Group Discussion(2nd) on the impacts of AI technologies on media & entertainment IG

Song Xu, China Mobile

AI Impact on MEIG: Landscape & Core Use Cases

AI Content Labeling Ecosystem

- **C2PA** – Content Credentials (cryptographic provenance) for images/video/audio
- **IPTC 2025.1** – New metadata fields for AI-generated content (prompt, model, Digital Source Type)
- **W3C AI Content Disclosure CG** – Structured syntax for AI-disclosure in HTML/text

MEIG Focus: Beyond Text

- Audio/Video, WebVR/XR, Volumetric Video – where **WebCodecs, WebGPU, WebNN, WebTransport** face new challenges

Three Core Use Cases (UCs)

1. **UC1** – Next-gen codecs (AVS-NG, MPEG DCVC) driven by neural networks
2. **UC2** – Immersive & enhanced experiences (volumetric video, super-res, frame interpolation, AI denoising, face masking)
3. **UC3** – Multimodal AI agents evolving from text to audio-visual on the Web

UC1 & UC2: Technical Demands on Web APIs

UC1 – Neural Codecs (AVS-NG, MPEG DCVC)

- WebCodecs currently supports only H.264/AV1/VP9 → **no native NN-codec acceleration**
- WebGPU → parallel execution of CNN ops (feature extraction, context propagation)
- WebNN → full operator-set mapping for DCVC, utilising NPU for efficient inference

UC2 – Volumetric Video & Real-time Enhancements

- **Data deluge**: 200k points/frame, 30 fps → 720 Mbps raw bandwidth (10–100× traditional video)
- WebGPU → massive parallel processing for point-cloud decompression (AVS PCC), rendering, and memory management
- WebNN → accelerates segmentation, neural rendering (2D→3D), and real-time AI enhancement (super-res, denoising, etc.)
- WebTransport → low-latency streaming for immersive, interactive experiences

UC3 & Summary: Key Challenges Ahead

UC3 – Multimodal AI Agents

- Agents (ChatGPT, Claude, Doubao, Qwen) moving from text-only to audio-visual reasoning
- Need **simultaneous** inference for vision (CNN/ViT), language (Transformer), and audio
- WebNN + WebGPU via **MLTensor** (buffer sharing) enables cross-model memory reuse
- Real-time interaction requires **WebTransport** low-latency pipelines + WebCodecs for encoding/decoding

Cross-Cutting Challenges

- **Extend WebCodecs** to support NN-based codecs
- **Optimise WebNN/WebGPU synergy** for zero-copy data flow
- **Integrate AI labelling** (C2PA/IPTC) automatically into media processing pipelines

Key Takeaway

AI transforms Web media from **codec to rendering to agent interaction** – WebGPU/WebNN/WebTransport are the foundational enablers, and MEIG must guide their evolution for immersive, trustworthy, and real-time AI media.

- AI related media use cases #123
https://github.com/w3c/media-and-entertainment/issues/123

- Impact of AI technologies on Media and Entertainment Interest Group's mission #3
https://github.com/w3c/webai-roadmap/issues/3