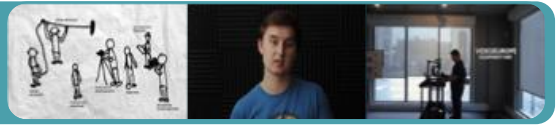
UNIVA | UNIFIED VIDEO CREATION AGENT

**User Input with Complex Task & Multi-Modal Input**
Create a prequel to the original video that introduces the backstory ……

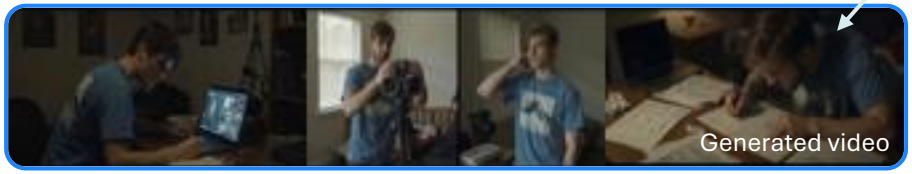


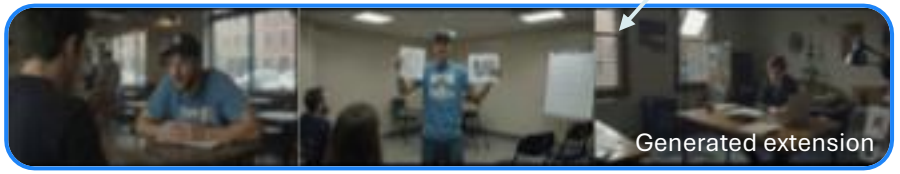
1. Video Understanding before Generation → 2. Seamless Generation → 3. Intention Understanding & Self-Reflection

First I will try to understand the video content. The video describes a boy narrates the process of his directorial creation…

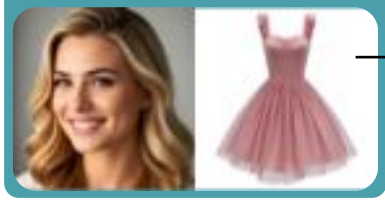
Now I fully understanded the video content, I will extract the relevant content to generate ……

Current video content don't satisfy the requirements, so I will extend this video ……



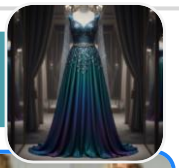

**User Input:**
Make the woman wear this dress, then turn a round to show

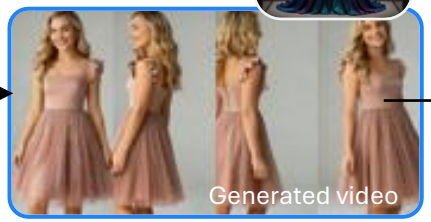


- ✓ Analyze to understand the woman's appearance and characteristics
- ✓ Analyze to understand the dress design and style
- ✓ Generate an image of the woman wearing the specified dress
- ✓ Generate a video of the woman in the dress turning around to show the outfit

Planner

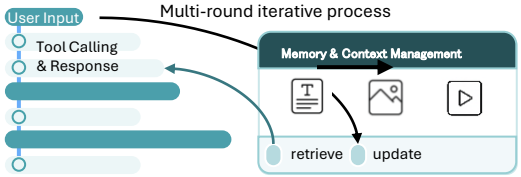
**Multi-Round Editing:**
Let she wear this dress.



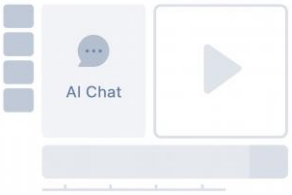



Plan & Tool Calling…

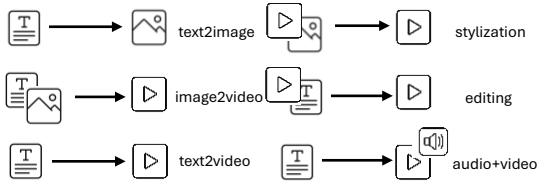

## Interactive, Immersive, Creative Video Generation




## User-Friendly Workspace




## Any-Conditioned Video Generation



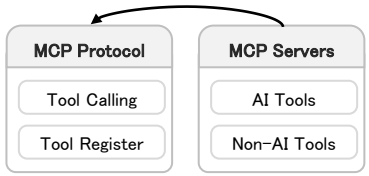

## Quality & Consistency


## Intention Understanding & Automation

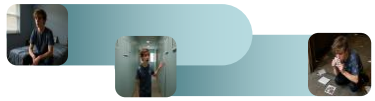
User Input: I want to stay in France <video>.

<thinking…> I guess you want to replace the background, so I segment the main object first, and use the editing tool to replace the background to Paris…

## Scalable: MCP-Based Framework

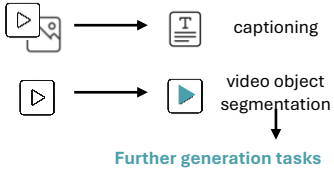



## Complex Video Generation



- User instruction with complex story
- Multi-scene/characters/camera pose changes
- Ultra long video generation/editing

## Comprehension & Perception




**Highly automated, interactive, proactive user experience**

**Comprehensive, industrial-level video production capabilities**