MVPrompt:
Building Music–Visual Prompts for AI Artists to Craft Music Video Mise-en-scène
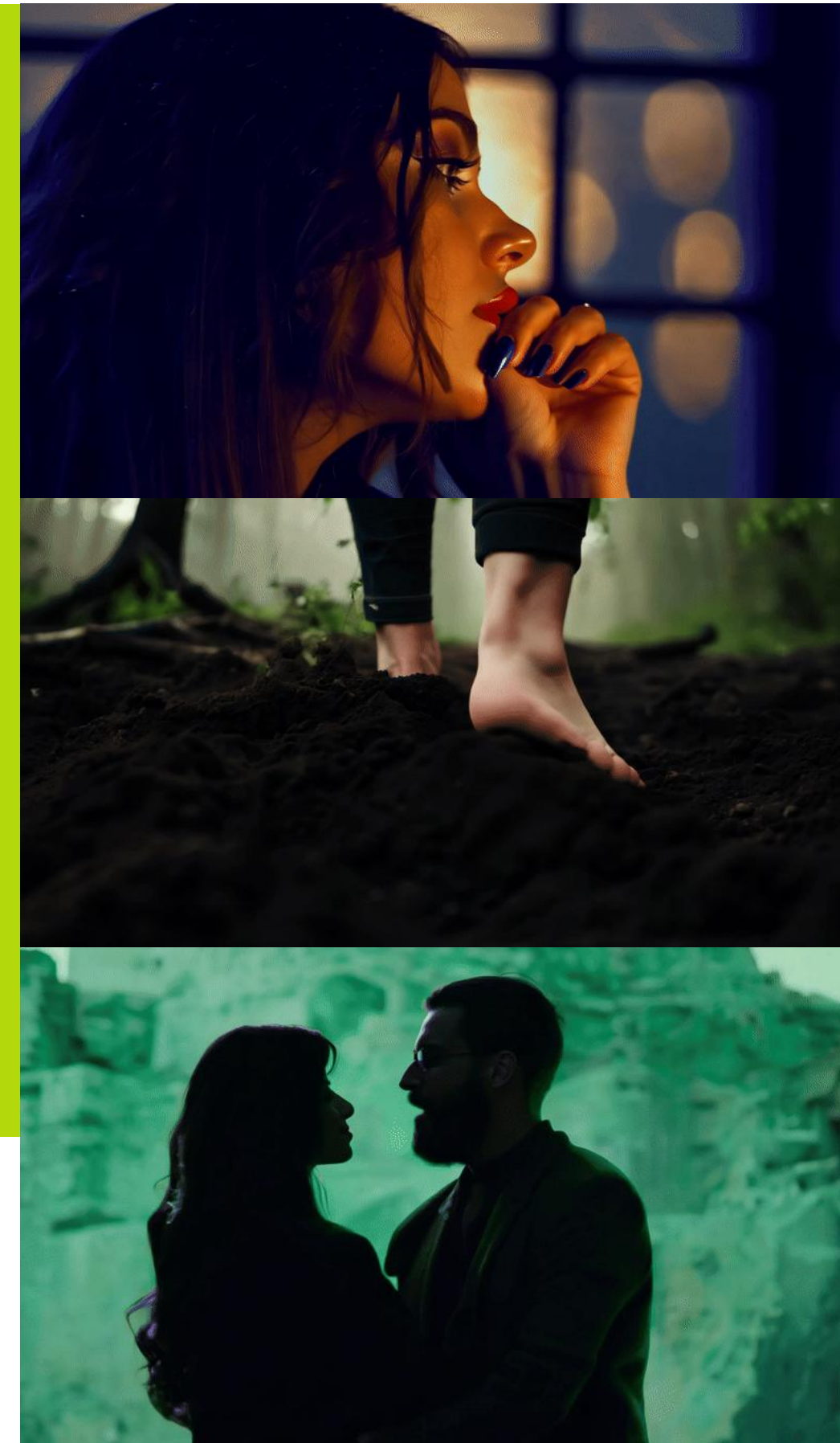
ChungHa Lee
Presenter

DaeHo Lee

Jin-Hyuk Hong*

GWANGJU INSTITUTE OF SCIENCE AND TECHNOLOGY
**Soft Computing & Interaction Lab**

INTRODUCTION

Music videos have traditionally been the domain of experts, but with text-to-video generative AI models, <u>AI artists can now create music video scene</u> more easily [37].

<u>Mise-en-scène,</u> rooted in film theory, focuses on harmoniously arranging all visual components within a scene to convey semantic meanings [57].

MVPROMPT: BUILDING MUSIC-VISUAL PROMPTS
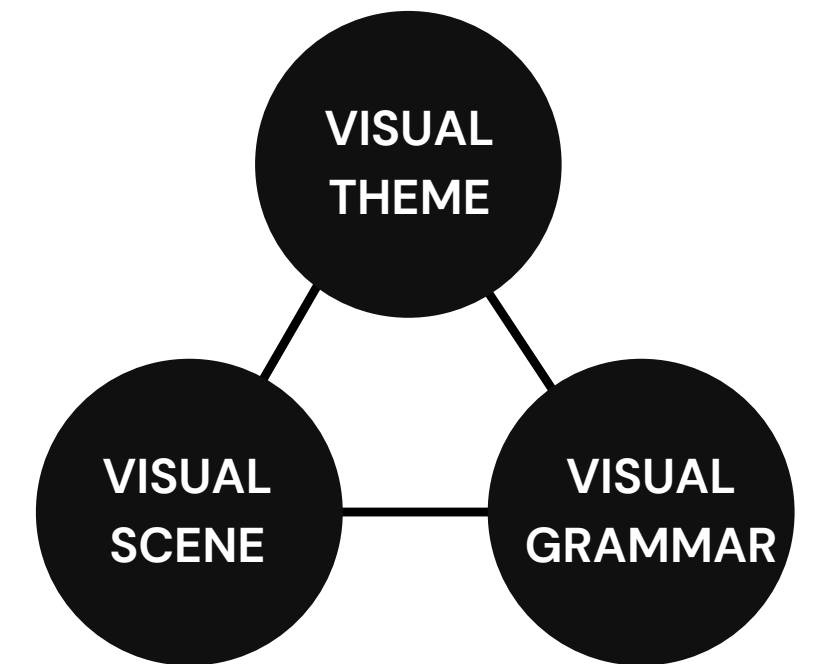FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE-EN-SCÈNE
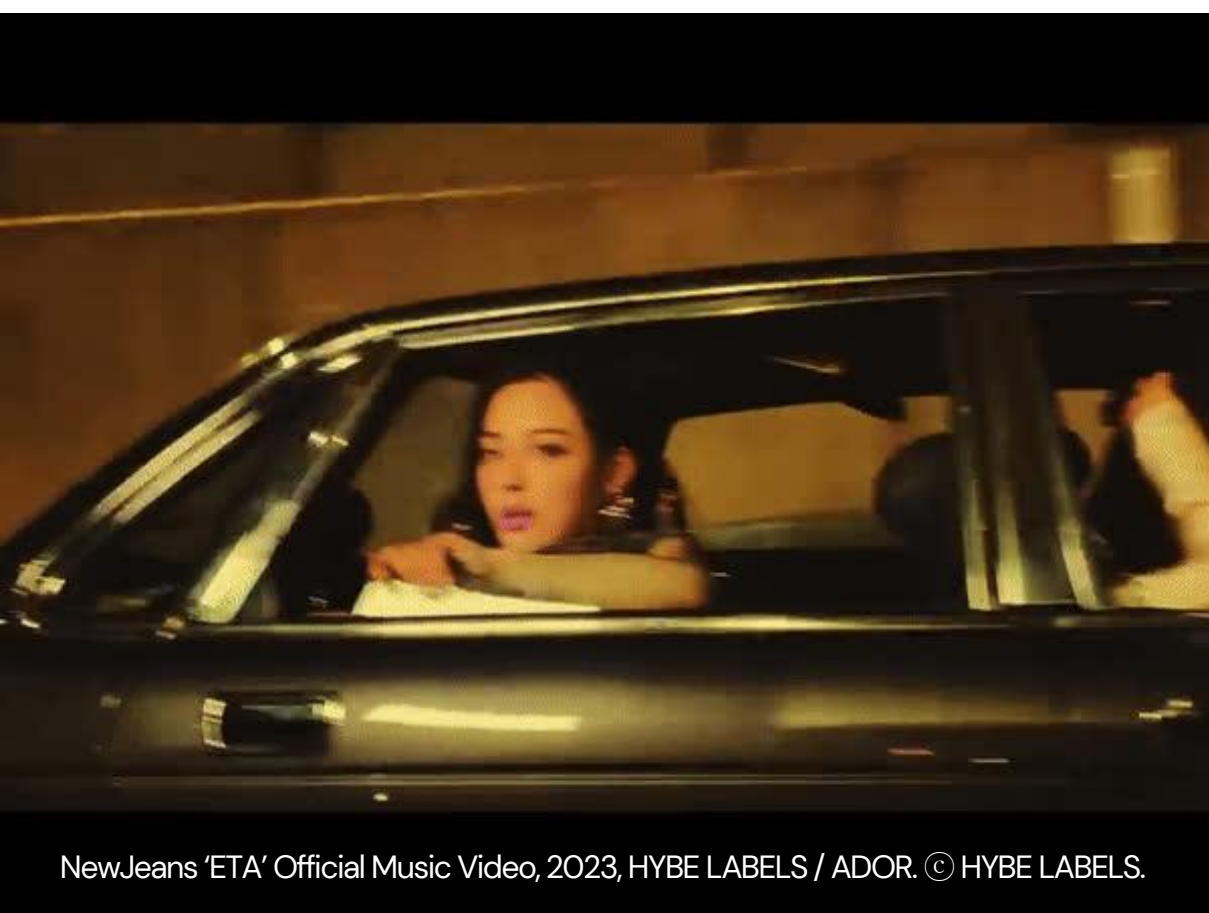
INTRODUCTION

