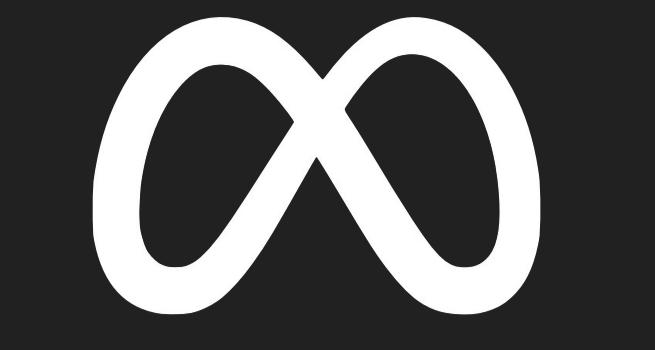


# LAVE: LLM-Powered Agent Assistance and Language Augmentation for Video Editing



UNIVERSITY OF  
TORONTO



Bryan Wang, Yuliang Li, Zhaoyang Lv, Haijun Xia, Yan Xu, Raj Sodhi

\*Full Paper to be presented at Thursday 4:10 pm

**LAVE Video Editing Agent**

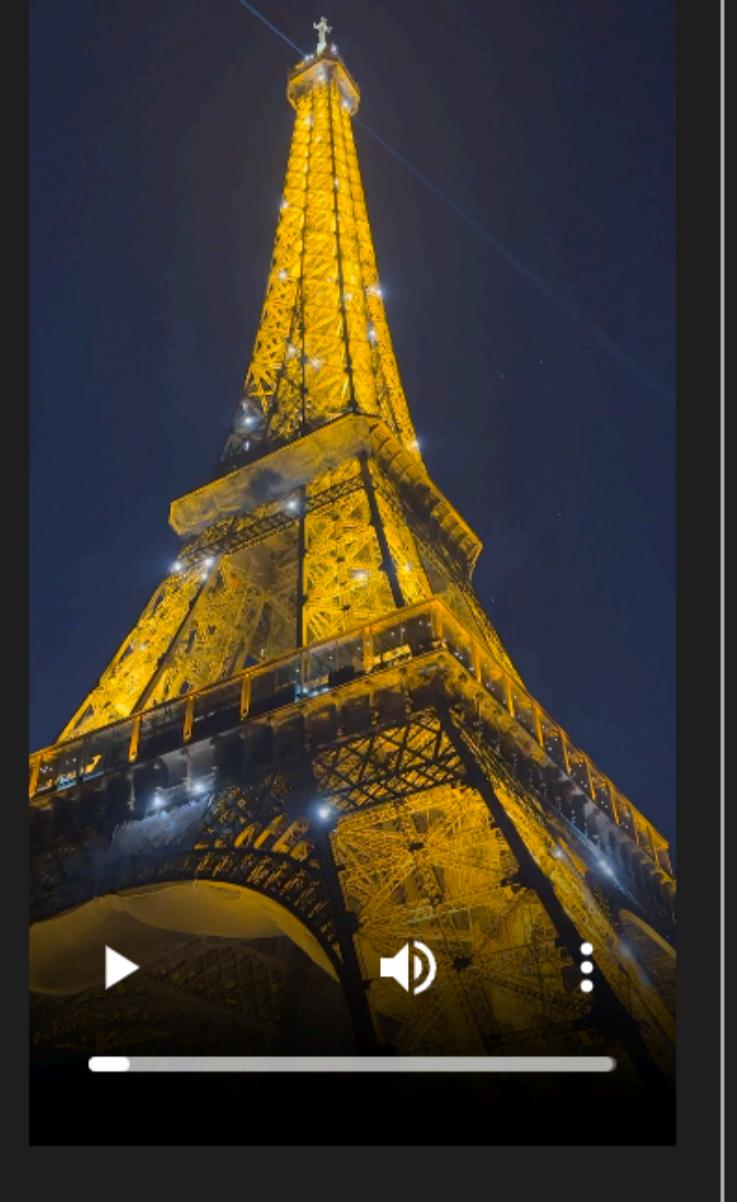
Agent: The current goal is User wants to find videos related to "eiffel tower".  
Let's think step by step. I will carry out the following actions to assist you:  
1. **Video Retrieval:** Search for videos related to "eiffel tower".  
Please type 'enter' to start executing the plan, or let me know how you want to change the plan!

