**3GPP TSG-SA WG4 Meeting #132 S4-250883**

**Fukuoka, JP, 19 – 23 May 2025**

**Source: China Mobile Com. Corporation**

**Title: [FS\_Beyond2D] Introduction of AI Generated Beyond 2D Content**

**Agenda item: 9.7**

**Document for: Agreement**

**1. Introduction**

The commercialization of AIGC has garnered significant interest from both academia and industry, it also shows promise in overcoming the challenges associated with Beyond 2D content production. At the last SA4#131-bis-e meeting, there was a proposal to address AI-generated Beyond 2D content in a separate new clause within the TR document.

**2. Proposal**

It is proposed to agree the following changes to the 3GPP draft TR 26.956 V0.4.0.

\* \* \* First Change (All New) \* \* \*

### 7.2.4 AI-Generated Beyond 2D Content

Creating and capturing high-quality Beyond2D content is often a labor-intensive task that demands substantial time, expertise, or specialized capturing tools/devices, which limits the widespread adoption of Beyond 2D media. Artificial Intelligence Generated Content (AIGC) leverages AI technologies to autonomously produce content. For example, in clause 7.2.2.2, AI-powered 2D to stereoscopic 3D video methodology was introduced, which effectively reduces the reliance on high-end capture devices. Beyond this, AIGC encompasses a range of emerging technologies, including: Image-to-dynamic Mesh Generation, text-to-dynamic Mesh Generation and 4D Scene generation, which will be introduced in the following sections. The commercialization of AIGC has attracted considerable attention from both academia and industry, driving innovation in Beyond 2D content creation, compression technologies, and quality assessment methodologies.

Figure 7.2.4-1 illustrates a reference workflow for AI-generated beyond 2D content. The workflow positions a large language model (LLM) at the core of logical reasoning, transforming different inputs, such as text, image, video, 3D models, actuator signals and etc into a unified tensor representation. After reasoning and inference by the LLM, the output tensor is mapped back to the target modality.



Figure 7.2.4-1 Workflow for AI-generated beyond 2D content

**- Representation:** The model should effectively represent and process different media types, such as text, images, video, and 3D models. Appropriate representation format should be selected for each type (e.g., CNNs for image features) to enable downstream processing and analysis.

**- Alignment:**Alignment refers to the process of matching and correlating data across different media types, enabling the model to comprehend their interrelationships. For instance, attention mechanisms can be employed to establish semantic correspondences between text and images.

**- Inference:** The model should be capable of inference capabilities, it can analyze and understand input data to extract useful information. A common approach is to leverage pre-trained large language modes to perform inference tasks.

**- Generation:**The model should be capable of generating new content, for instance, creating 3D mesh from text prompts.

**- Evaluation:** Assessing model performance (include both subjective methodologies and objective metric) is critical to ensure output relevance and reliability.

\* \* \* End of Changes \* \* \*