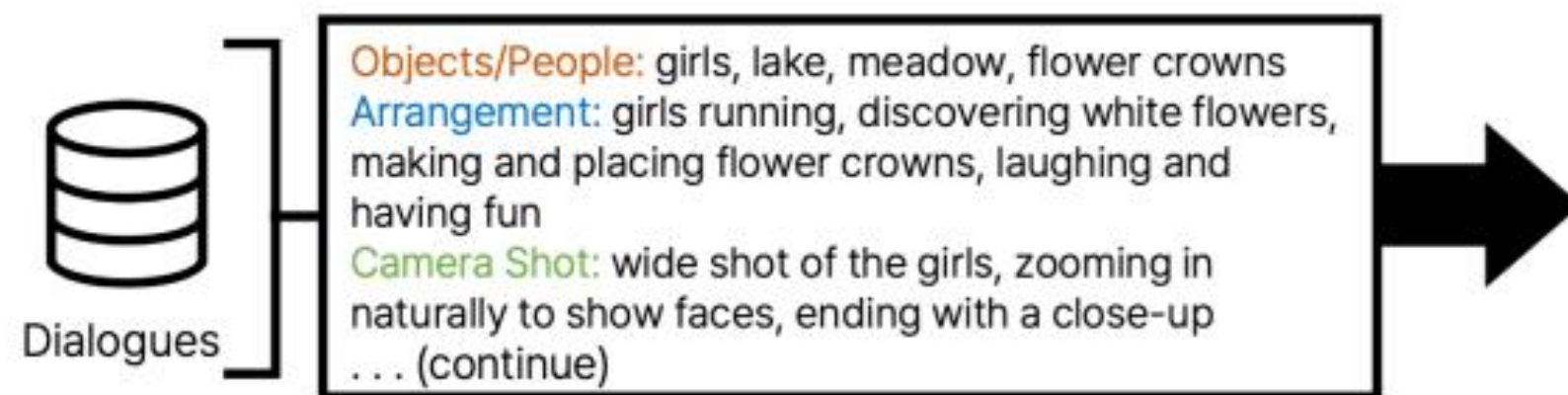NewJeans 'ETA' Official Music Video, 2023, HYBE LABELS / ADOR. © HYBE LABELS.

Videos convey dynamic information, making it difficult to refine them without expertise in video production [36, 89] – especially in music videos where maintaining visual coherence and conveying music is essential [79].

Accurately reflecting the desired mise–en–scène remains challenging for AI artists without specialized knowledge [20], highlighting the need for prompt supportive tools.

MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE

In our study, we introduce MVPrompt (Music-Visual Prompt), a prompt-supportive tool designed to support AI artists in making music video scenes with music-visual mise-en-scène.

VISUAL THEME

VISUAL SCENE

VISUAL GRAMMAR

Three Mise-en-scène Features for Music Video Scene Prompts

Objects/People: girls, lake, meadow, flower crowns
Arrangement: girls running, discovering white flowers, making and placing flower crowns, laughing and having fun
Camera Shot: wide shot of the girls, zooming in naturally to show faces, ending with a close-up
. . . (continue)

Dialogues

CoT-based four phases of the creative dialogue

Desired music video scene with MVPrompt

MVPROMPT: BUILDING MUSIC-VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE-EN-SCÈNE

RELATED WORKS

01
GenAI Video Production and Creative Needs of AI Artists

02
Mise-en-scène in Music Videos

03
Needs for Prompt Supportive Tools for Text2Video

MVPROMPT: BUILDING MUSIC-VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE-EN-SCÈNE

GenAI Video Production and Creative Needs of AI Artists

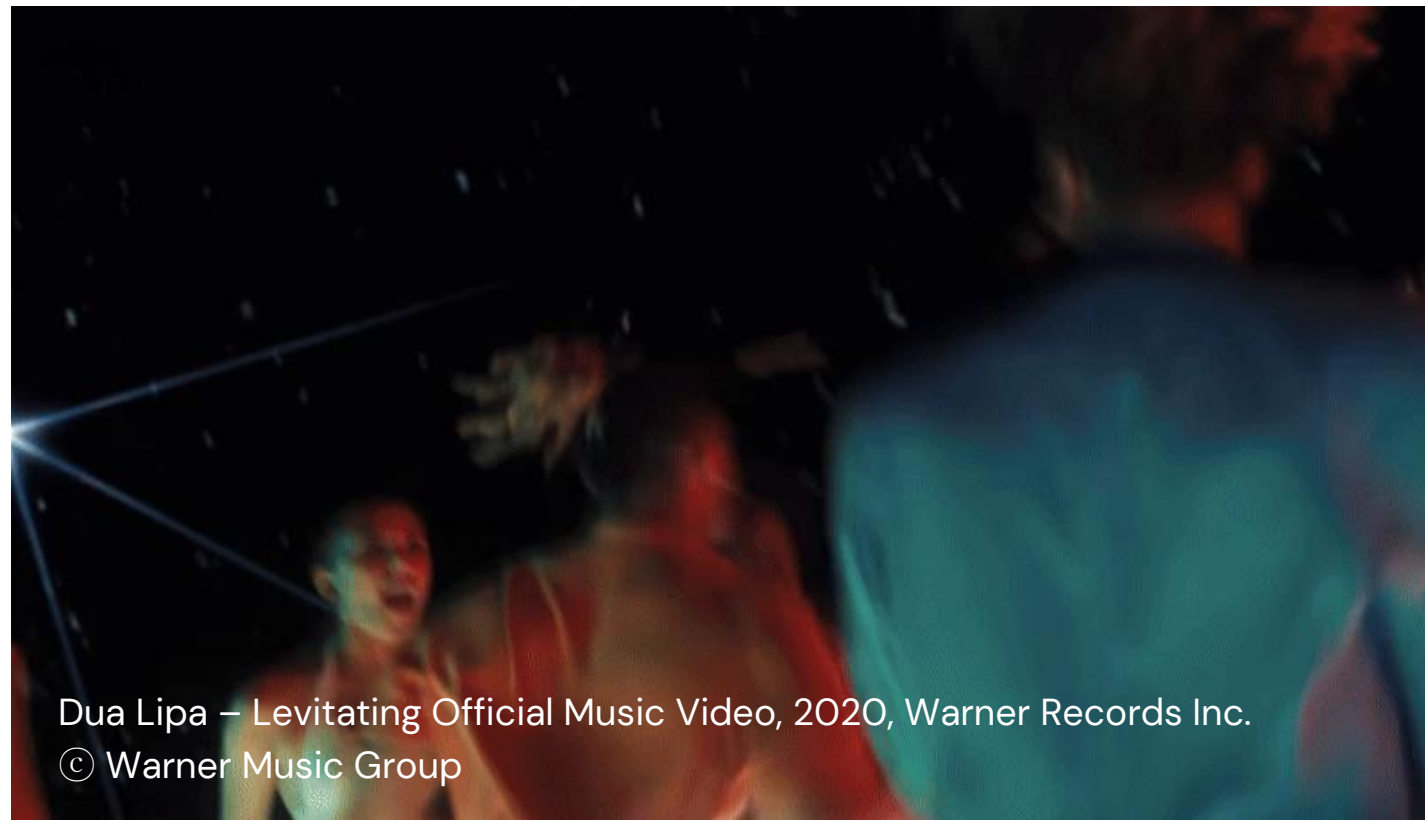Joy Buolamwini / Mario Klingemann / Songwen Chung

AI artists are human creators who leverage GenAI to engage in artistic process [3, 10, 77].

Recent GenAI video tools allowed for high-quality video creation, to produce longer videos [41, 50].

Mise-en-scène in Music Videos

Dua Lipa – Levitating Official Music Video, 2020, Warner Records Inc.
© Warner Music Group

e.g., ***Fast-tempo music*** *paired with dynamic visuals conveys energy and excitement*

Music videos extend auditory sensations into a visual form, offering a medium to convey the experience of music [26, 82].

The Congruence-Association Model [79]

Audio-Visual Synchrony [15, 18]

MVPROMPT: BUILDING MUSIC-VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE-EN-SCÈNE

Mise-en-scène in Music Videos

Billie Eilish – Birds of a Feather Official Music Video, 2024, Darkroom/Interscope Records. © Darkroom/Interscope Records.

*e.g., **Slow-tempo music** paired with smooth, flowing visuals conveys calmness*

