Une image contenant texte

Description générée automatiquement ISO/IEC JTC 1/SC 29/WG 2 N447

**ISO/IEC JTC 1/SC 29/WG 2  
MPEG Technical requirements   
Convenorship: SFS (Finland)**

**Document type:** Output Document

**Title:** Strategy on MPEG-AI

**Status:** Approved

**Date of document:** 2025-04-04

**Source:** ISO/IEC JTC 1/SC 29/WG 2

# Expected action: None

# Action due date: None

**No. of pages:** 5 (with cover page)

**Email of Convenor:** igor.curcio@nokia.com

**Committee URL:** <https://sd.iso.org/documents/ui/#!/browse/iso/iso-iec-jtc-1/iso-iec-jtc-1-sc-29/iso-iec-jtc-1-sc-29-wg-2>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 2**

**MPEG TECHNICAL REQUIREMENTS**

**ISO/IEC JTC 1/SC 29/WG 2 N447**

**Online – April 2025**

|  |  |
| --- | --- |
| **Title** | **Strategy on MPEG-AI** |
| **Source** | **WG 2, MPEG Technical requirements** |
| **Status** | **Approved** |
| **Serial Number** | **25070** |

The MPEG-AI initiative envisions a future where Artificial Intelligence (AI) and Neural Networks (NN) become the cornerstone of multimedia coding, processing, and standardization, revolutionizing how we interact with digital content. As multimedia services evolve and demand unprecedented levels of interactivity, adaptability, and embedded intelligence, AI's role in transforming content creation, encoding, and distribution has become indispensable. This integration is advancing within ISO/IEC JTC1/SC29, where MPEG is crafting a visionary framework—MPEG-AI—that brings together a family of standards leveraging AI’s potential to shape a new generation of multimedia experiences.

In this strategic mission, MPEG-AI positions AI technologies at the heart of multimedia evolution through two transformative pathways. First, AI serves as a powerful tool for multimedia coding, enabling next-generation representations that exploit neural networks’ strengths. This includes generating intelligent content descriptors and achieving groundbreaking efficiencies in video, audio, and graphics compression, an innovation poised to elevate streaming, broadcast, and interactive experiences. Second, MPEG-AI champions AI as an engine for content analysis, where multimedia data is optimized for intelligent systems, facilitating seamless processing, interpretation, and decision-making by AI models. This paradigm shift ensures multimedia is encoded not only for human consumption but also for machine-driven applications, laying a foundation for AI-augmented ecosystems such as autonomous systems and adaptive metaverse environments.

MPEG-AI standards are meticulously crafted to embrace AI at multiple levels, whether fully integrated into the processing chain or strategically employed within hybrid architectures. This flexibility allows for a synergy of AI and traditional methodologies, maximizing performance while balancing computational demands. Key technological domains under the MPEG-AI umbrella include innovative multimedia representation, where AI-driven encoding techniques enable efficient content compression, feature extraction, and adaptive descriptors across video, audio, graphics. Likewise, AI-based analysis and processing techniques are being standardized to enhance content reconstruction, optimize distribution networks, and enable intelligent interaction with multimedia data. Supporting technologies, including neural network model optimization and metadata schema for AI processes, further enable scalable and efficient deployment.

The current MPEG-AI strategy builds on a solid foundation established by three pioneering standards: ISO/IEC 15938-13:2015, ISO/IEC 15938-15:2019, and ISO/IEC 15938-17:2024. These standards were instrumental in setting the stage for the intelligent multimedia solutions envisioned within MPEG-AI, representing the initial milestones in MPEG’s journey toward integrating AI into multimedia representation and analysis.

This journey began with ISO/IEC 15938-13:2015, entitled “*Compact descriptors for visual search”*, a part of the MPEG-7 family, which introduced compact descriptors for visual search (CDVS). This standard created a robust framework for image description and visual content matching, addressing challenges such as partial occlusions and viewpoint changes. Its impact on efficient, interoperable visual search applications marked a key step in the convergence of visual data and intelligent retrieval systems, which later influenced MPEG-AI’s goals.

