

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	AI 超高清媒体封装与传输协议	制定 <input checked="" type="checkbox"/> 修订 <input type="checkbox"/>	被修订标准号	
采用国际/内标准名称 (中文)	AI 超高清媒体封装与传输协议			
申请单位	1. 中国移动通信集团有限公司 2. 咪咕文化科技有限公司 3. 上海交通大学 4. 鹏城实验室 5. 中兴通讯股份有限公司 6. 上海网仕科技有限公司 7. 马栏山音视频实验室	联系人	陈丽丽、毕蕾	
手机	陈丽丽: 13524661159 毕蕾: 18066077313	Email	陈丽丽: chenlili@migu.chinamobile.com 毕蕾: bilei@migu.chinamobile.com	
牵头单位	咪咕文化科技有限公司			
计划起止时间	2026 年 6 月-2027 年 6 月			
<p>立项的目的、意义或必要性</p> <p>【标准背景】</p> <p>超高清视频产业正经历从“高清化”向“智能化”和“沉浸化”的深刻变革。以 4K/8K 为代表的超高清视频已进入规模化应用阶段，而人工智能技术的爆发式发展，正在重塑视频内容的生成、传输、理解和呈现方式。与此同时，语义编码、3DGS/4DGS/7DGS、世界模型等技术的发展对媒体封装和传输也提出了全新要求。</p> <p>然而，当前主流媒体封装和传输协议（如 MP4、FLV、MKV、HLS、DASH 等）主要围绕传统音视频媒体设计，缺乏对 AI 超高清媒体封装与传输的支持。</p> <p>【行业需求】</p> <p>随着 AI 与超高清视频的深度融合，行业已经涌现出大量新兴应用场景，对 AI 超高清媒体封装与传输提出了明确且迫切的需求：</p> <p>1、多维信息的融合承载需求</p> <p>在沉浸式视频、AIGC 音视频加嵌、AI 导播、AI 推理数据应用、面向 ROI 的 AI 智能编码等典型场景中，除了传统的视频和音频轨道，系统还需同步传输 AIGC 生成音视频，AI 数据包括图像识别结果、语音识别结果、视频推理结果、ROI 区域掩码、以及 AIGC 提示词等，以及传感器数据包括相机参数、采集时间等多维信息。</p> <p>目前 AI 数据和传感器数据通常通过 SEI 嵌入的方式实现，或独立传输的方式实现。由于 SEI 是编码中可选字段，存在数据大小受限、易丢失、编辑困难等问题，无法满足 AI 数据的灵活、可扩展的传输需求。而独立传输方式下缺乏非时序数据和音视频时序数据的关联机制定义。</p> <p>行业迫切需要一种能够将传统音视频与 AI 信息、传感器信息等一体化封装的统一容器格式，从而降低端到端系统的复杂度，避免数据在传输链路中的割裂与丢失。</p>				

2、多维信息的数据协同需求

在多维度信息的融合承载场景中，音视频、AIGC 音视频、AI 数据、传感器数据等来源各异、格式不同，但必须在业务逻辑上实现紧密耦合。目前多链路独立传输的机制，系统复杂、同步误差大，AI 信息和音视频的呈现、标注、依赖、映射等协同关系缺乏定义。行业迫切需要一种对“多源异构、强协同”类数据之间的关联机制的原生支持，能够在一体化封装的框架内，满足多维度信息的数据之间的精细化协同需求。

3. 多维信息的数据传输需求

在多维度信息的融合承载场景中，AI 数据（如目标检测坐标、预估的速度、关键事件、雷达图、人物模型参数、ROI mask、AIGC 提示词等）必须与特定视频帧实现毫秒级的精准对齐。同时，来自采集端的相机参数、深度信息、时间信息等非时序元数据，也需要在采、编、播全链路中保持完整性并与音视频内容同步。行业急需支持内置精确的时序模型与非时序信息精准同步的封装传输协议，确保非时序元数据与媒体内容在采集、传输、处理、呈现各环节始终保持高度同步，同时支持不同类型的数据对应不同的 QoS 映射以及底层网络映射。