Music videos extend auditory sensations into a visual form, offering a medium to convey the experience of music [26, 82].
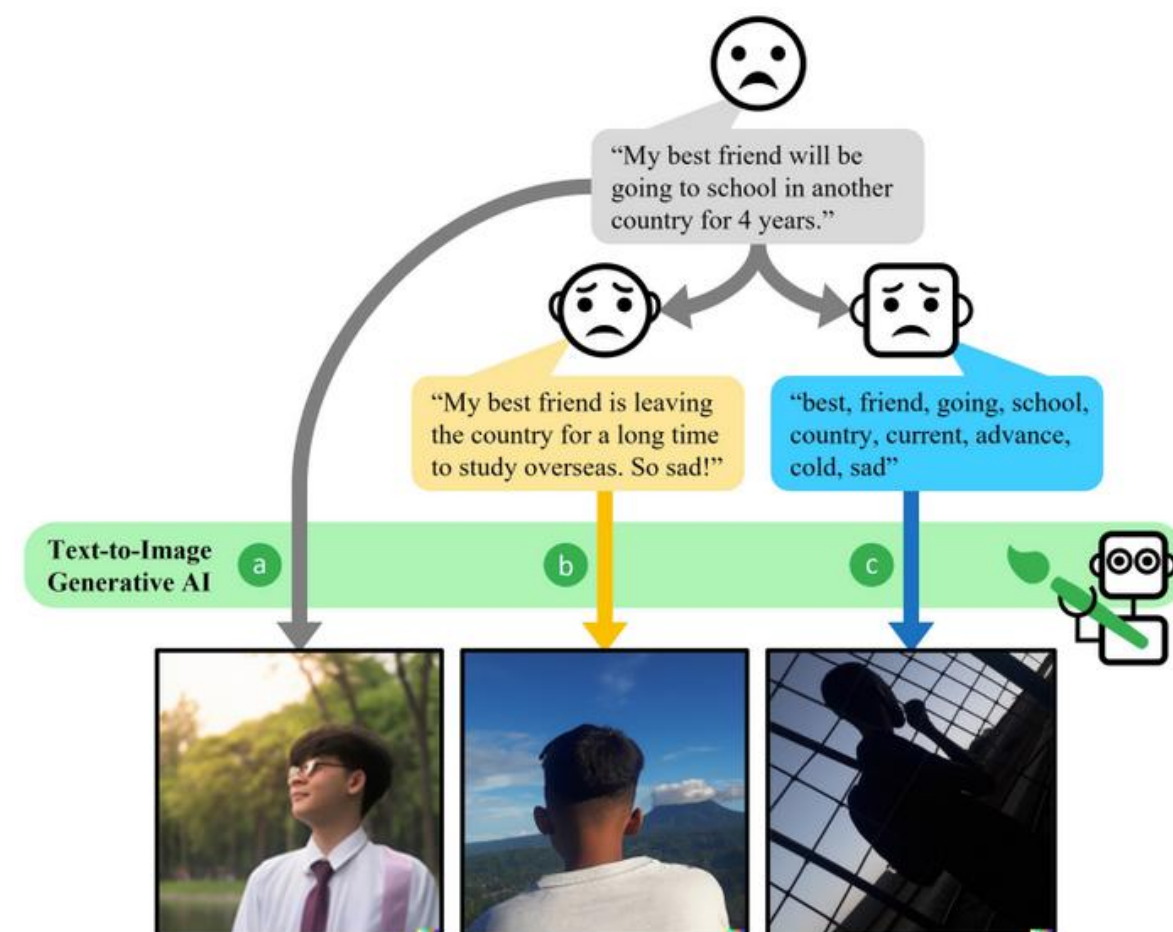
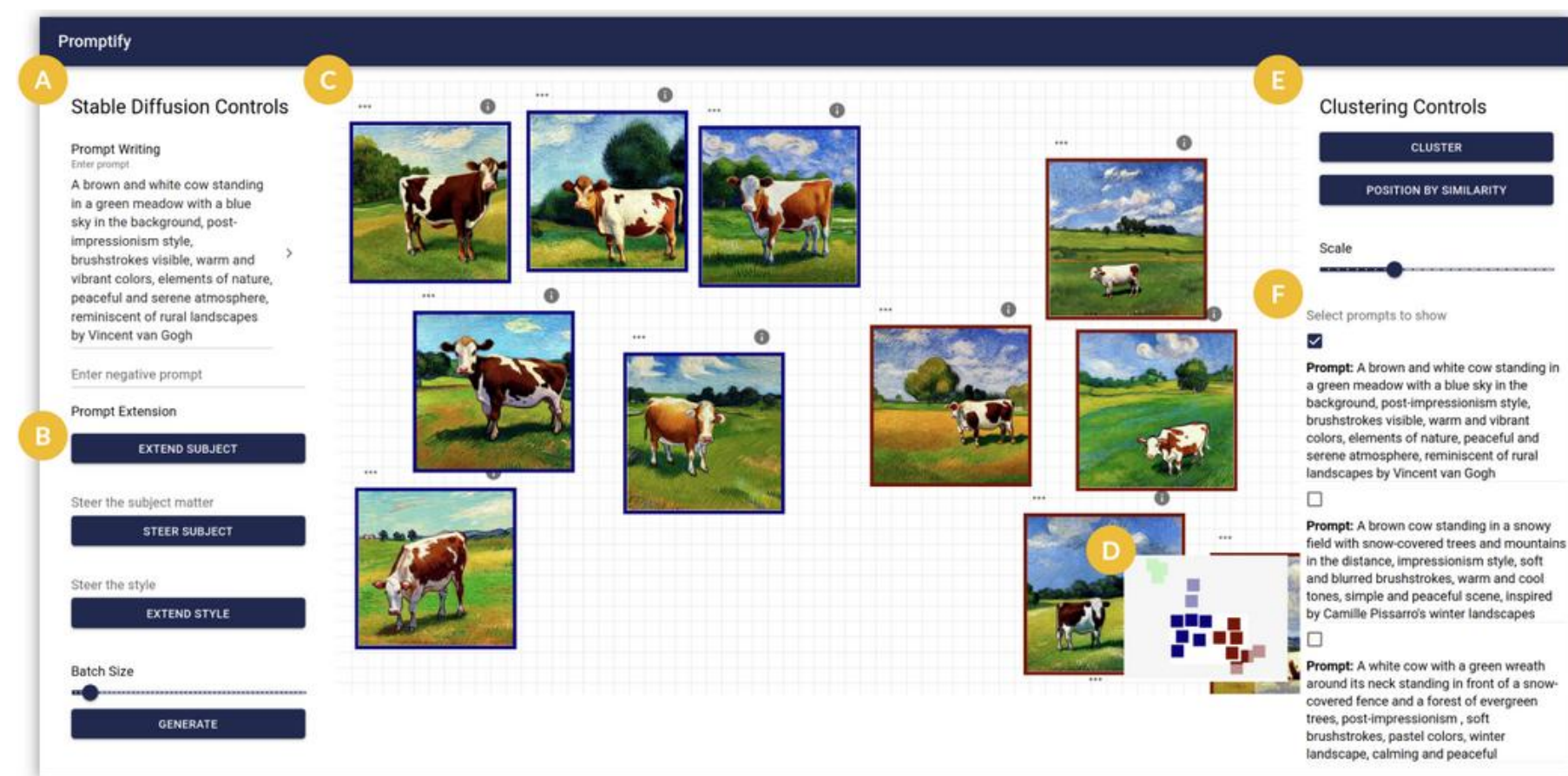The Congruence-Association Model [79]

Audio-Visual Synchrony [15, 18]

08

MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE

Needs for Prompt Supportive Tools for Text2Video

RePrompt, Wang *et al.,*
CHI 2023

Promptify, Brade *et al.,*
UIST 2023

MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE

# Needs for Prompt Supportive Tools for Text2Video

Magic Prompt (Runway)
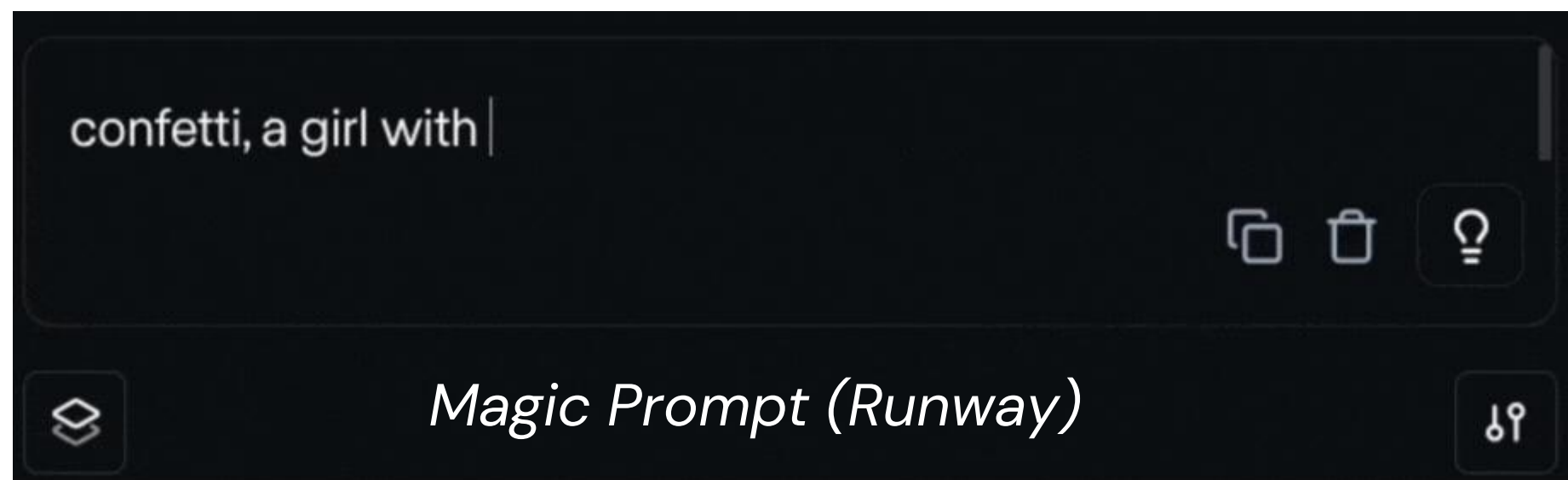
Enhancing text prompt with AI may lead to the ***one-size-fits-all solution*** [56]

Unlike previous studies and services, our research adopts a more personalized approach by capturing the creative intentions of AI artists.

DESIGN WORKSHOP

Participants and Procedure

- With seven music video experts (via Zoom)

- Experts had over three years of experience planning, directing, filming and editing music videos respectively (See Table 1 for the details)

60 min   *60 min*   *by researchers*
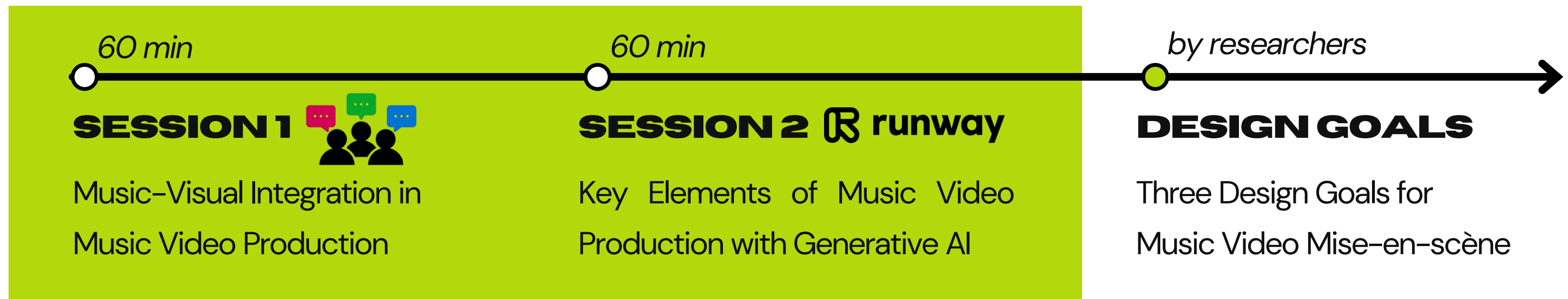
SESSION 1

Music-Visual Integration in

Music Video Production

SESSION 2 runway

Key Elements of Music Video

Production with Generative AI

DESIGN GOALS

Three Design Goals for

Music Video Mise-en-scène

Music–Visual Integration in Music Video Production

Pinterest

[FILMGRAB]

PLANNING

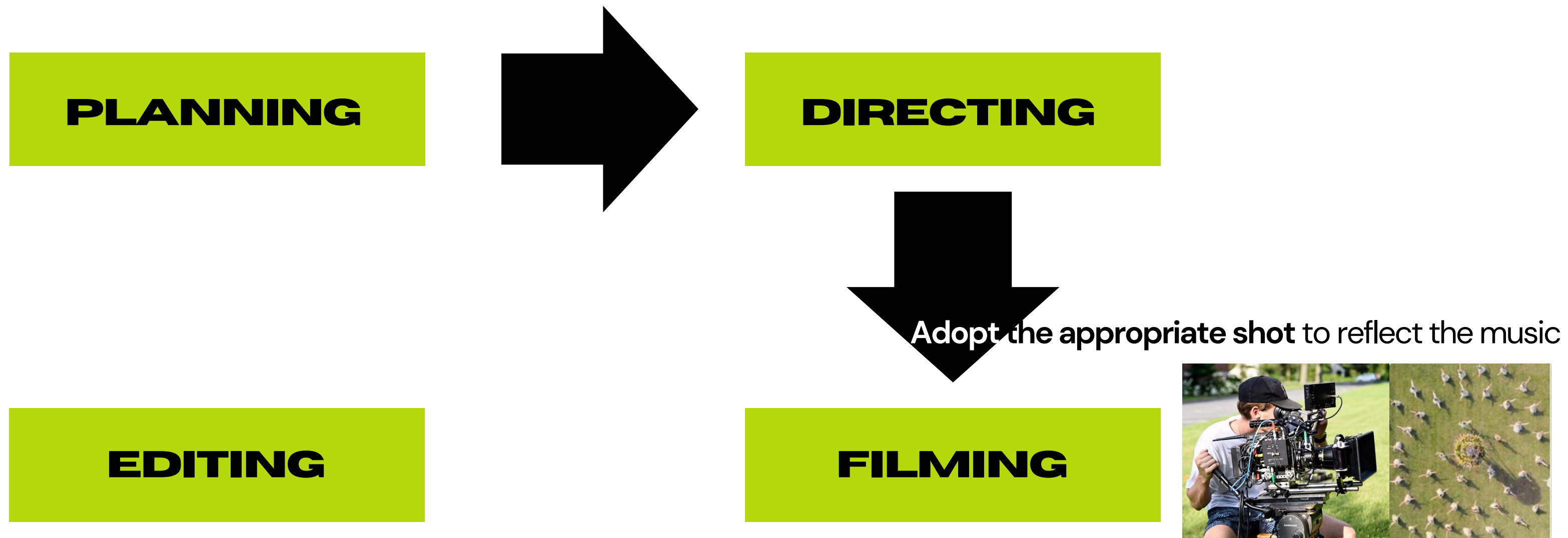**Explore images and videos**
to deepen the key visual theme

DIRECTING

EDITING

FILMING
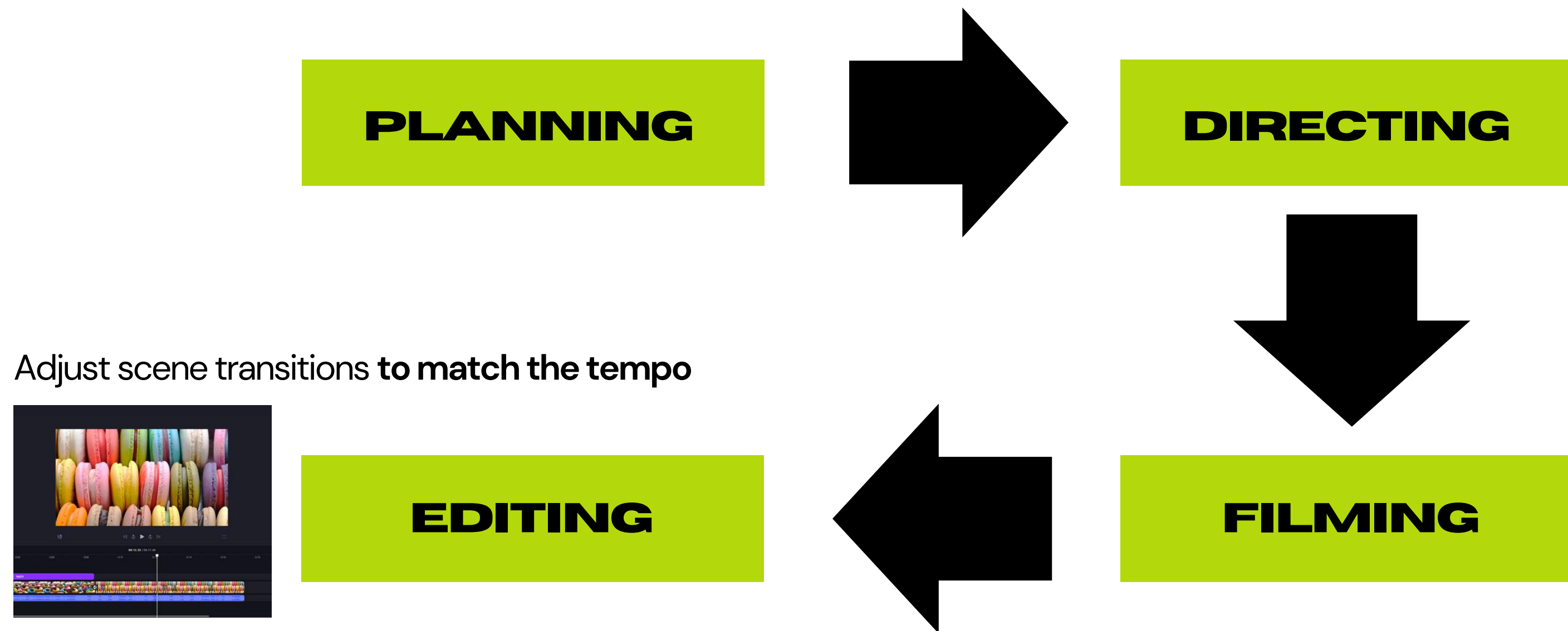
Music–Visual Integration in Music Video Production

PLANNING

DIRECTING

Adopt the **appropriate shot** to reflect the music

EDITING

FILMING

Key Elements of Music Video Production with Generative AI

Suggestion of Music-Driven Theme

e.g., Dark and intense music ➔ Striking and contrasting colors

Prompt Specification for Music Video Scene

Lighting ➔ **natural light at 2 PM**

➘ **artificial light in a dark alley**

Instructional Guidance on Arrangement and Camera Shots

Providing explanations of film grammar terms like *"rule of thirds"* or *"over-the-shoulder shot"*

Dialogue-based Creative Context for Mise-en-scène

*"**Like layering a cake**, it is important to start by clarifying the overall theme, ... , after that, the conversation can move on to more specific elements." – E2*

DESIGN WORKSHOP – **RESULT**

Mise–en–scène Features for Music Video Scene Prompts

01
Visual Theme

The key visual that determines the overall tone and manner, represented as the **aesthetic theme**

02
Visual Scene

The core elements that compose the scene: **objects, background, colors, and lighting**

03
Visual Grammar

The compositional elements that decide how components should be presented: **arrangement and camera shots**

DESIGN WORKSHOP – **RESULT**

Design Goals for MVPrompt

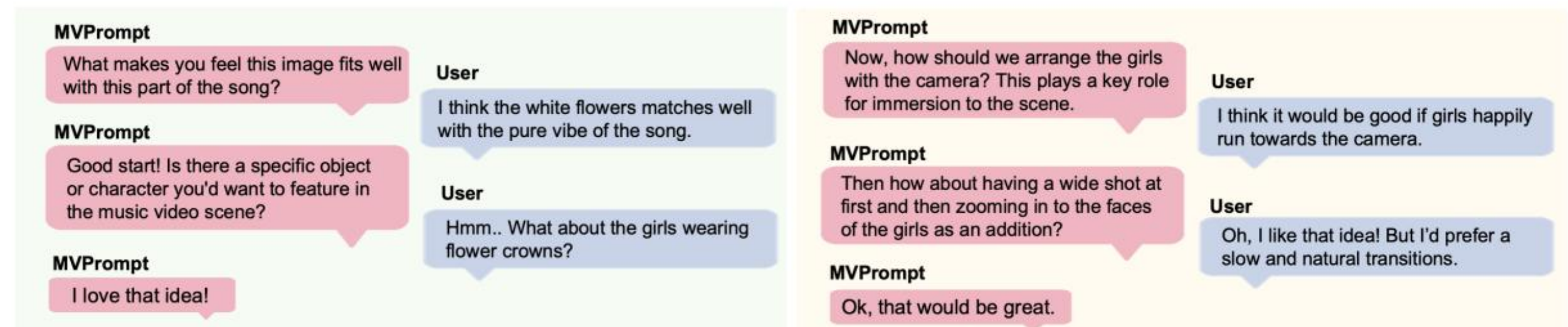| DG	STAGE	DESCRIPTION
1	**Planning & Directing**	Recommend visual themes and images based on the musical features
2	**Filming & Editing**	Support concretizing visual scenes and grammar via dialogue
3	**All Stages**	Easily guide users to determine visual mise-en-scène step-by-step

MVPROMPT

**Stage 1.
Visual Theme**

**Stage 2.
Visual Scene
& Grammar**

A. Ideation on the Visual Theme

Bloomcore

.mp3 file

metadata
(title, lyrics)

Music

Bloomcore
Vectorgarden
Dreamy

Show **music-related themes**
based on the music analysis

Let users to explore **key visual images** from the Pinterest
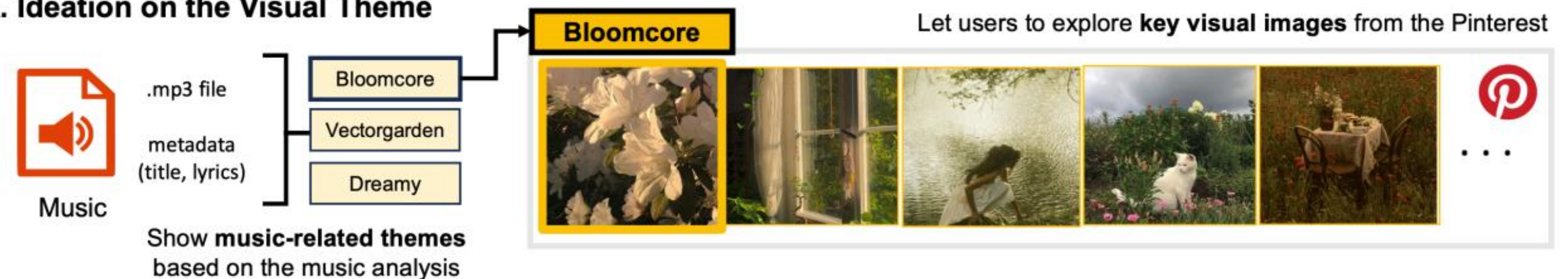
B. Interactive Dialogue Visual Scene & Grammar

MVPrompt
What makes you feel this image fits well
with this part of the song?