**User: proceed**

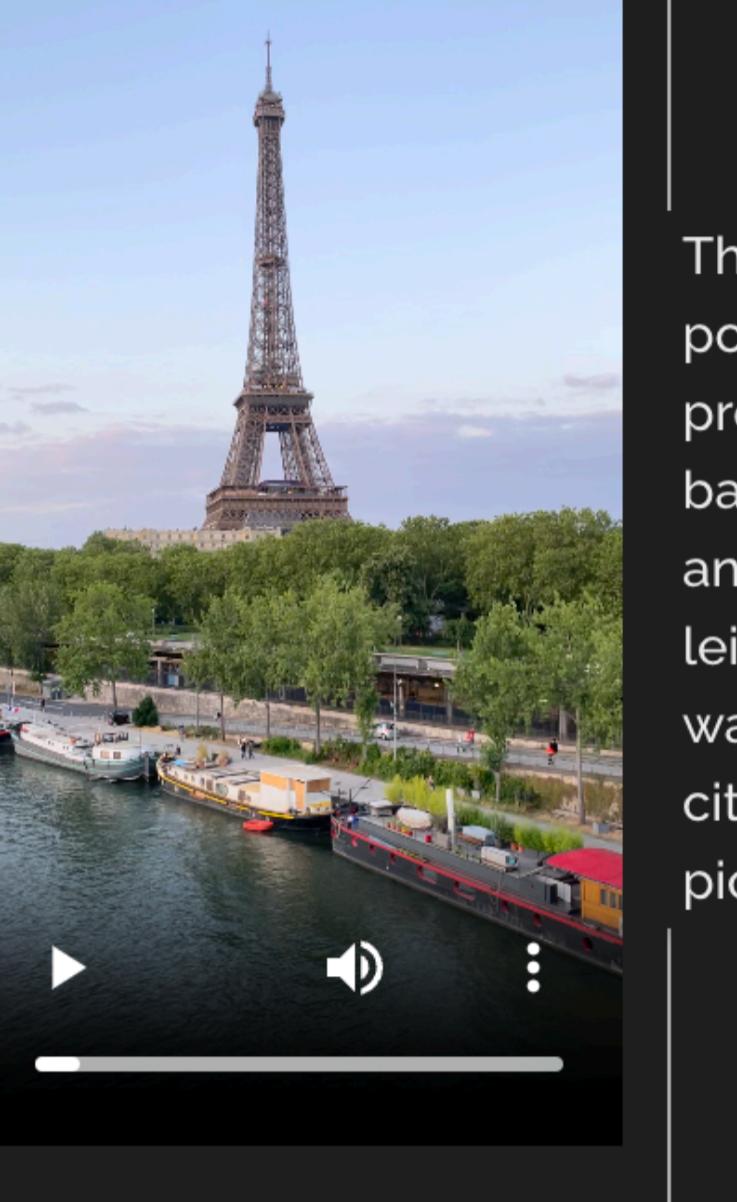
Agent: Sure, I've retrieved the relevant videos based on your query: eiffel tower.  
All actions have been completed. Feel free let me know what else can I help.

