

2026 START Program CFP Brief

THEME: **03. Multimedia**

SUB-THEME: **3.1. Intelligent Multi-Modal Systems for Next-Generation Display**

Context/ Overview

AI advancements in vision, language, audio, ambient sensing, and multimodal reasoning are transforming modern displays into intelligent systems capable of far more than visual rendering. Next-generation displays—TVs, monitors, XR/AR headsets, mobile screens, automotive displays, and public signage—are becoming interactive, perceptive, and adaptive. They not only optimize picture quality but also enable natural on-device intelligence, including gesture, voice, spatial awareness, device-to-user interaction, and contextual UI adaptation.

Problem Statement

As displays evolve toward adaptive and interactive systems, several challenges remain:

- 1. Multimodal complexity**
Future displays must interpret visual content, audio cues, ambient conditions, gestures, voices, and user context.
- 2. Dynamic real-world environments**
Lighting, movement, room geometry, user distance, and interactions shift constantly.
- 3. On device interaction constraints:**
Gesture recognition, voice understanding, UI adaptation, and proximity sensing must work with low latency on tight hardware budgets.
- 4. Hardware limitations**
Display SoCs face strict performance and memory constraints, especially for 4K/8K/HDR content and always-on sensing.
- 5. Lack of holistic evaluation**
Perceptual quality and interaction quality require new, integrated ways to measure user experience.

This CFP seeks solutions that expand both **adaptive visual intelligence** and **natural, on device interaction intelligence** on display platforms.

Objectives & Scope

Proposals should aim to:

1. Advance **multimodal display intelligence**, incorporating vision, audio, ambient sensing, gestures, voice, and contextual cues.
2. Support **next generation interaction models** (gesture control, voice UI adaptation, proximity awareness, spatial interaction).
3. Develop **adaptive, context-aware display applications** across PQ, ambient response, personalized UI, accessibility, and immersive modes.
4. Innovate in **efficient on device AI** for interactive sensing and high-resolution content pipelines.
5. Enable **feedback-driven or user aware adaptation loops** for continuous improvement.
6. Provide strong **technical feasibility, evaluation plans, and deployment pathways**.

Scope & Focus Areas(*)

This CFP welcomes foundational research, applied research, prototype systems, data pipelines, and deployable solutions. Submissions should fall under one or both of the following tracks:

Track 1: Multimodal Reasoning and Intelligent Display Applications

This track focuses on advanced reasoning and adaptive capabilities that enhance visual quality, UI behavior, and user-device interaction (but not limited to):

1. **Multimodal Scene Understanding & Context Awareness**
 - Vision + audio + ambient sensor fusion for display adaptation
 - Content genre and intent understanding (movies, sports, games, apps)
 - User presence, distance, gaze region, or interaction state detection
2. **Adaptive Picture Quality & Rendering Intelligence**
 - Scene-aware dynamic tone mapping, HDR, contrast, and color adaptation
 - Content aware detail enhancement, denoising, motion handling
 - Display environment alignment (ambient light, color temperature)
3. **Personalized & Feedback-Driven Display Intelligence**
 - Adaptive UI and Accessibility features based on user behavior
 - Preference learning for PQ, sound modes, UI layouts
 - Multi-user profile recognition
 - Closed-loop personalization based on usage patterns
4. **Intelligent Media & Experience Pipelines**
 - Smart upscaling, interpolation, frame synthesis
 - AI-enhanced rendering for XR or immersive displays
 - Mood-aware or context-aware visual presentation

Track 2: Resource Constrained AI for Embedded and Embodied Devices

This track focuses on enabling real-time multimodal reasoning and interaction intelligence on display SoCs with strict compute, memory, and latency requirement. Key topic areas include (but not limited to):

1. Efficient AI Models for Displays & Interaction

- Lightweight architectures, including ultra-compact visual encoders
- Adaptive Inference Mechanisms including dynamic layer skipping, token pruning, conditional computation, and early exit strategies
- Compression strategies that preserve reasoning capability under aggressive pruning, quantization & distillation
- Real-time inference on limited compute
- Display-side personalization and preference updates

2. On Device Multimodal Processing

- Edge vision, audio, and sensor fusion
- Low-power situational reasoning and planning
- Lightweight continual learning respecting compute limits

※ The topics are not limited to the above examples and the participants are encouraged to propose other original ideas.

Requirements of the Proposal

Proposals should address the following five elements:

1. Detailed Technical Description

- Technical contribution and novelty
- Model/system architecture (PQ, interaction, or multimodal pipeline)
- Modalities used (vision, audio, ambient sensors, gestures, etc.)
- Relationship to existing display intelligence techniques

2. Technical Feasibility

- Compute, memory, or power assumptions
- Expected performance and scalability
- Early experimental evidence (if available)
- Risks and mitigation strategies

3. Validation Plan

- Datasets, benchmarks, or test environments
- Metrics for accuracy, quality, latency, robustness, etc.
- Plans for user studies (if applicable)
- Real-world or simulated deployment scenarios

4. Implementation Plan

- Milestones and stages
- Required resources and dependencies
- Integration approach with existing systems
- Proposed timeline and deliverables

5. Expected Impact

- Target domains (displays, robotics, agents, consumer devices, etc.)
 - Benefits to users, developers, or industry
 - Scalability across device classes or use cases
-

END OF DOCUMENT