MVPrompt
Good start! Is there a specific object
or character you'd want to feature in
the music video scene?

MVPrompt
I love that idea!

User
I think the white flowers matches well
with the pure vibe of the song.

User
Hmm.. What about the girls wearing
flower crowns?

B-1. The dialogues for specifying **visual scene**
(objects, background, color, lighting) prompt

MVPrompt
Now, how should we arrange the girls
with the camera? This plays a key role
for immersion to the scene.

MVPrompt
Then how about having a wide shot at
first and then zooming in to the faces
of the girls as an addition?

MVPrompt
Ok, that would be great.

User
I think it would be good if girls happily
run towards the camera.

User
Oh, I like that idea! But I'd prefer a
slow and natural transitions.

B-2. The dialogues for specifying **visual grammar**
(arrangement, camera shot) prompt

16

MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE

MVPROMPT

System Structure

MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE
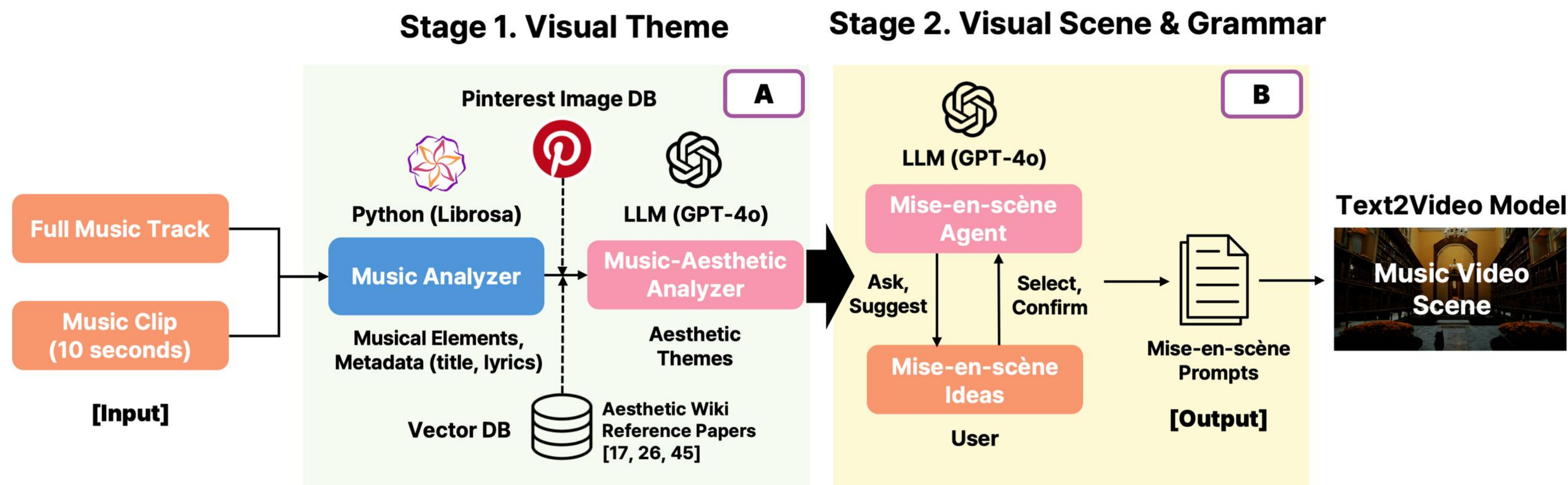
MVPROMPT

4–Steps for Creative Dialogue

Phase 2: **Select**

Phase 4: **Confirm**

Initial Object

Specified Object

Confirmed Object

Phase 1: **Ask**

"What object would you like to feature in the scene? The object plays an an essential role of symbolism in music video."

roses

bugs

Bugs gathering around the roses, bothering them

A scene where bugs approach and bother the roses, but the roses remain strong and resilient.

"Roses and bugs **are a great idea!** [Compliment]
What about bugs gathering around the roses? [Suggestion]
This can make the scene more dramatic, [Reasons behind]
showing how someone stays strong
and doesn't give up, even in tough situations."

Phase 3: **Suggest**

MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE

USER STUDY DESIGN

Study Protocol & Setup

SYSTEM	VECTOR DB	MUSIC ANALYZER	STAGE 1 AESTHETIC ANALYZER	STAGE 2 4–STEPS DIALOGUE
MVPrompt	O	O	O	O
w/o Visual Theme	O	O	X	O
w/o Visual Scene & Grammar	O	O	O	X
Baseline	O	X	X	X

We designed an ablation study to examine the impact of individual stages of MVPrompt.
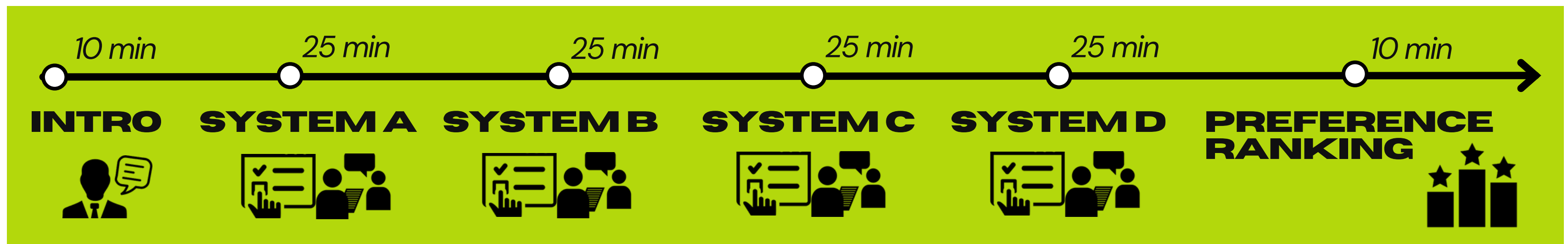
MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE

USER STUDY DESIGN

Participants

- Based on the principles of "*low thresholds, high ceilings, and wide walls*" [71, 76], we recruited **24 AI Artists** who actively use GenAI tools for video production as job or personal projects.

- Participants had over six months of experience in GenAI video production (See Table 3 for the details).

- With a **within-subject design**, we used complete counterbalancing, each participant engaging **120 minutes**.

MVPROMPT: BUILDING MUSIC-VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE-EN-SCÈNE

USER STUDY DESIGN

Methodology ◪ **runway** (Gen–3 Alpha model)

Timeline: INTRO (10 min) → SYSTEM A (25 min) → SYSTEM B (25 min) → SYSTEM C (25 min) → SYSTEM D (25 min) → PREFERENCE RANKING (10 min)

We compared user's creative experiences across the four system utilizing…

- **Creativity Support Index (CSI) [11]**

- **Experience of Creating Music Video Scene Prompts [68]**

- **Learnability Survey**

- **Preference Ranking**

FINDINGS

Friedman Test for non-parametric testing
Benjamini-Hochberg for multiple comparison

Creativity Support Index
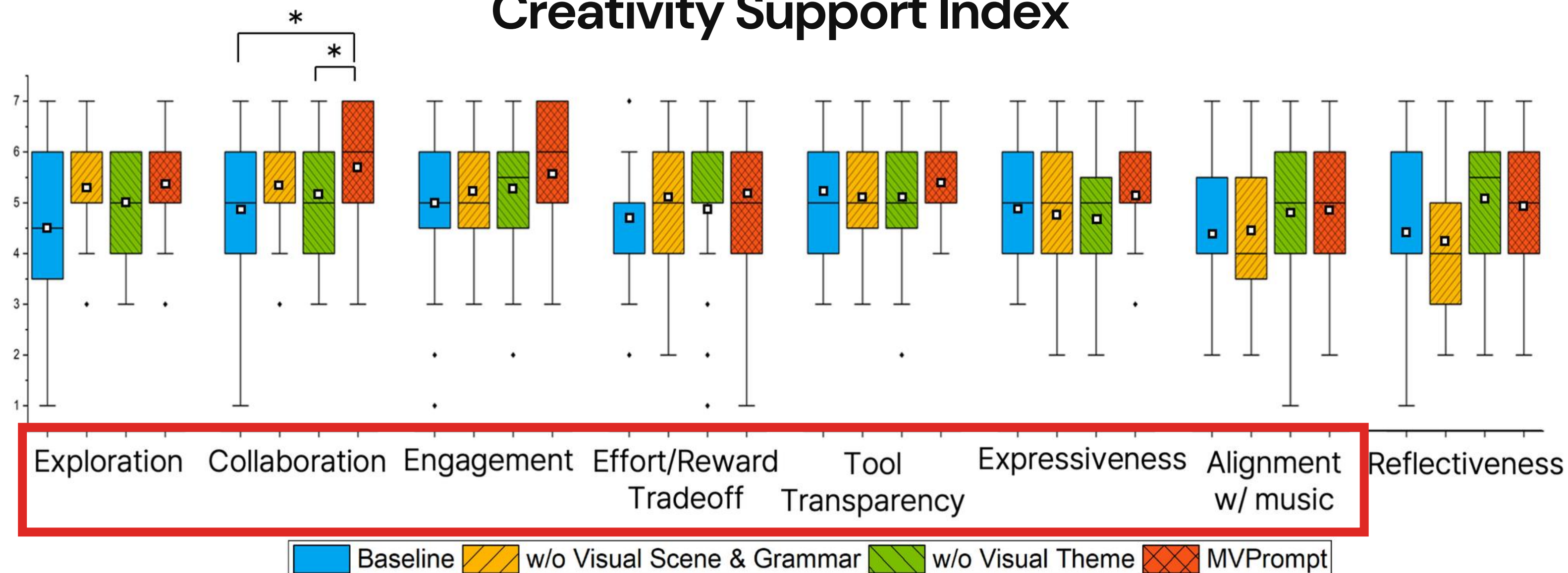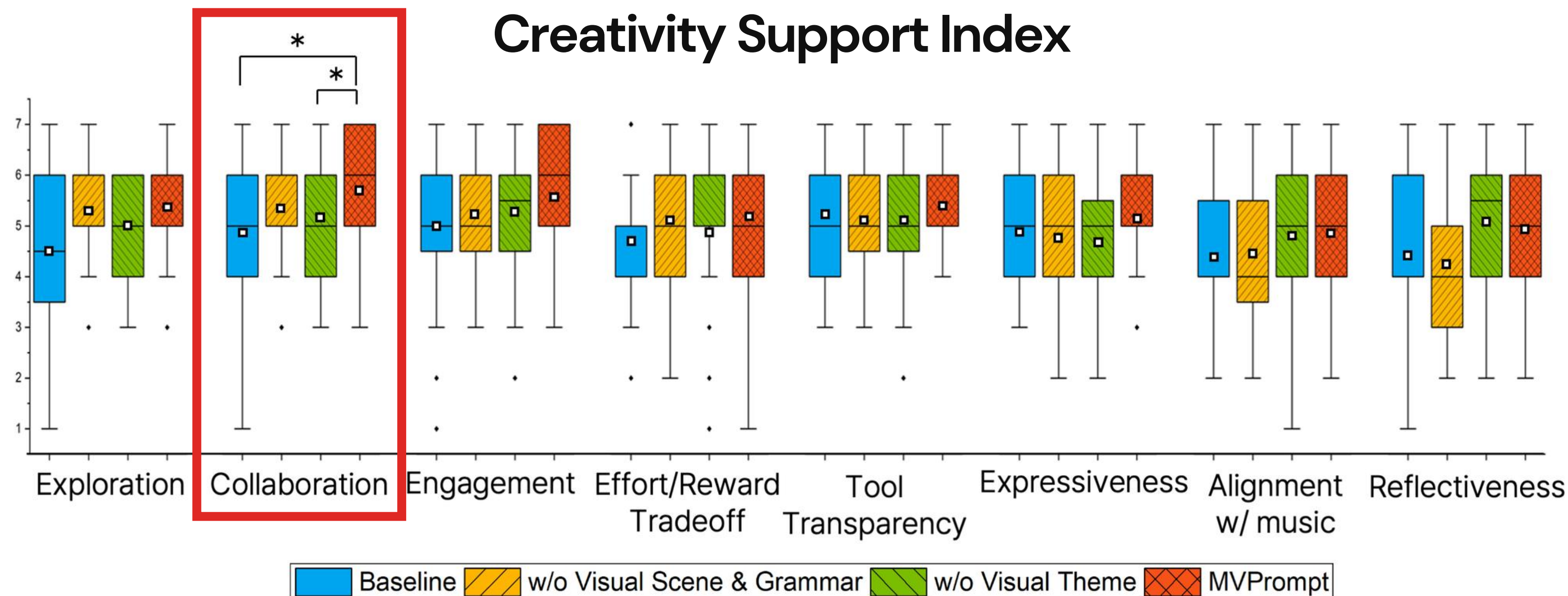
Participants rated MVPrompt as the best system they collaborated in the process.

MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE

Collaboratively Incorporating Music-Visual Mise-en-scène

Chosen aesthetic theme: **Lovecore**

| PHASE 1: ASK | PHASE 2: SELECT | PHASE 3: SUGGEST | PHASE 4: CONFIRM |

*"It helped me **to decide on prompts step by step** by prompting me to think about elements I only had vague ideas about"* – P7

MVPROMPT: BUILDING MUSIC-VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE-EN-SCÈNE

Collaboratively Incorporating Music-Visual Mise-en-scène

Chosen aesthetic theme: **Lovecore**

PHASE 1: ASK	PHASE 2: SELECT	PHASE 3: SUGGEST	PHASE 4: CONFIRM

*"It helped me **to decide on prompts step by step** by prompting me to think about elements I only had vague ideas about"* – P7

MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE

Collaboratively Incorporating Music-Visual Mise-en-scène

CHI 2025
April 26–May 1, 2025 in Yokohama, Japan

Chosen aesthetic theme: **Lovecore**

PHASE 1: ASK	PHASE 2: SELECT	**PHASE 3: SUGGEST**	PHASE 4: CONFIRM

*"It helped me **to decide on prompts step by step** by prompting me to think about elements I only had vague ideas about"* – P7

MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE

Collaboratively Incorporating Music-Visual Mise-en-scène

Chosen aesthetic theme: **Lovecore**

PHASE 1: ASK	PHASE 2: SELECT	PHASE 3: SUGGEST	PHASE 4: CONFIRM

*"It helped me **to decide on prompts step by step** by prompting me to think about elements I only had vague ideas about"* – P7

FINDINGS – **CREATIVITY SUPPORT INDEX**

Collaboratively Incorporating Music-Visual Mise-en-scène

P7 's scene generated with MVPrompt

Aesthetic Theme: Lovecore
Objects/People: Hamsters holding heart-shaped guitars, wearing sunglasses, with fluffy fur, each playing instruments
Background: Apartment living room with pink wallpaper, pink mood lamp in the back
Color Scheme: Pastel pink, pop pink, white, with gold and silver accents
Lighting: General brightness, soft and diffused lighting
Arrangement: Triangular arrangement of hamsters playing instruments
Camera Shot: Full shot of the apartment, moving through a window to show the living room

MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE

Collaboratively Incorporating Music-Visual Mise-en-scène

Baseline received the lowest average score across six CSI items, including *Collaboration.*

P6 's scene generated with Baseline
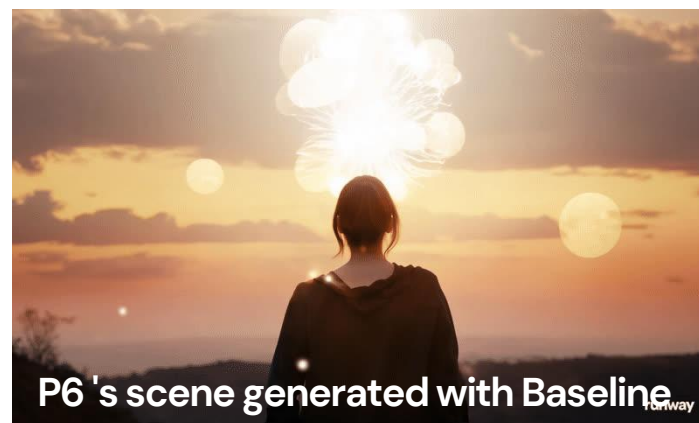
*"**I had imagined a dark and intense mood for this song**, but the system suggested a symbol of light. **It didn't match the music at all.**"* – P6

*"It suggested elements for the music video, but **it was hard to explore how to turn those elements into moving scenes**."* – P20

MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE

FINDINGS – **CREATIVITY SUPPORT INDEX**

Shaping Expressive Visuals in Each Stages

Stage 1. Visual Theme

Stage 2. Visual Scene & Grammar

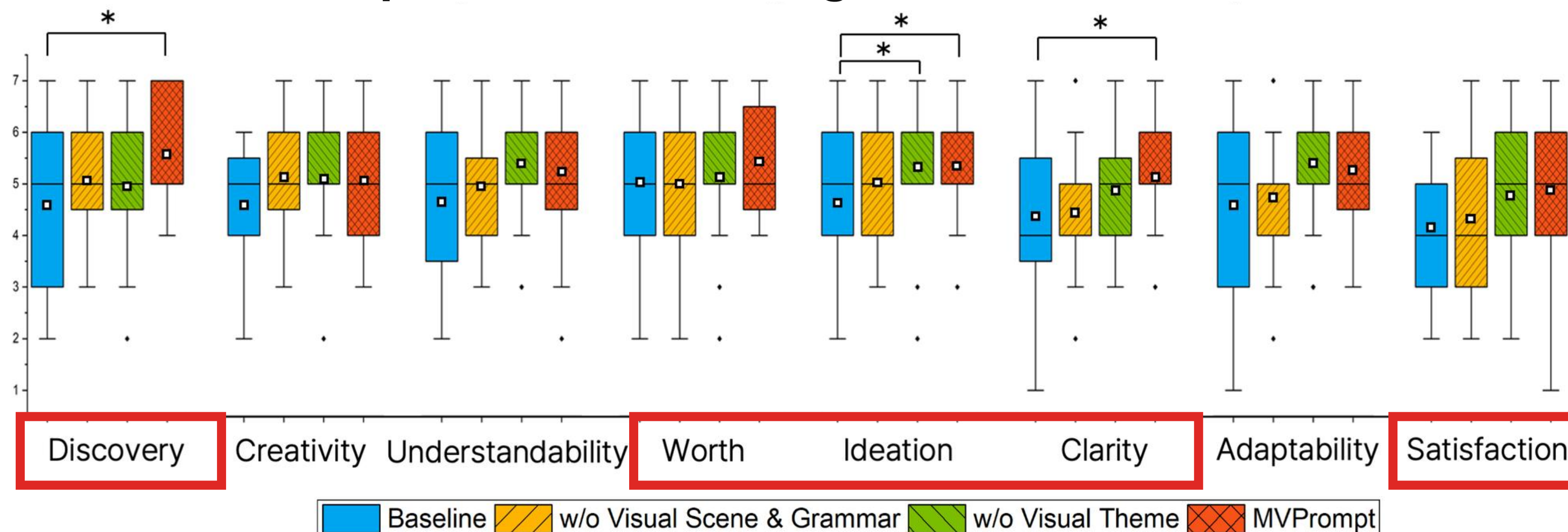**Defining a theme from the music helped to determine the <u>tone and manner of the scene</u>.**
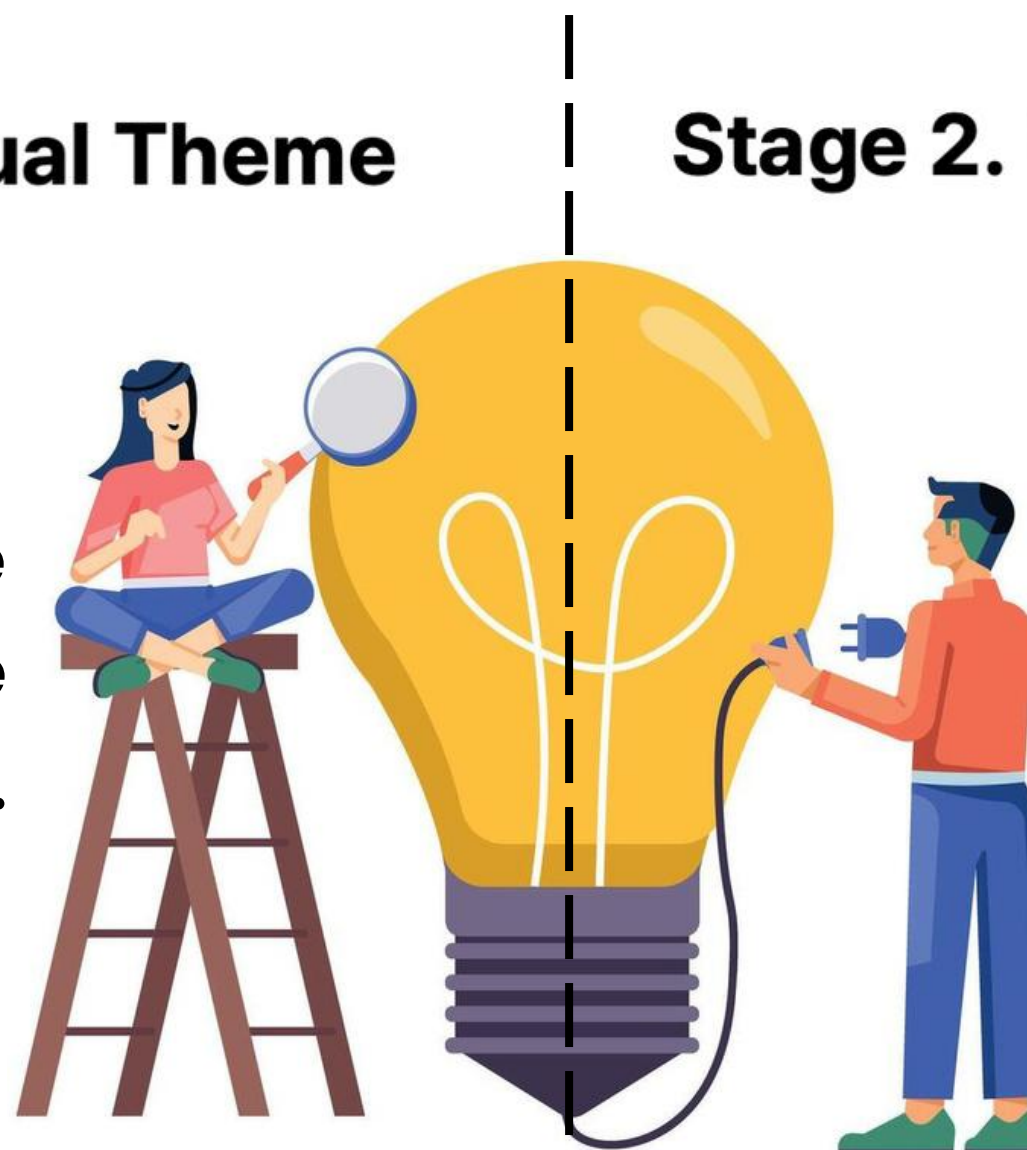
<u>4–step dialogue</u> assisted artists easily adjust scene while understanding creative intent.