Type a message... Send

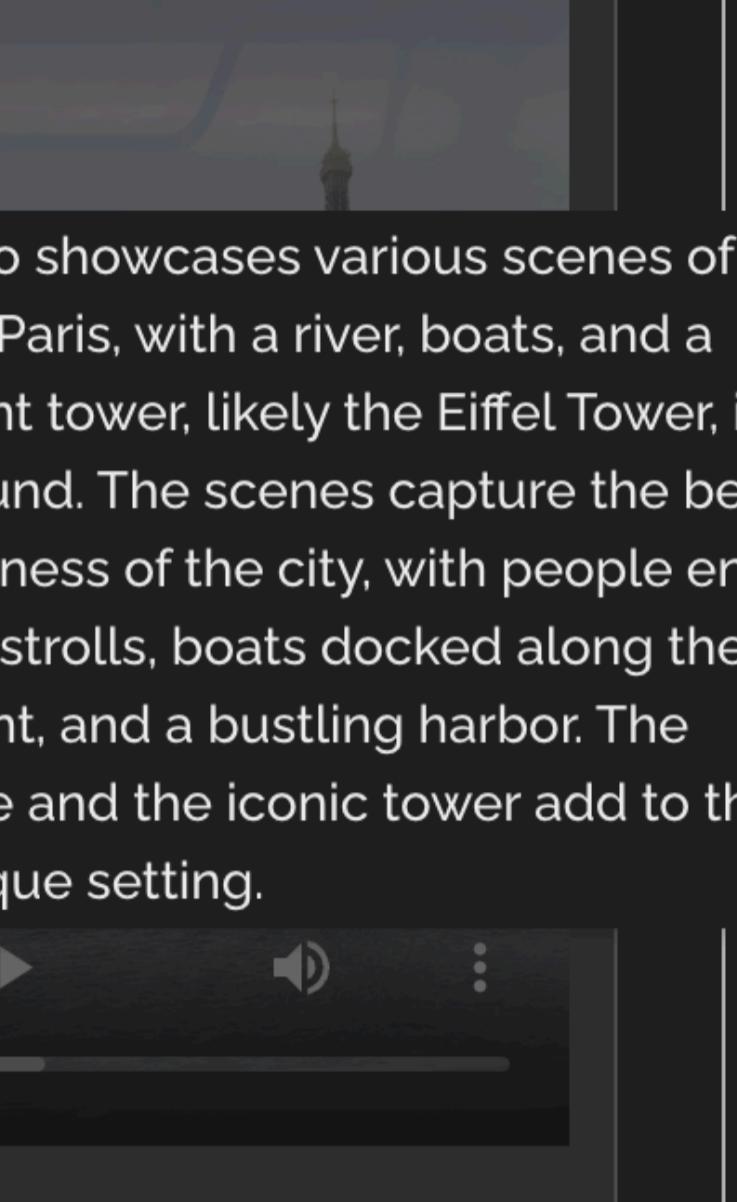
**Video Gallery** Add to Timeline Select/Deselect All



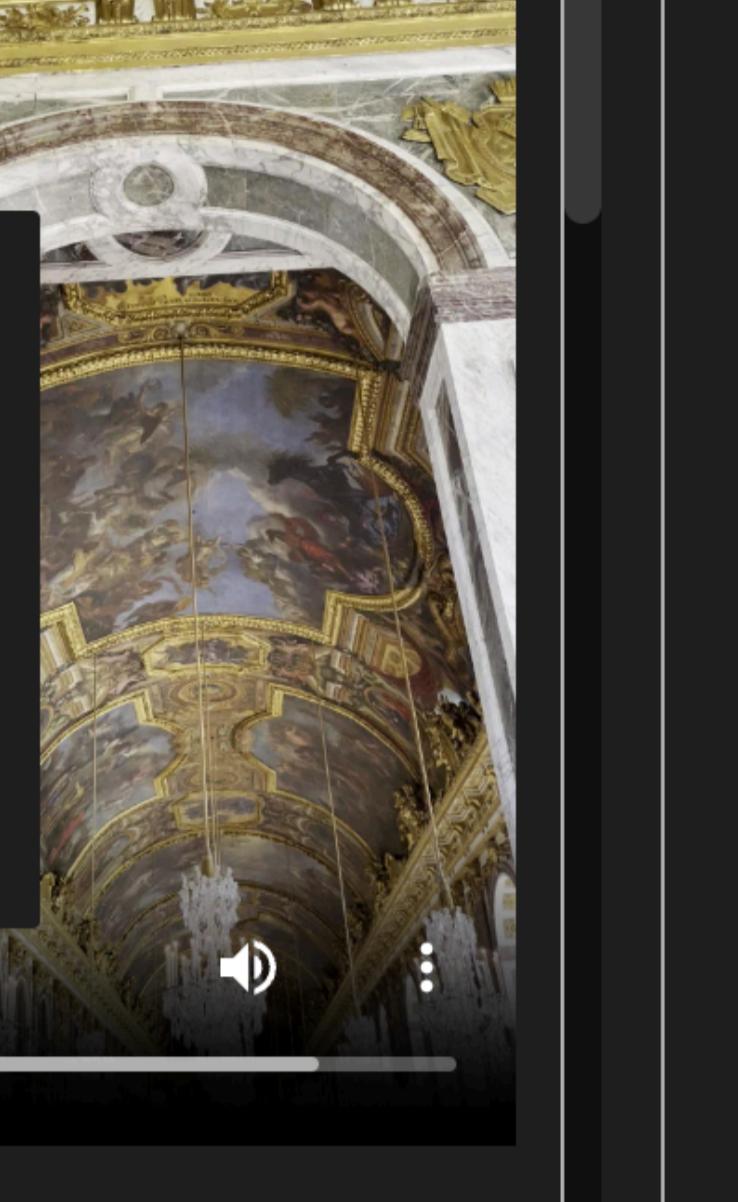
Eiffel Tower Illuminated at Night (9s)