4、面向未来演进的灵活扩展需求

随着 3DGS/4DGS/7DGS 与沉浸式交互、视频语义编码与 AIGC 世界模型、大规模海量数据与超大码率等新一代技术的即将规模化应用，以及云边端协同、双流/多流协同传输、空天地一体融合通信（5G/6G、卫星、终端直连等）的演进，协议必须具备良好的前瞻性扩展能力。行业期待一个既能兼容现有主流传输协议和编码格式，又能通过预留扩展机制平滑支持未来新兴媒体类型和新兴融合通信传输的标准化方案，避免因标准频繁更迭带来的产业升级成本。

【目的及意义】

AI 超高清媒体封装与传输协议定位为 AI 超高清媒体的应用层协议，兼容主流编码格式和传输协议，支撑多样化应用场景，构建开放、高效、可演进的标准化体系。

本协议支持在应用层标准化 AIGC 音视频和 AI 数据如识别、推理、理解、AIGC 提示词等结构化数据、以及传感器数据的封装与传输，满足 AI+平面视频，自由视角等场景需求，可基于 HTTP、RTP 等协议实现传输，并确保协议未来可演进，为构建下一代媒体网络传输奠定标准基础。

后续将通过协议扩展，支持 8K 视频分片、球幕视频、视频语义编码、AI 信息压缩编码，3DGS/4DGS/7DGS 传输、双流/多流传输等场景需求。

目前尚无专门针对 AI 超高清媒体封装与传输的标准，本立项有望填补这一空白，解决产业核心痛点，引领产业发展。本协议自主定义，具备自主演进能力，能有效支撑 AI 超高清产业发展，为赛事、广电、安防、车载、VR/AR 等场景提供标准化的 AI 超高清媒体封装传输方案，加速 AI 超高清媒体的规模化应用。

适用范围或主要技术内容

【适用范围】本文件适用于 AI 超高清媒体的应用场景。

【主要技术内容】本文件规定了 AI 超高清媒体包括视频、音频、AI 信息、传感器数据的封装与传输协议。涵盖数据模型、数据封装、数据传输等内容。

主要技术内容包括：

数据模型：定义 AI 超高清媒体（AUM）中媒体内容的组织结构和关联关系。在 ISO BMFF 标准框架的基础上进行扩展，通过新增轨道类型、样本入口类型和关联机制，实现对传统媒体、AI 信息及传感器数据的统一封装与表达。通过扩展机制确保协议未来可演进。

数据封装：定义 AI 超高清媒体（AUM）的封装格式。在 ISO BMFF 规范基础上进行扩展，通

过'aumf'标签标识AUM文件,通过'aume'盒子提供资产标识、序列号和能力声明,AUM支持时序媒体、非时序媒体以及多种媒体类型的混合封装。可扩展设计预留演进能力。

数据传输:定义基于HTTP和RTP协议的传输适配机制,解包和打包规则等,通过AUM提示轨道提供RTP和HTTP传输的打包指引。可扩展设计预留未来演进能力。

国内外情况简要说明

国际相关标准与技术现状

在国际标准领域,目前尚无专门针对AI超高清媒体的封装与传输的标准。

主流媒体封装格式(如MP4、FLV、MKV)主要围绕传统音视频媒体设计,缺乏对AI超高清媒体的支持。ISO/IEC制定的DASH标准是国际主流的自适应流媒体传输协议,定义了媒体呈现描述和分段传输机制。但DASH主要关注传统音视频的分片、码率自适应和传输调度,其封装格式依赖ISO BMFF(MP4),未包含AI信息,无法实现AI信息与视频帧的紧密耦合。HLS分片机制缺乏AI数据封装能力,难以满足AI信息与视频帧的紧密耦合需求。

综上所述,国际上在AI超高清媒体封装传输领域尚存标准空白,亟需填补。

国内相关标准与技术现状

在国内,SMT作为我国自主定义的智能媒体传输协议,支持异构网络的传输。但标准仍以传统音视频为核心对象,未明确AI信息等的封装与传输。

国家广播电视总局发布了GY/T 420-2025《超高清音视频编码码流网络传输封装技术要求》行业标准,规定了AVS2、AVS3、Audio Vivid等编码格式在传输流(TS)、智能媒体传输(SMT)等协议中的封装技术要求。

综上,国内已有较完善的超高清编码和封装标准基础,但尚无专门针对AI超高清媒体封装与传输的标准,本立项有望填补这一空白。

申请立项单位意见



备注: