

附件 2：联盟标准制修订立项申请书

世界超高清视频产业联盟标准制、修订立项申请书

| | | | | |
|---|--|---|--------------------------|---|
| 项目名称 | 超高清公共显示系统 内容播控安全要求 | 制定 <input checked="" type="checkbox"/> 修订 <input type="checkbox"/> | 被修订 标准号 | 无 |
| 采用国际/内标准名称（中文） | | 无 | | |
| 申请单位 | 1. 深圳市洲明科技股份有限公司 2. 北京百度网讯科技有限公司 3. 深圳蓝普科技有限公司 4. 中国电子技术标准化研究院 5. 北京匠数科技有限公司 6. 深圳创维-RGB 电子有限公司 7. 中联超清（北京）科技有限公司 8. 西安诺瓦星云科技股份有限公司 | 联系人 | 谭胜淋 | |
| 手机 | 0755-29918999 | Email | tanshenglin@unilumin.com | |
| 牵头单位 | 深圳市洲明科技股份有限公司 | | | |
| 计划起止时间 | 2026 年 3 月至 2027 年 2 月 | | | |
| <p>随着各地信息化建设的逐步深入，公共显示系统的应用越来越广泛。由于目前缺乏对公共显示系统有效的安全管控措施，屡屡出现公共显示系统播放不良、敏感、恶性内容的安全事故，造成恶劣的社会影响，损害公民、法人和组织的权益，破坏社会秩序和公共利益，甚至危害国家安全。</p> <p>习近平总书记在全国网络安全和信息化工作会议上指出：“没有网络安全就没有国家安全”，网络安全已经上升到了国家战略安全的高度。特别是在当下，境内外敌对势力造谣煽动，诋毁攻击我国政治体制，严重损害党和国家形象，侵蚀党的群众基础，党中央更是明确“要加强互联网内容建设，建立网络综合治理体系，营造清朗的网络空间”。</p> <p>公安部于 2019 年 5 月下发了《2019 年公安机关网络安全执法检查工作方案》，政府其他机关部门相应转发了网络安全工作要求。在网络安全四大检查项目中，明确指出将公共场所显示屏，即铁路、民航、社会治安、车站、医院、商场等重点区域、重点部门的公共显示屏系统作为重点检查对象。国家广播电视总局、教育部等各地方各部门也纷纷出台关于加强公共场所电子显示屏内容安全管理的通知。</p> <p>2020 年 3 月 1 日起施行的《网络信息内容生态治理规定》指出，各级网信部门应当会同有关主管部门，建立健全的工作机制，协同开展网络信息内容生态治理工作，定期对本行政区内网络信息内容治理情况进行评估。2020 年 5 月 22 日起，国家网信办在全国范围内启动为期 8 个月的 2020 “清朗”专项行动，</p> | | | | |



此次专项行动全面覆盖各类网络信息传播渠道，集中清理各类违法和不良信息。

2025年1月13日，国务院印发《公共安全视频图像信息系统管理条例》（中华人民共和国国务院令 第799号），规范公共安全视频系统管理，维护公共安全，明确公共安全视频系统的建设要求，公共安全视频系统管理单位的运行安全职责及视频图像信息使用要求，电信业务经营者对视频图像信息传输的安全管理义务，以及设计、施工、检验、验收、维护等单位对视频图像信息的保密义务。

根据《中华人民共和国网络安全法》《中华人民共和国刑法》相关规定，网络服务提供者不履行法律、行政法规规定的信息网络安全管理义务，有可能触犯“拒不履行信息网络安全管理义务罪”，对单位判处有期徒刑，并对其直接负责的主管人员和其他直接责任人员，依照规定进行处罚。

目前，显示屏内容及播控安全普遍缺乏有效的技术防护手段，通常采用“人工管理的事前审核 & 事后处理”手段，无法实现内容播放过程中的实时审核及控制，因此存在很大的安全隐患。

1、安全管理孤岛化，被动式管控思想。传统管理手段通常是采用显示屏配套的控制系統，系統除了控制显示屏，不作他用。该手段虽可降低被外部不法分子入侵的风险，但却无法有效防止内部管理人员误操作或蓄意篡改播放内容等不良行为。

2、显示屏存在低安全高风险。随着智能设备的普及和快速发展，越来越多的显示屏能通过无线投屏的方式呈现内容。据调查，为了安装维护方便，可使用无线投屏的播控系统往往未设置太复杂的接入密码，甚至未对默认密码进行修改，无形中降低了犯罪成本。显示屏容易被不法分子作为媒介入侵到服务器、数据库等内部重要的资产，危及服务系统的数据安全，导致各类敏感信息泄露。

3、管控技术手段严重滞后。缺少对显示屏播控系统操作痕迹的记录手段，一旦出现显示屏安全事故，无法记录异常事件相关的操作日志，无法查询到违规操作和被入侵的痕迹，导致安全事件无法溯源上报。

综上所述，目前超高清公共显示系统存在低安全性、安全管理孤岛化、被动式管控思想、管控技术手段严重滞后的现状。一旦超高清公共显示系统的播控终端或播控电脑受到黑客攻击、人为误操作或蓄意更换篡改播放内容，无法做到内容播放过程中的实时审核及控制，可能会造成极其不良的社会影响和后果。

综上，制定《超高清公共显示系统内容播控安全要求》标准，规范超高清公共显示系统内容安全防护和管控势在必行。

本标准规定了超高清公共显示系统的内容播控安全系统的系统组成和播控流程、播控设备要求、安全管理要求。

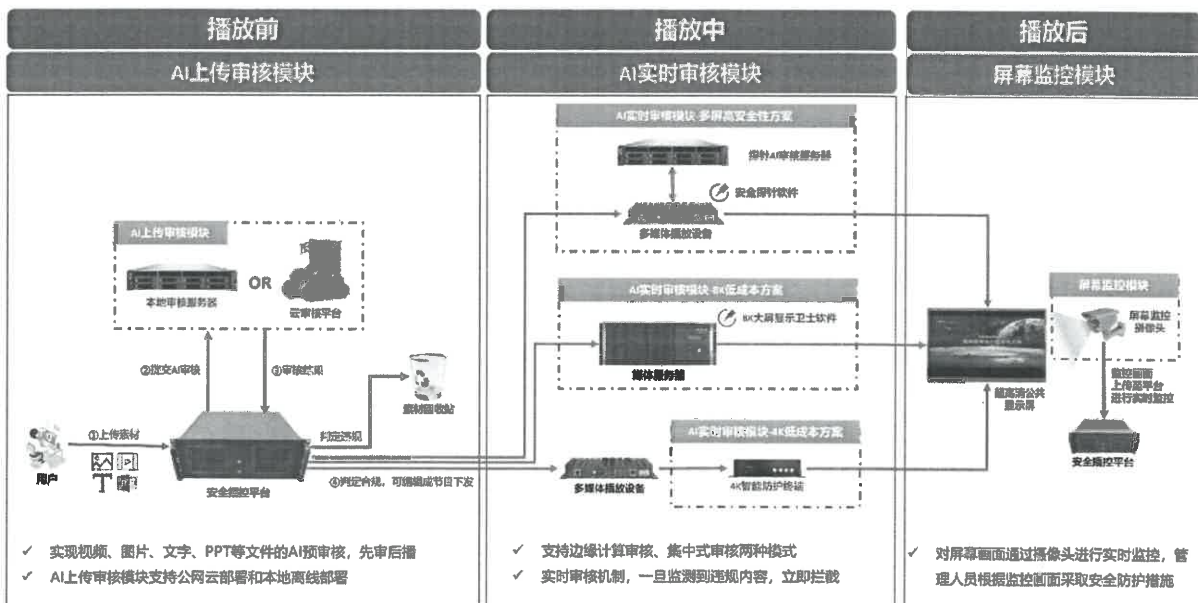
本标准适用于超高清公共显示系统的内容播控安全系统的规划、建设及运行、播控设备的监督、管理。其它电子屏播放安全可参考。

本标准框架总体如下：

- 1 范围
 - 2 规范性引用文件
 - 3 术语和定义
 - 4 缩略语
 - 5 系统组成和播控流程
 - 5.1 系统组成
 - 5.2 播控安全流程
 - 6 播控设备要求
 - 6.1 播控设备功能要求
 - 6.2 播控设备性能要求
- 附录 A（资料性）应用场景
- 参考文献

本标准提出的技术内容总体如下：

一、系统组成



超高清公共显示系统内容播控安全系统由播控平台、上传审核服务器、实时审核探针、实时审核终端、屏幕监控设备构成，各单元各司其职、协同联动，实现播控全流程安全管控。系统组成见上图，各设备功能定义如下：