27

FINDINGS

Friedman Test for non-parametric testing
*Benjamini-Hochberg for multiple comparison*

Experience on Making Music Video Scene

MVPrompt had the highest average score in five items
: Discovery, Worth, Ideation, Clarity, and Satisfaction.

MVPROMPT: BUILDING MUSIC-VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE-EN-SCÈNE

FINDINGS – **EXPERIENCE ON MAKING MUSIC VIDEO SCENE**

Balancing Guidance and Creative Freedom in Ideation
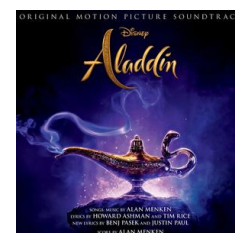
Aesthetic Theme
: **Cyberpunk**

P1's scene generated with MVPrompt
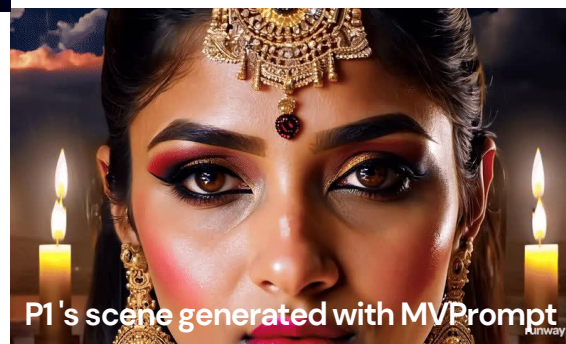
Refining mise-en-scène elements through dialogue supported idea generation, but MVPrompt's combined approach was more effective.

*"The suggest felt like **a non-traditional take**, so I accepted the aesthetic theme and made my scene through dialogue."* – P1

However, we also found that participants felt MVPrompt's theme image suggestion and dialogue features limited their freedom in conversation.
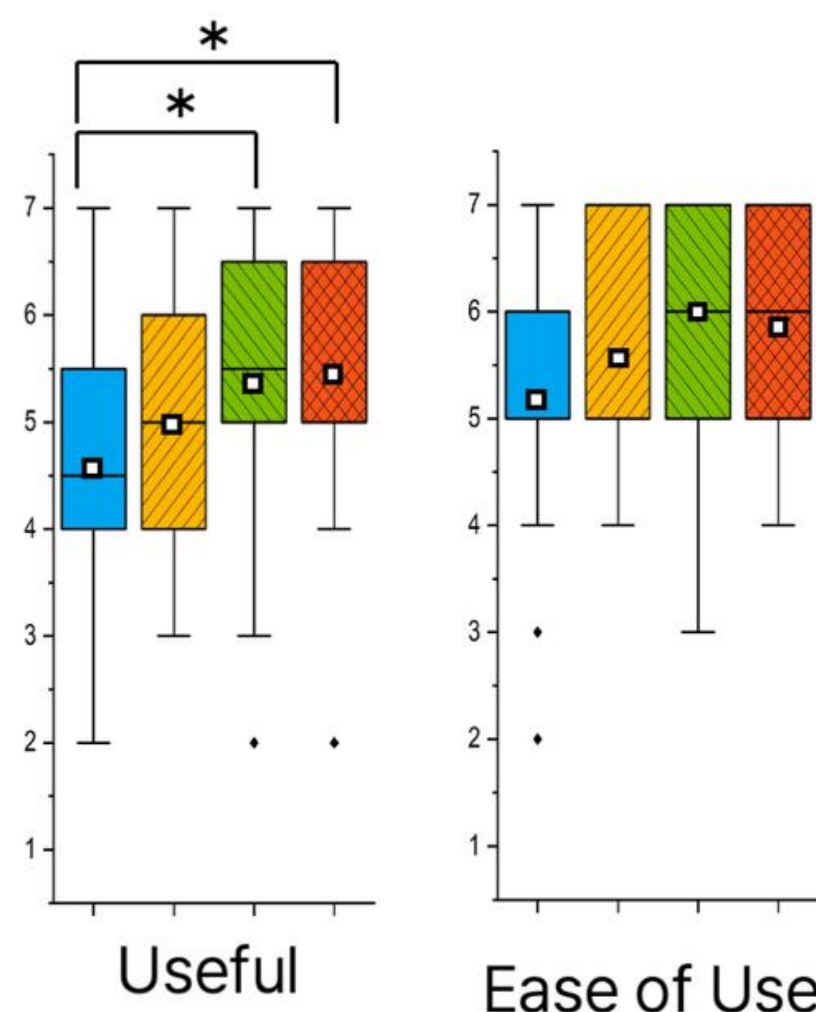
*"Although it's convenient to receive guidance on prompts, **I find it challenging to lead the conversation my way or reflect my creative elements.**"* – P18

MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE
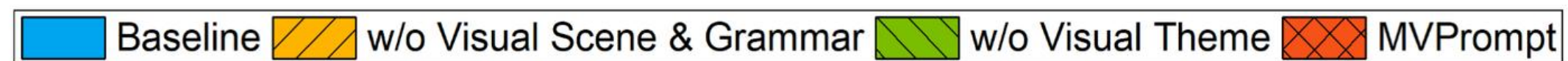
Useful and Easy Mise-en-scène Guidance

While MVPrompt and w/o Visual Theme was significantly more useful than the Baseline, w/o Visual Theme was the easiest system among the four systems to use.

*"The aspects related to **cinematography are diverse and professional**,*

so the system's suggestions were helpful." – P4

Participants found it challenging to use the systems without guided dialogue – feeling it less useful to make music video scenes.
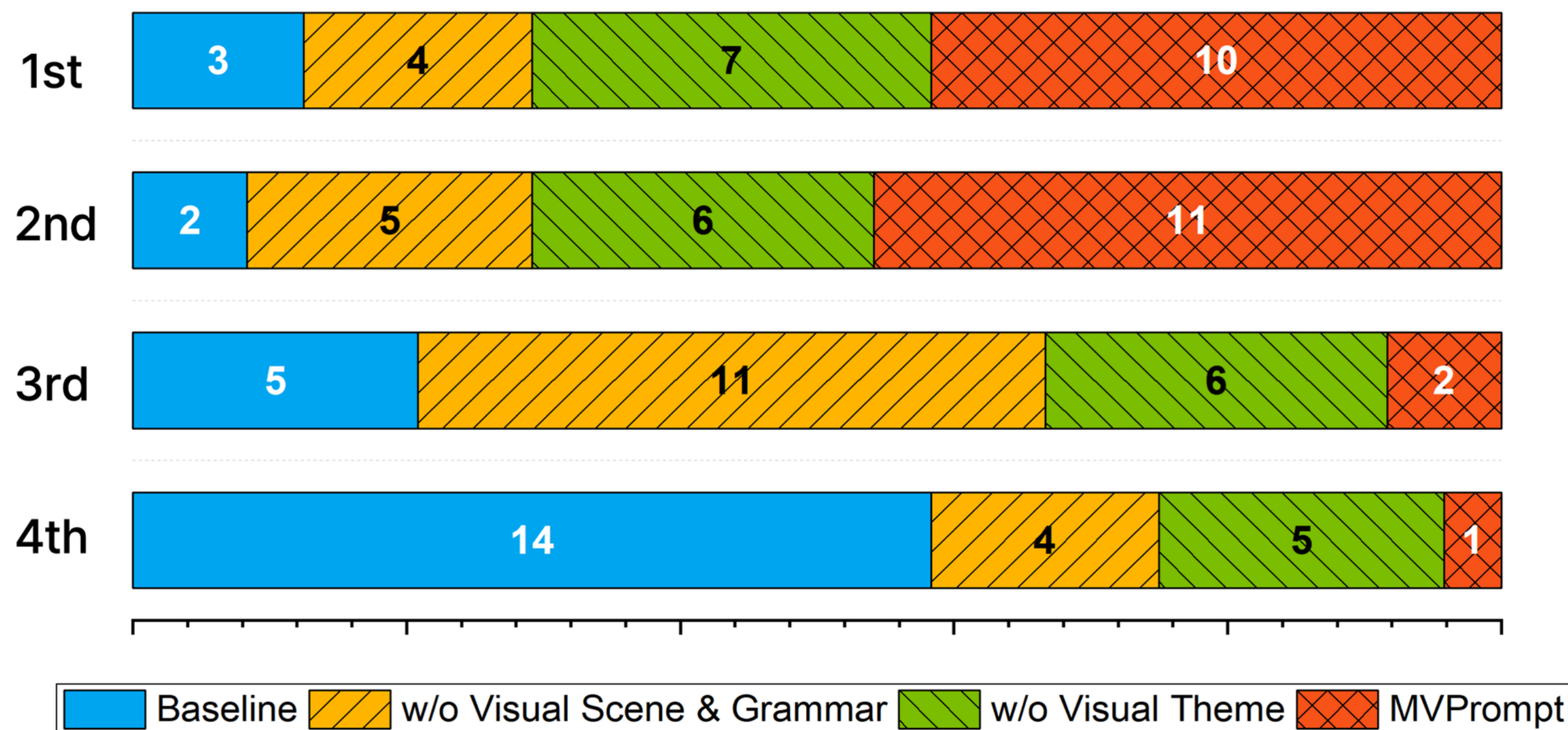
"It provided a bulk list of elements to include in the prompt,