The next leap came with ISO/IEC 15938-15:2019, called “*Compact descriptors for video analysis”*, which extended these capabilities to video content. With optimized descriptor sizes and processing efficiencies, this standard supported large-scale video search and retrieval, proving invaluable for managing complex video databases such as broadcast archives. It addressed the unique demands of dynamic scenes, enabling reliable matching across varied conditions and laying a critical foundation for future multimedia standards in intelligent video analysis.

ISO/IEC 15938-17:2024, entitled “*Compression of neural networks for multimedia content description and analysis”*, introduced Neural Network Coding (NNC), establishing a powerful toolset for compressing and distributing neural network parameters in a way that complements existing network topologies. This framework for parameter compression enables adaptive, scalable deployment of neural models, facilitating intelligent processing across a broad spectrum of devices and use cases. NNC is particularly significant for MPEG-AI, as it optimizes AI model deployment, ensuring efficiency in AI-powered multimedia applications.

These previous standards collectively formed the backbone of the MPEG-AI initiative. By building on these early advances, MPEG-AI is grouping now a new series of standards that fully harness the potential of AI-driven multimedia analysis, coding, and content retrieval. This evolution from initial descriptors to advanced AI integrations underscores MPEG-AI’s vision: to create a multimedia ecosystem that seamlessly combines AI with multimedia, delivering data-driven insights, real-time analysis, and transformative multimedia experiences across applications and industries.

MPEG is currently developing three critical standards and preparing to initiate two more, all of which collectively advance our mission of integrating Artificial Intelligence (AI) and Neural Networks (NN) into multimedia coding and processing, setting the stage for a new era in intelligent multimedia applications.

The foundational *Part 1*, ISO/IEC TR 23888-1, titled “*Vision and scenarios”*, serves as the cornerstone of this initiative. This standard lays out a strategic perspective on the role of AI and NN technologies in multimedia coding. It outlines a forward-looking vision on the potential of AI for multimedia, identifying key technical challenges and assumptions that will guide future work. This foundational standard articulates the essential scenarios where AI-driven multimedia can have transformative effects, highlighting the specific gaps that subsequent standards will address to fulfill the vision of AI-enhanced multimedia systems.

Building upon this groundwork, *Part 2*, ISO/IEC 23888-2, “*Video coding for machines*”, addresses the distinct requirements of machine-centric applications. Unlike traditional video coding, which prioritizes human viewing experiences within bit-rate constraints, Video Coding for Machines (VCM) caters to the needs of AI systems that process vast amounts of data in real time. This standard reflects a paradigm shift in video coding to support intelligent platforms such as connected vehicles, video surveillance, and smart cities, where the data scale and latency requirements surpass those suited for human consumption. VCM anticipates scenarios in which machines communicate directly to execute complex tasks autonomously. However, this standard also recognizes cases, such as in surveillance, where a human operator may intermittently need to analyze specific decompressed video streams, ensuring that VCM supports both human and machine-centric interactions when necessary.

In *Part 3*, “*Optimization of encoders and receiving systems for machine analysis of coded video content”*, MPEG-AI focuses on refining video encoding and decoding processes to better serve machine analysis applications. Traditional video compression standards like HEVC and VVC were designed primarily for human viewers, resulting in encoding techniques optimized for the human visual system. However, as AI and machine learning systems have advanced, it has become clear that the visual cues vital for human perception differ from those required for effective machine analysis. This standard seeks to bridge this gap by proposing methods to optimize encoders and receiving systems for machine analysis, ultimately improving machine-based interpretation without altering the foundational video compression standards. The technical report provides detailed examples of technologies and methodologies that can enhance machine analysis performance, paving the way for more effective AI-driven video processing across applications.

Two additional explorations complete the current scope of the MPEG-AI initiative.

