ACM Multimedia Topics of Interest

1. Introduction and History

For many years before 2024, ACM Multimedia was structured in four themes, namely: Engaging Users with Multimedia, Multimedia Systems, Multimedia Content Understanding, and Experience. In 2024 and 2025, the theme Multimedia in the Generative AI Era was added. Each theme was divided into three sub-themes as follows:

- Multimedia in the Generative AI Era
  - Multimedia Foundation Models
  - Generative Multimedia
  - Social Aspects of Generative AI
- Engaging Users with Multimedia
  - Emotional and Social Signals
  - Multimedia Search and Recommendation
  - Summarization, Analytics, and Storytelling
- Experience
  - Interactions and Quality of Experience
  - Art and Culture
  - Multimedia Applications
- Multimedia Systems

    - Systems and Middleware
    - Transport and Delivery
    - Data Systems Management and Indexing
- Multimedia Content Understanding
    - Multimodal Fusion
    - Vision and Language
    - Multimedia Interpretation

Since the beginning of 2026, the hierarchy between themes and sub-themes has been removed, and only a list of topics of interest and their description remains, as shown in the next section.

2. Topics List and Description

The list of topics is shown here in alphabetical order.

Art and Culture

Papers under this topic area should develop techniques that enable effective engagement of the public with art and other forms of cultural expression, balancing between sophisticated computational and engineering techniques and artistic and cultural purposes. Topics include (i) digital artworks, including hybrid physical digital installations; dynamic, generative, and interactive multimedia artworks; (ii) computational tools to support creativity, cultural preservation, and curation.

Data Systems Management and Indexing

Papers under this topic area should address performance issues related to data management and indexing to support multimedia access at a large scale, including browsing, searching, recommendation, analysis, processing, and mining. Topics include scalable systems and indexing techniques that support multimedia access and analytics.

Embodied and Immersive Multimedia

This topic explores multimedia models and methods that bridge digital media with real-world embodiment, enabling richer, more adaptive, more context-aware multimedia experiences. Specifically it focuses on technologies designed either for intelligent agents that perceive, act, and interact in the real world or for humans immersed in simulated/augmented/extended environments. This theme emphasizes embodied models that integrate vision, language, audio, haptics, and other signals, virtual humans, and multimodal agents capable of grounded reasoning and purposeful action immersed in a wide range of environments.

Emotional and Social Signals

This area focuses on the analysis of emotional, cognitive and interactive social behavior in the spectrum of individual to small group settings. It calls for novel contributions with a strong human-centered focus specializing in supporting or developing automated techniques for analyzing, processing, interpreting, synthesizing, or exploiting human social, affective and cognitive signals.

Interactions and Quality of Experience

Papers on this topic should address human-centered issues. Topics include (i) novel interaction techniques and modalities for accessing, authoring, and consuming multimedia data, (ii) design and

implementation of novel interactive media, (iii) new methodologies, models, and metrics to understand and/or measure multimedia quality of experience.

Multimedia Applications

Papers under this topic area should push the envelope of how multimedia can be used to improve the user experience in a rich and meaningful manner for a specific application domain. We solicit papers that design, implement, and evaluate applications that employ multimedia data in surprising new ways or in application scenarios that user experience remains challenging based on today's state-of-the-art, such as immersive telepresence, distance education, and metaverse. Applications can include entertainment and information in diverse areas such as education, healthcare, wellbeing, governance, etc.

Multimodal Fusion

In the real world, some problems are addressable only through a combination of multiple media and/or modalities. This area seeks new insights and solutions of how multi-perspective media information should be fused and embedded for novel problems as well as innovative systems.

Multimedia Generative and Foundation Models

This topic unifies advances in multimedia foundation models and modern generative modeling. It covers large-scale multimodal foundation models that integrate several modalities; new architectures and alignment strategies for cross-modal processing; and fundamental insights into training, scaling, and adapting such models. It also includes generative approaches that enable the creation of realistic and diverse multimedia content. Emphasis is placed on interactive, controllable, and personalized generation systems that leverage the combined power of foundation models and generative techniques.

Multimedia Interpretability and Explainability

This area seeks novel processing of media-related information in any form that can lead to new ways of interpreting, explaining, or creating multimedia content. Examples include processing of visual, audio, music, language, speech, or other modalities for interpretation, understanding, and generation. A particular emphasis is placed on approaches that provide transparent, interpretable, and explainable insights into how multimedia systems derive their conclusions. The topic encourages research that advances both the interpretability and the explainability of multimedia models, enabling more trustworthy, accountable, and user-centered multimedia technologies.

Multimedia and Language

This topic investigates the integration of natural language with other modalities such as images, audio, video, and sensor data. It encompasses methods for multimodal processing enabling the understanding, description, and interaction with the world using both linguistic and perceptual information. Areas of interest include vision-language and audio-language models, multimodal dialogue systems, captioning, multimodal question answering.

Multimedia Search and Recommendation

To engage users in information access, search, and recommendation requires not only understanding of data but also user and context. This area calls for novel solutions for user-centric multimedia search

and recommendations, in either automatic or interactive mode, with topics ranging from optimization, user intent prediction, to personalized, collaborative or exploratory algorithms.

Multimedia Reasoning

This topic focuses on models and algorithms that can perform logical, causal, temporal, or spatial reasoning over multimodal content such as images, videos, audio, and text. This includes tasks that require understanding relationships between entities across modalities; drawing inferences that are not explicitly stated; performing step-by-step reasoning over multimodal evidence; integrating symbolic reasoning with neural multimedia representations; answering complex queries about multimedia collections; and enabling robust, explainable reasoning.

New Forms of Media Content

Papers under this topic area should propose, discuss and analyze multimedia experiences harnessing new forms of media content (e.g. VR, AR, MR, XR, 360°, live-streaming, haptics, olfactory, gustation, etc.); those that are consumed in diverse ways including multiple devices, platforms and modalities; and those that arise from AI generation.

Responsible Multimedia

The success of multimedia models and technology requires deep thinking into its societal and ethical impact. This topic calls for research works in promoting privacy, security, fairness, transparency, interpretability, and explainability of multimedia models, technology, and applications.

Summarization, Analytics, and Storytelling

The information underlying multimedia is by nature multi-perspective. Allowing efficient multi-perspective and context-adaptive information access remains an open problem. This area calls for new and novel solutions that can compose, link, edit and summarize multimedia data into a compact but insightful, enjoyable and multi-perspective presentation to facilitate tasks such as multimedia analytics, decision making, searching and browsing.

Transport and Delivery

Papers under this topic area should address improvements to multimedia transport and delivery mechanisms over a network, focusing on performance, reliability, and QoE at scale. Topics include protocols and architectural solutions that optimize the flow of large-volume, time-sensitive multimedia data, and systems for intelligent content distribution and in-network content placement.

3. Modifications and Versions

The first version of this document dates from December 2025. However, most of the topics above are based on years of discussion and efforts to adapt the topics to the new state-of-the-art and new trends. Several people contributed to these evolutions. The conference organisers can propose new topics they want to put forward or any other changes to the Steering Committee **before** advertising them in the web page and related calls for contributions.