Scenic Views of a City with River, Boats, and Tower (11s)



Scenic Views of a City: Capturing the Essence of Paris (7s)



Exploring a Grand Ornate Room (2s)

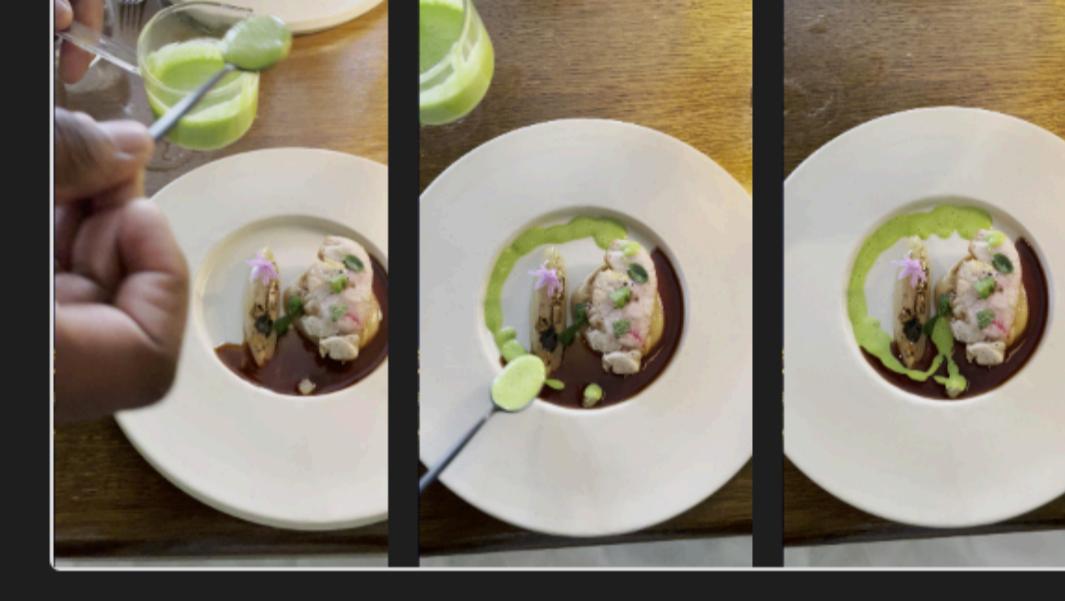
**Timeline** Play Delete Clear All Undo



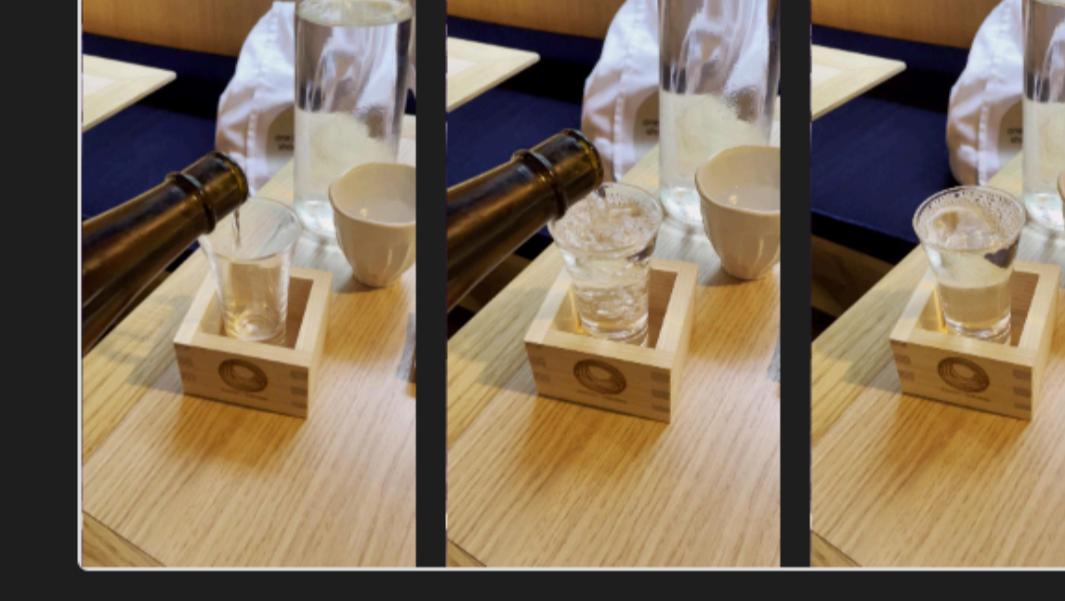
Delicious and Refreshing Drinks on a Counter (3s)



