分辨率、低噪声、色彩还原准确、动态范围扩展的视频。
29		音效增强	终端实时生成清晰度增强、声场优化及三维空间音效渲染的音频。
30		人机交互	终端实时响应语音、手势、姿态等多模态输入，完成频道切换、音量调节、内容搜索、播放控制及个性化设置。



序号	一级场景	二级场景	场景说明
31		智能问答	依据用户自然语言提问，实时提供节目信息、演员资料、剧情简介或相关背景知识，并支持多轮对话及上下文理解。

A.2 主要风险

人工智能技术在策划创作场景的应用可能存在事实不准确、价值导向偏差或表述不规范等风险，如杜撰事件、歪曲历史、偏离社会主义核心价值观、使用不准确的政治和历史表述、对民族宗教等敏感议题呈现不当立场，或输出缺乏依据的医学、法律等专业领域信息。

人工智能技术在制作播出场景的应用可能存在视频剪辑过程不可控、历史影像修复增强画面失真、字幕或翻译内容偏离原意、随意制作违法违规有害信息等风险，如画面识别不准确、衔接不连贯、叙事逻辑不一致、误用敏感内容、输出内容失真、面部相似度不足、纹理失真、时代元素错配、引入原始资料未包含的信息、损失原始影像质感等。

人工智能技术在传输覆盖场景的应用可能存在新型故障识别失效、接口与集成安全缺陷、人机协同策略可信度不足等风险。

人工智能技术在集成分发场景的应用可能存在算法可解释性不足、信息茧房与内容偏见、违规内容传播、用户隐私泄露、特定人群权益受损等风险，如推荐逻辑不被用户理解导致信任度下降，恶意生成内容传播造成不良社会影响，以及影响未成年人价值观和老年用户个人权益受损等。



参考文献

- [1] 中华人民共和国网络安全法
- [2] 中华人民共和国数据安全法
- [3] 中华人民共和国个人信息保护法
- [4] 中华人民共和国广告法
- [5] 中华人民共和国著作权法
- [6] 生成式人工智能服务管理暂行办法
- [7] 人工智能生成合成内容标识办法
- [8] 互联网信息服务算法推荐管理规定
- [9] 互联网信息服务深度合成管理规定
- [10] 微短剧发展管理办法