***making it hard to determine which ones were important*.** *"* – P23

Legend: Baseline | w/o Visual Scene & Grammar | w/o Visual Theme | MVPrompt

MVPROMPT: BUILDING MUSIC-VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE-EN-SCÈNE

DISCUSSION

01

Transferring Expert's Knowledge to AI Artists

02

Dynamics between Artists' Creativity and AI Recommendations

03

Limitations & Future Work

MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE

DISCUSSION #1

Transferring Expert's Knowledge to AI Artists

Because mise-en-scène broadly applies to the aesthetic composition of video content [21, 31, 57], our framework can be extended beyond music videos to include film, dramas, and ads.

Meet Meta Movie Gen, META


Battalion, Gen AI Director Dave Clark

Whipnotic AD Competition 2nd Place, Justin Tolley

The method of providing creative context may vary by the type of the content, but general structure remains applicable across various types of video production.

MVPROMPT: BUILDING MUSIC-VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE-EN-SCÈNE

Dynamics between Artists' Creativity and AI Recommendations

Existing research suggests that artists tend to accept AI recommendations because AI tools enhance productivity in the creative process and provide novel inspiration [7, 90].
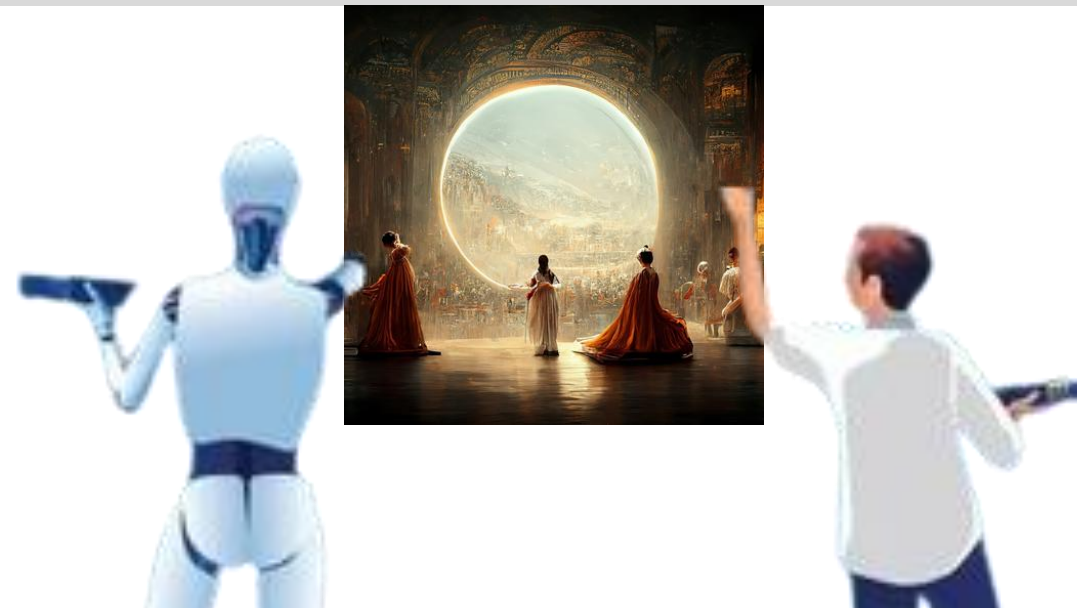
However, our findings raised the importance of designing AI tools that support artist in expanding unique creative visions without overshadowing their autonomy.

MVPROMPT: BUILDING MUSIC–VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE–EN–SCÈNE

Limitations & Future Work

- As each AI artist created eight scenes within 120 minutes, future research could explore a broader range of music to better understand the system's capabilities.

- As we focused on text2video workflows, future research should investigate working with image–based video production to provide a broader perspective on its potential use cases.

CONCLUSION
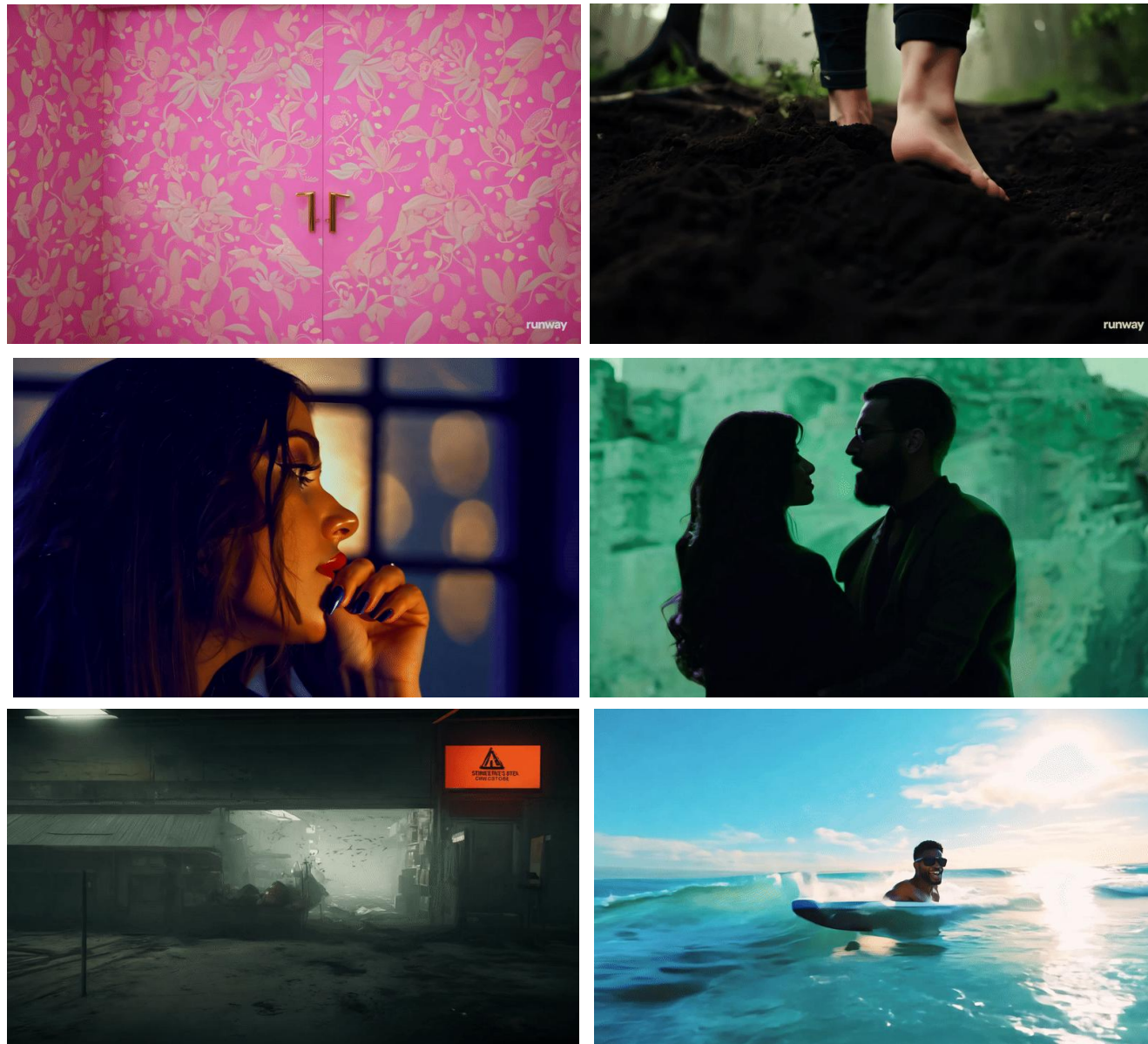
- We introduce MVPrompt, a tool that helps AI artists create music video scenes by structuring mise-en-scène into themes, scenes, and grammar.

- Through a user study with 24 artists, we found that integrating music-visual mise-en-scène enhances both creativity and the overall experience of video production.

MVPROMPT: BUILDING MUSIC-VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE-EN-SCÈNE

Thank you for your attention!

ChungHa Lee
Presenter

✉ chunghalee@gm.gist.ac.kr
🏠 chunghalee.com

DaeHo Lee

✉ leedaeho@gm.gist.ac.kr
🏠 lee-daeho.github.io

Jin-Hyuk Hong*

✉ jh7.hong@gist.ac.kr
🏠 sci.gist.ac.kr

**GWANGJU INSTITUTE OF SCIENCE AND TECHNOLOGY
SOFT COMPUTING & INTERACTION LAB**

MVPROMPT: BUILDING MUSIC-VISUAL PROMPTS
FOR AI ARTISTS TO CRAFT MUSIC VIDEO MISE-EN-SCÈNE