Scenes of Appetizing Outdoor Dining Tables (3s)



Elegant Meals on White Plates (11s)



Scenes of Wine and Champagne Pouring into a Glass (6s)



Scenic Views of a City: Capturing the Essence of Paris (7s)



Artistic Portraits in a Museum (4s)

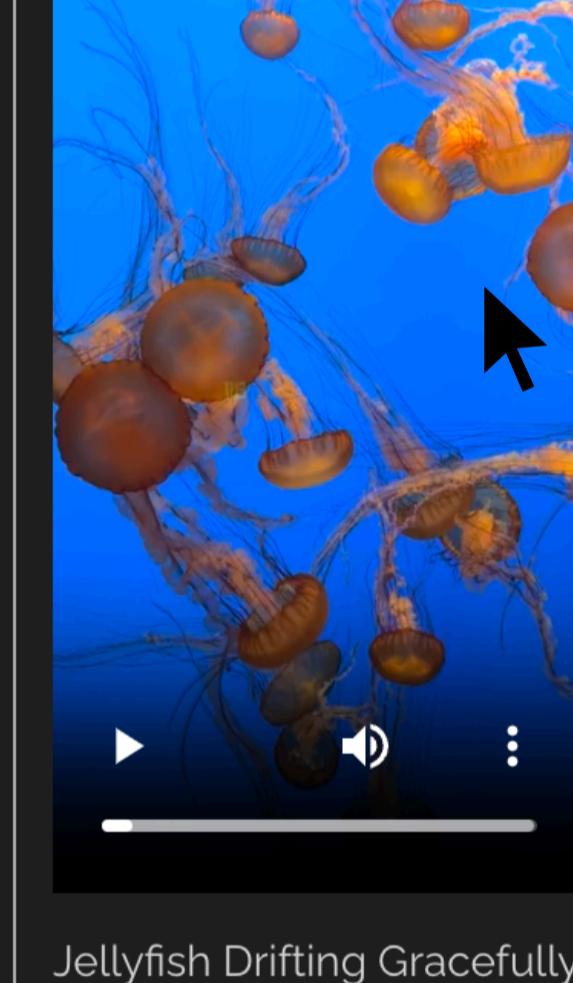
## Plan-and-Execute LLM Agents

**User:** I want to make a video for my france trip. Start from city vibes, to food, then end with my sister's wedding

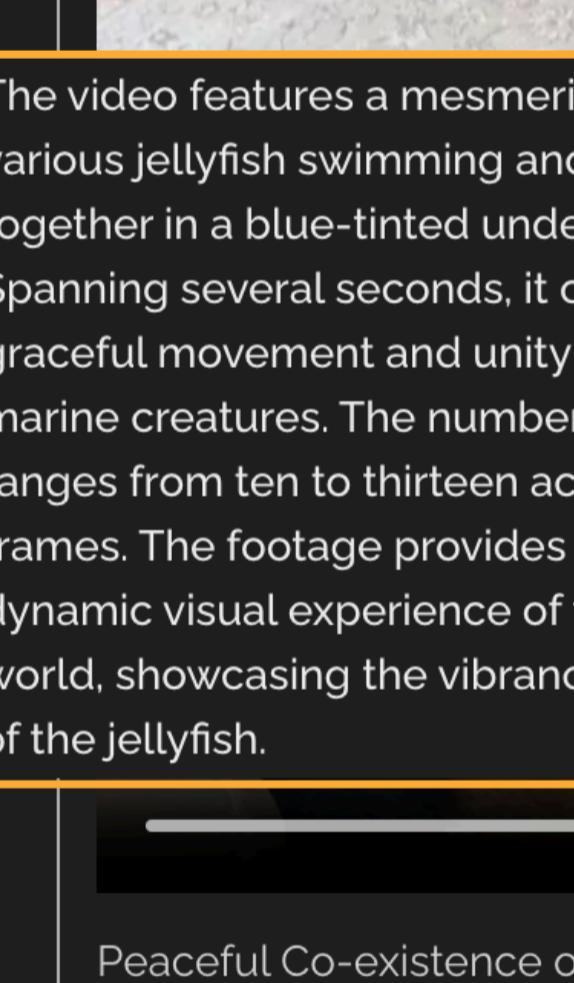
Agent: User's high level goal is to make a video for their trip to France, starting with city vibes, moving to food, and ending with their sister's wedding.  
Let's think step by step. I will carry out the following actions to assist you:  
1. **Video Retrieval:** Retrieve videos based on the query "city vibes in France". This will include videos showcasing the atmosphere, landmarks, and street scenes in various French cities.  
