2025 START Program
CFP Brief

THEME: **03. Artificial Intelligence**

SUB-THEME: **3.1. Multimodal Conditioned Video Generation**

Context/ Overview

Text-to-video generation systems (such as OpenAI's Sora, Google's Veo) have piqued research community as well as general public's interests. Such systems can pave the way for enabling rich creative expression and have applicability in personalized content creation, education, healthcare and synthetic data generation. However, generating contextually rich, semantically grounded and visually coherent videos is extremely challenging. Furthermore, the generated videos are short and require inordinate amount of compute. Solving these challenges will enable a number of use-cases for grounded physical intelligence beyond supporting the creative community.

Problem Statement

Existing video generation models often rely on text-only conditioning, limiting their capabilities in terms of integrating other sources of conditioning such as an existing video, 3D model of an actor, multiple videos of a set etc. The lack of multimodal conditioning limits controls over visual details, motion trajectories and narrative/storyline to visual alignment. Addressing these challenges will require new large-scale datasets, novel model architectures, training and evaluation methodologies.

Objectives & Scope

The goal of this call for proposals is to foster research in multimodal conditioned video generation which address integration of multiple modalities for video synthesis with particular emphasis on temporal coherence and semantic consistency. We are also interested in curating large-scale datasets and evaluating various aspects of video generation and editing.

Specific Topics & focus areas*

1. Vision, language, storyline, person conditioned video generation and editing.
2. Large-scale training and evaluation benchmarks
3. Novel ML architectures, tokenizers and training techniques for video generation especially in low-resource setting
4. Temporal consistency for video generation

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

END OF DOCUMENT