1. 播控平台：作为系统核心管控端，实现播放内容的信息发布，同时对全系统安全设备进行统一运维、配置及管理。
2. 上传审核服务器：接收播控平台管理员上传的图片、视频、文本、PPT 等待审核素材，调用内置内容审核模型开展敏感内容检测，完成检测后向播控平台反馈最终检测结果。
3. 实时审核终端：实时读取播控服务器输出的视频流数据并进行画面帧采样，通过内置内容审核模型检测画面帧敏感内容；检测正常则输出原视频流至显示终端，检测异常则将原视频流替换为预设安全画面后输出。
4. 实时审核探针：实时读取播控服务器输出的视频流数据并进行画面帧采样，通过网络将画面帧传输至探针审核服务器；经探针审核服务器内置模型检测后接收反馈结果，检测正常输出原视频流，检测异常则替换为预设安全画面输出至显示终端。
5. 屏幕监控设备：实时采集超高清公共显示系统的播放画面，将监控画面传输至播控平台，实现多终端画面的平台集中查看与监控。

二、系统功能

1、上传审核

- 1) 播控平台集成 AI 审核功能，支持对视频、图片、文字、PPT 等各类素材文件进行自动化敏感违规内容检测。
- 2) 可对图片、文本、视频、文档等媒体文件开展多维度检测过滤，精准识别色情、涉政、暴恐、违禁等不良画面及信息特征。
- 3) 支持系统级敏感库自定义配置，可添加自定义敏感词、敏感人脸信息，实现对自定义敏感人脸、敏感人物的精准检测。

2、实时审核

- 1) 具备 7×24 小时全天候无人值守的实时审核机制，可实时获取显示终端当前播放画面，自动检测画面中的敏感违规内容。
- 2) 支持播放视频全维度敏感内容检测，覆盖涉黄、涉暴恐、涉政、违禁等图像类敏感内容，及涉政、涉黄、涉暴恐、违禁和用户自定义的文字类敏感内容。
- 3) 具备异常内容快速处置能力，检测到敏感内容后立即触发安全防护措施，将播放画面替换为预设安全画面；待检测到正常画面后，可自动恢复原视频流内容输出。

3、屏幕监控

- 1) 支持通过摄像头等设备对多个显示终端的播放画面进行同步实时监控，实现多终端画面的全域采集。
- 2) 播控平台集成屏幕监控模块，可集中展示区域内所有显示终端的实时播放状态，提升异常事件的感知与发现效率。
- 3) 支持自定义创建多套异常处置预案，预案可配置切换安全节目、屏幕断电、切断画面信号、向管理人员推送告警通知等防护措施；发现终端异常画面时，管理人员可通过播控平台一键调用预案，缩短异常处置响应时间。

三、系统流程：

超高清公共显示系统内容播控安全系统遵循“播放前上传审核 – 播放中实时检测 – 播放后监控屏幕”的全链路管控流程，具体流程如下：

1. 素材上传审核：在播控平台 Web 端，上传人员通过基于 HTTP/HTTPS 协议的文件上传接口将本地播放文件上传至播控平台，平台通过素材审核接口将文件及元数据提交至内容审核服务器。审核服务器完成 AI 审核后，通过 AI 审核结果推送接口向播控平台推送“合规 / 违规”结果及违规详情。审核通过的素材归档至合规素材库，审核不通过的素材被系统拦截并反馈违规原因。

2. 信息发布：在播控平台 Web 端，管理人员选定合规播放文件及播控电脑并确定下发后，平台通过播放文件下发接口将合规播放文件下发至播控电脑，完成播放前的内容分发。

3. 媒体播放：播控电脑通过播放文件解析接口解析播放文件，提取音视频流并输出画面数据。

4. 视频流实时检测：播放服务器输出视频流后，实时审核终端通过 HDMI 或 DP 高清视频接口同步读取视频流并进行画面帧采样，通过实时审核终端内置的模型开展实时敏感内容检测；检测正常则输出原视频流，检测异常则立即替换为预设安全画面输出。

5. 屏幕监控核验：屏幕监控设备实时采集各显示终端的实际播放画面，通过 IP 网络将画面传输至播控平台监控模块进行集中展示，管理人员可实时核验终端播放状态。

6. 异常处置：实时审核环节检测到异常，系统自动触发安全防护；监控环节发现异常，管理人员可通过播控平台一键调用预设处置预案，采取对应防护措施；异常解除后，系统自动恢复原内容正常播放。