2. **Video Retrieval:** Retrieve videos based on the query "food experiences in France". This will include videos showcasing French cuisine, dining experiences, and food markets.  
3. **Video Retrieval:** Retrieve videos based on the query "wedding scenes". This will include videos showcasing wedding ceremonies, receptions, and celebrations.  
4. **Storyboarding:** Sequence the retrieved videos in the current timeline based on the storyline of your trip to France. The sequence will start with city vibes, move to food experiences, and end with your sister's wedding.  
Please type 'enter' to start executing the plan, or let me know how you want to change the plan!

## Language Augmented Video Gallery

**Video Gallery** Add to Timeline Select/Deselect All



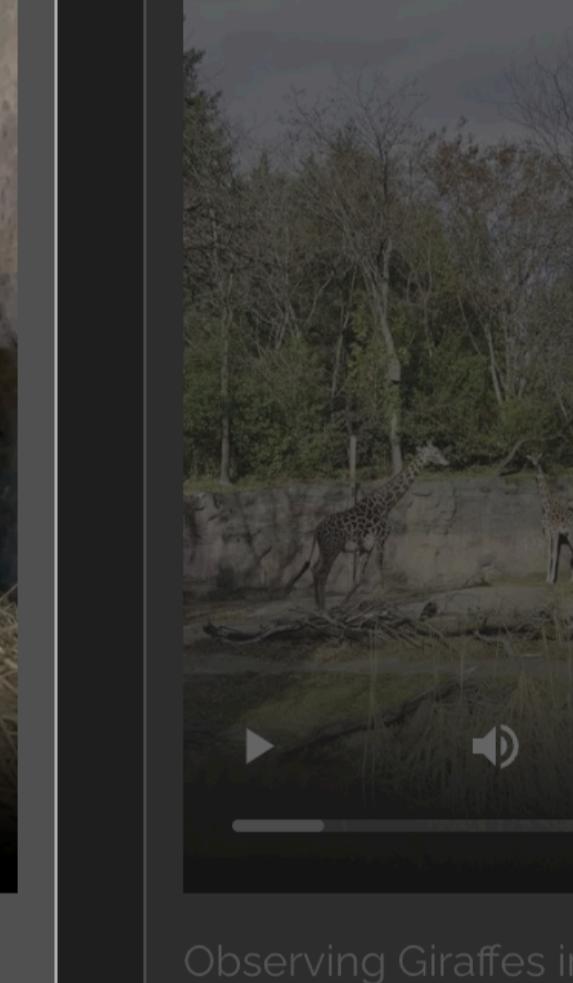
Jellyfish Drifting Gracefully in the Ocean (10s) A