The potential *Part 4*, “*Feature coding for machines*”, aims to create a framework for efficiently compressing intermediate feature data used in machine vision tasks, such as object detection, instance segmentation, and tracking. Unlike traditional video coding, which is optimized for human perception, this standard is tailored for machine-based analysis, supporting collaborative architectures that distribute processing between edge devices and cloud systems. By compressing feature data instead of raw video content, the potential Part 4 enhances data efficiency, computational offloading, and privacy protection for high-data environments like intelligent transportation and smart cities.

The potential *Part 5*, “*AI-based point cloud coding*”, focuses on efficient compression of 3D spatial data to support applications in immersive media and autonomous systems. This standard introduces a versatile coding framework that supports both dense point clouds for immersive experiences and sparse point clouds from LiDAR sensors. Using a hybrid approach that combines learning-based geometry coding with traditional attribute coding, the potential Part 5 technologies achieve substantial bitrate savings while ensuring high visual quality. This approach aligns with MPEG-AI’s vision of adaptable, high-quality 3D content compression, enabling scalable applications from immersive environments to intelligent transportation.

|  |  |  |  |
| --- | --- | --- | --- |
| **Project Number** | **Title** | **Status** | **Online** |
| ISO/IEC 15938-13:2015 | Compact descriptors for visual search | Published | www.iso.org/standard/65393.html |
| ISO/IEC 15938-15:2019 | Compact descriptors for video analysis | Published | www.iso.org/standard/75399.html |
| ISO/IEC 15938-17:2024 | Compression of neural networks for multimedia content description and analysis | Published | www.iso.org/standard/85545.html |
| ISO/IEC 23888-1 | Vision and scenarios | In progress | www.iso.org/standard/87589.html |
| ISO/IEC 23888-2 | Video Coding for Machines | In progress | www.iso.org/standard/88879.html |
| ISO/IEC 23888-3 | Optimization of encoders and receiving systems for machine analysis of coded video content | In progress | www.iso.org/standard/89045.html |
| Potential Part 4 of ISO/IEC 23888 | Feature Coding for Machines | In progress |  |
| Potential Part 5 of ISO/IEC 23888 | AI-Based Point Cloud Coding | In progress |  |
| ISO/IEC 23093-6 | Internet of Media Things | In progress |  |

**Staying aligned with the MPEG Mission**

MPEG’s mission is rooted in a commitment to defining standards that enable the efficient representation, processing, and delivery of digital media. This mission extends beyond the traditional realms of images, audio, and video to encompass emerging forms of digital information and metadata, ensuring that multimedia systems remain adaptable to the ever-evolving technological landscape.

In this context, the integration of AI technologies into MPEG’s framework represents a natural evolution of its core objectives. AI has the potential to revolutionize the way media is created, processed, and experienced by introducing unprecedented levels of intelligence and adaptability. By embedding AI into the heart of multimedia systems, MPEG envisions a future where media coding is not only efficient but also responsive to both human and machine needs, bridging the gap between content intended for human consumption and data optimized for intelligent systems.

This vision sees AI as a transformative force that enhances every stage of the multimedia lifecycle. From intelligent compression techniques that preserve quality while minimizing data requirements to advanced metadata systems that enable seamless interaction between users and machines, AI enriches the ecosystem by making it smarter, faster, and more capable of addressing complex challenges. In this future, media becomes more than just a means of communication; it evolves into an adaptable and context-aware entity, capable of shaping and responding to the environments in which it is deployed.

MPEG also recognizes the importance of safeguarding the integrity and reliability of multimedia systems as they incorporate AI. With a focus on synchronization, security, and privacy, the mission ensures that these advanced systems remain trustworthy and aligned with user expectations. Quality of Experience and system performance continue to be central to this vision, ensuring that the integration of AI technologies enhances user satisfaction while maintaining the efficiency and interoperability that define MPEG’s standards.

By embracing AI, MPEG positions itself as a leader at the intersection of media and intelligence, crafting a multimedia future where creativity, technology, and innovation converge. This vision invites stakeholders across industries to join in shaping this transformative journey, building a new era of digital ecosystems defined by intelligence, adaptability, and immersive possibilities.