7. 全流程记录：系统对素材上传、审核、播控、检测、监控、处置等全环节操作及数据进行日志记录，实现全流程可追溯、可核查。

四、系统技术指标：

本章节针对系统三个核心应用场景，明确各场景核心技术指标及测试方法，参考边缘端实时审核场景

的详细规格，兼顾专业性、可落地性，贴合各场景实际应用需求，确保指标定义清晰、测试方法可操作。

(1) 播放前：上传审核（本地 AI 审核服务器）

本场景核心为本地 AI 审核服务器对上传图片进行合规审核，聚焦审核精度、处理效率，确保上传内容合规性，为后续播控流程奠定安全基础。

| 指标类别 | 指标名称 | 指标定义 | 参考值 | 说明 | 测试方法 |
|--------|------------|---|---------------------|--|--|
| 审核精度指标 | 上传图片审核准确率 | 本地 AI 服务器对上传图片审核时，正确识别合规/违规内容的数量占总上传图片数量的比例 | $\geq 95\%$ | 涵盖各类违规视觉元素、敏感内容，可识别篡改、拼接的违规图片，保障上传内容合规 | 1. 选取 10000 张含合规、违规（含篡改）的上传图片作为测试集；2. 导入本地 AI 服务器审核，统计正确识别数量；3. 按公式计算准确率，重复测试 3 次取平均值。 |
| | 审核误判率 | 正常合规的上传图片，被误判为违规或疑似违规的数量占总合规图片数量的比例 | $\leq 5\%$ | 降低合规内容误判概率，避免正常内容无法上传，保障播控流程顺畅 | 1. 选取 10000 张合规上传图片作为测试集；2. 导入服务器审核，统计误判数量；3. 按公式计算误判率，重复测试 3 次取平均值。 |
| | 文字提取及过滤准确率 | 对上传图片中的文字提取后，正确过滤违规文字、敏感表述的比例 | $\geq 95\%$ | 可识别同义词替换、谐音变体等规避审核行为，确保图片中隐藏违规文字被检出 | 1. 选取 500 张含违规文字、合规文字的上传图片；2. 统计正确提取并过滤违规文字的数量；3. 计算准确率，重复测试 3 次取平均值。 |
| 处理效率指标 | 单张图片审核耗时 | 单张图片从导入服务器到输出审核结果的总耗时 | $\leq 100\text{ms}$ | 适配批量上传场景，避免审核耗时过长影响上传效率，贴合实际操作需求 | 1. 选取 100 张不同规格（含 4K 超高清）上传图片；2. 记录单张审核耗时，统计平均值、最大值；3. 确保平均值满足参考 |

| | | | | | |
|--|----------|--------------------|-----------|---------------------------------|---|
| | | | | | 值要求。 |
| | 批量上传审核效率 | 服务器每分钟可完成审核的上传图片数量 | ≥600 张/分钟 | 满足批量上传审核需求，提升工作效率，适配多终端、多内容上传场景 | 1. 准备 6000 张上传图片，批量导入服务器；2. 记录完成全部审核的总耗时；3. 计算每分钟审核数量，重复测试 2 次取平均值。 |

(2) 播放中：实时审核

本场景核心为边缘盒本地实时审核播控内容，聚焦实时性、审核精度及边缘端性能，实现违规内容快速阻断，适配超高清播控实时需求。

| 指标类别 | 指标名称 | 指标定义 | 参考值 | 说明 | 测试方法 |
|---------|------------|------------------------------------|--------------------|----------------------------------|---|
| 审核精度指标 | 实时审核准确率 | 边缘盒对实时播控的超高清内容（图像/视频帧），正确识别违规内容的比例 | ≥95% | 适配超高清内容，可快速识别各类违规视觉元素，契合实时审核场景需求 | 1. 导入 10000 帧 4K 超高清合规、违规视频帧；2. 边缘盒实时审核，统计正确识别数量；3. 计算准确率，重复测试 3 次取平均值。 |
| 响应时间 | 违规内容阻断响应时间 | 检测到违规内容后，边缘盒触发阻断机制至终端停止播发的耗时 | ≤100ms | 快速阻断违规内容，避免违规信息在公共终端扩散 | 1. 输入违规超高清视频；2. 记录“检测到违规”至“终端停播”的耗时；3. 重复测试 50 次，取平均值。 |
| 边缘端性能指标 | 边缘盒算力要求 | 边缘盒支持 AI 推理任务的计算能力，以 TOPS 为单位 | ≥6 TOPS(基于 RK3588) | NPU，满足实时审核多模块协同算力需求 | 1. 用测试工具，在边缘盒运行 AI 推理任务；2. 记录 NPU INT8 算力，重复测试 3 次取平均值。 |