Peaceful Co-existence of Cats and Dogs in a Cozy Living Room (24s) B



Lion Activities in a Zoo Enclosure (18s) C

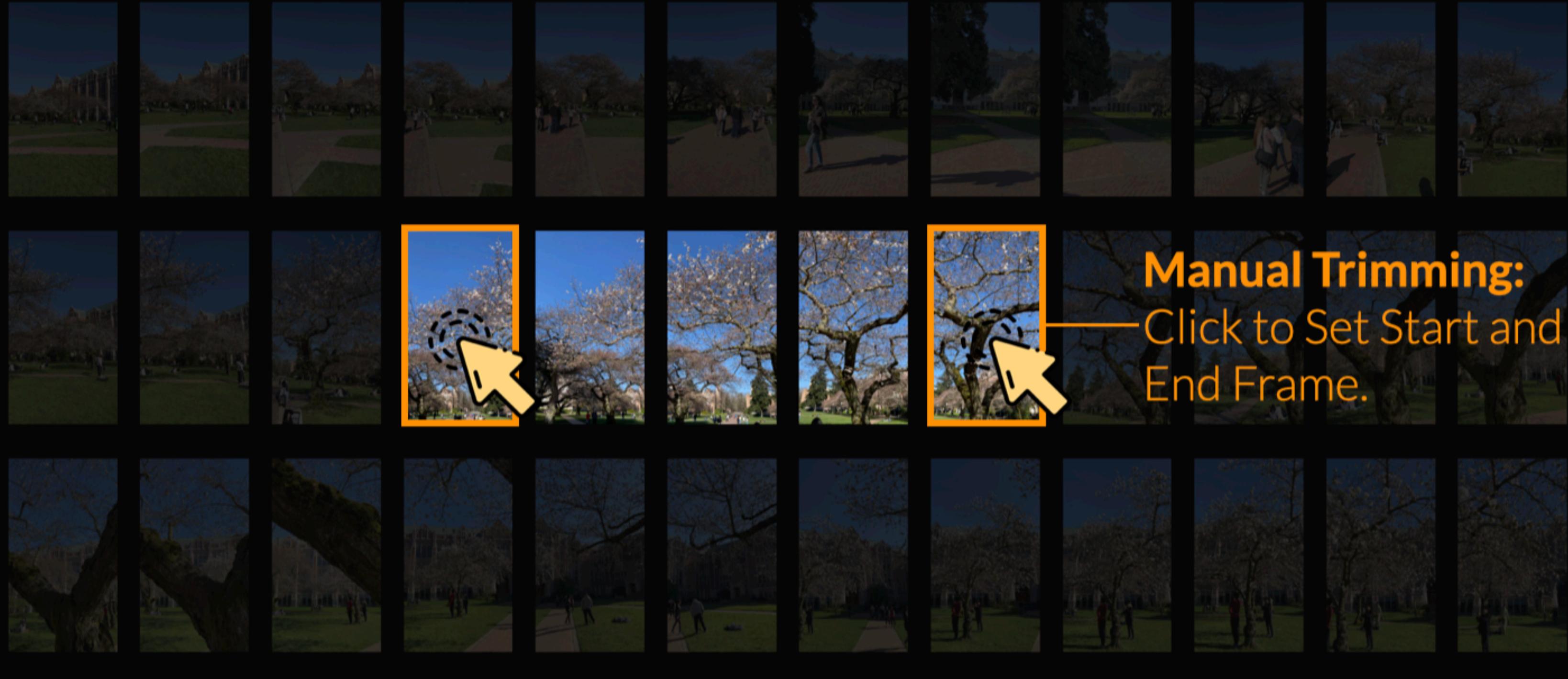


Observing Giraffes in a Zoo (14s)

## LLM-based Clip Trimming

Trimming for video: A Sunny Day at the Campus Park Preview

**A** Instruction: Each frame represents 1 second of video. You can click on the thumbnails to select the start and end frame, or use the text box below to specify your trimming command.



**Manual Trimming:**  
Click to Set Start and End Frame.

**B** Rationale: The cherry blossom tree is the main focus of the video from timestamp 15 to 19.

**C** Give me five seconds of video focusing on the cherry blossom tree nearby

**LLM Trimming:** Type Command for LLM Assistance. Send

LAVE is a video editing tool that integrates an LLM agent to collaborate with users. When users input an editing goal, the agent creates an action plan with editing functions to help achieve the goal. The system also supports direct manipulation, complementing agent automation and ensuring user control.

LAVE's LLM Editing Functions include:

- **Idea Brainstorming:** Generating video editing ideas.
- **Footage Overviewing:** Summarizing the video corpus.
- **Video Retrieval:** Searching for videos with language.
- **Storyboarding:** Sequencing clips based on narratives.
- **Clip Trimming:** Trim clips using language command.



Project Page



Demo Video



ArXiv