(3) 播放后：屏幕监视景（摄像头+本地 AI 审核服务器）

本场景核心为摄像头拍摄公共显示终端画面，回传至本地 AI 审核服务器，核对是否为已审核通过内容，聚焦识别精度、实时性及环境适应性，实现播后闭环监管。

| 指标类别 | 指标名称 | 指标定义 | 参考值 | 说明 | 测试方法 |
|---------|---------------|---------------------------------|----------------------|---|---|
| 回传响应指标 | 画面回传时间 | 摄像头拍摄画面回传到服务器所需的时间 | ≤10 秒 | <ol style="list-style-type: none"> 1. 摄像头拍摄标准测试画面，记录画面采集完成时间 T_1； 2. 记录服务器完整接收并可解析该画面的时间 T_2； 3. 计算回传时间 $T=T_2-T_1$，连续测试 10 次取平均值 | 测试需在实际部署网络环境下进行，排除网络拥堵等异常干扰 |
| 核对精度指标 | 画面核对准确率 | 摄像头拍摄画面回传服务器后，正确核对为已审核通过内容的比例 | ≥95% | 可适配画面轻微亮度、角度变化，准确区分已审核/未审核内容 | <ol style="list-style-type: none"> 1. 预设 10 组已审核通过画面，摄像头拍摄各 500 帧； 2. 回传服务器核对，统计正确核对数量； 3. 计算准确率，重复测试 3 次取平均值。 |
| 环境适应性指标 | 亮度适应性 | 不同环境亮度下，服务器对拍摄画面的正确核对率 | 50lux-1000lux 下 ≥80% | 适配公共场景白天、傍晚等不同亮度环境，不影响核对精度 | <ol style="list-style-type: none"> 1. 调节环境亮度至 50lux、500lux、1000lux； 2. 各拍摄 500 帧画面，回传核对； 3. 统计各亮度下的正确核对率。 |
| 应急响应指标 | AI 检测应急处置响应时间 | 服务器 AI 检测到异常画面后，触发并执行应急处置预案的总耗时 | ≤3 秒 | <ol style="list-style-type: none"> 1. 预设 10 组典型异常画面，模拟 AI 检测触发场景； 2. 记录 AI 输出异常判定结果的时间 T_1； 3. 记录应急处置动作（告警、联动等）执行完成的时间 T_2； 4. 计算响应时间 $T=T_2-T_1$，重复测试 10 次取 | 含 AI 推理、指令下发、设备执行全链路耗时，需在实际业务环境下验证 |

| | | | | | |
|---|--|--|--|-----|--|
| | | | | 平均值 | |
| <p>与《超高清公共显示系统内容播控安全要求》相关的标准如下：</p> <p>GB/T 22239-2019 信息安全技术 网络安全等级保护基本要求、GB/T 22240-2020 信息安全技术 网络安全等级保护定级指南：两个标准主要涵盖了信息安全通用性要求及分类规则、系统维护要求等内容，本标准沿用上述两个标准对信息安全产品的技术指标要求。</p> <p>T/UWA 037—2025 基于人工智能的多媒体信息发布系统技术要求：描述了基于人工智能的多媒体信息发布系统的业务流程，主要关注数据安全（数据完整性、数据保密性、数据时效性）、访问安全及网络安全，并不涉及播放内容的安全性，与本标准相关性较弱。</p> <p>DB31/T 1661-2025 公共区域电子屏播控安全管理要求：该标准要求建立多媒体信息审核机制、具备多媒体信息的延迟播放功能、可留存一定时间内的已播放多媒体文件。上述要求主要聚焦事中防护，并不涉及事前审核和事后监控，且并未对设备技术指标提出要求，本标准作为上述标准的补充。</p> <p>国外并未将公共显示屏内容安全作为重点关注事项，因此暂无显示屏播控安全的相关国外标准。</p> | | | | | |
| <p>申请立项单位意见</p> <p>建议制定联盟标准《超高清公共显示系统内容播控安全要求》。</p> | | | | | |
| <p>备注：</p> | | | | | |



