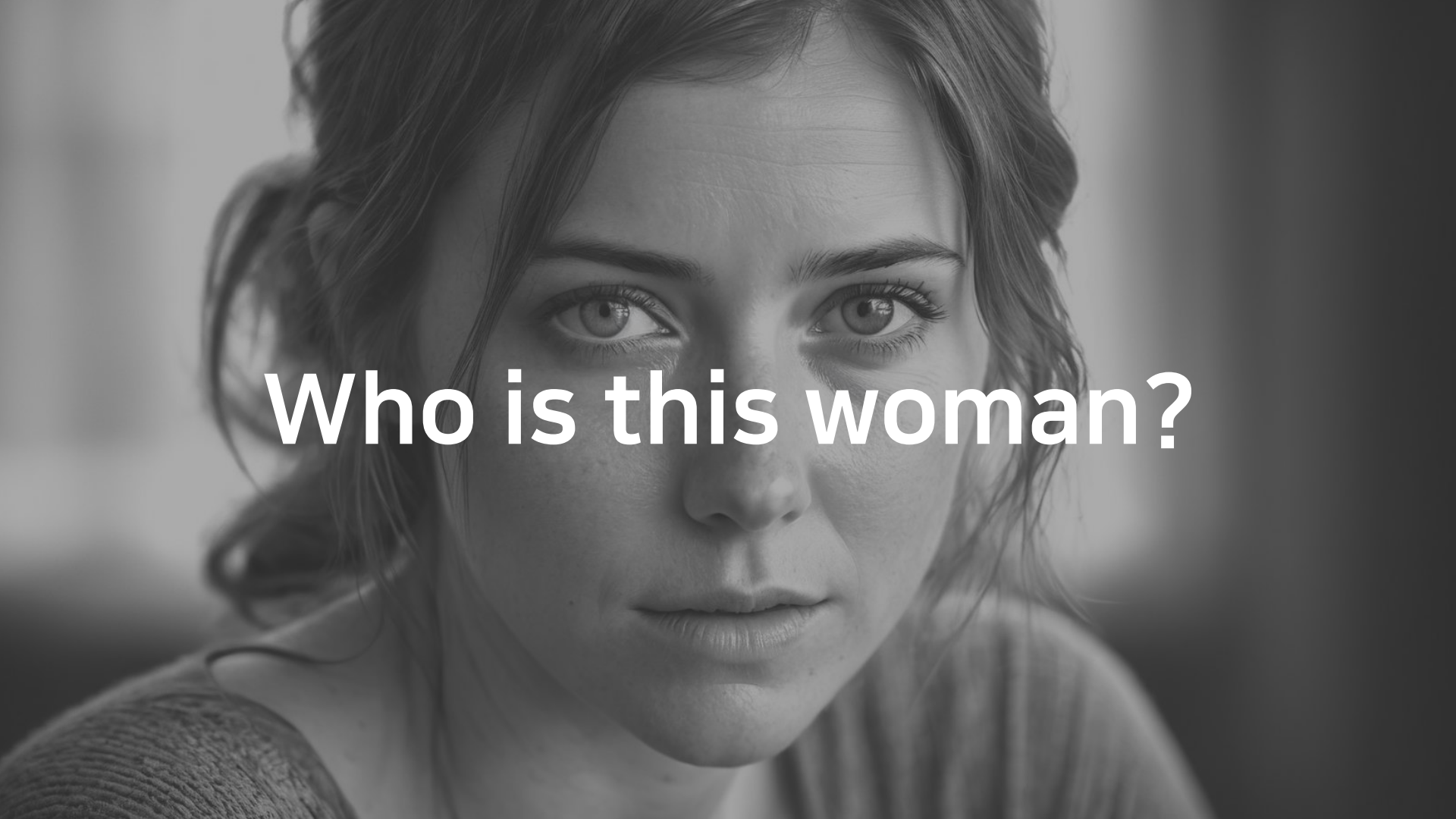
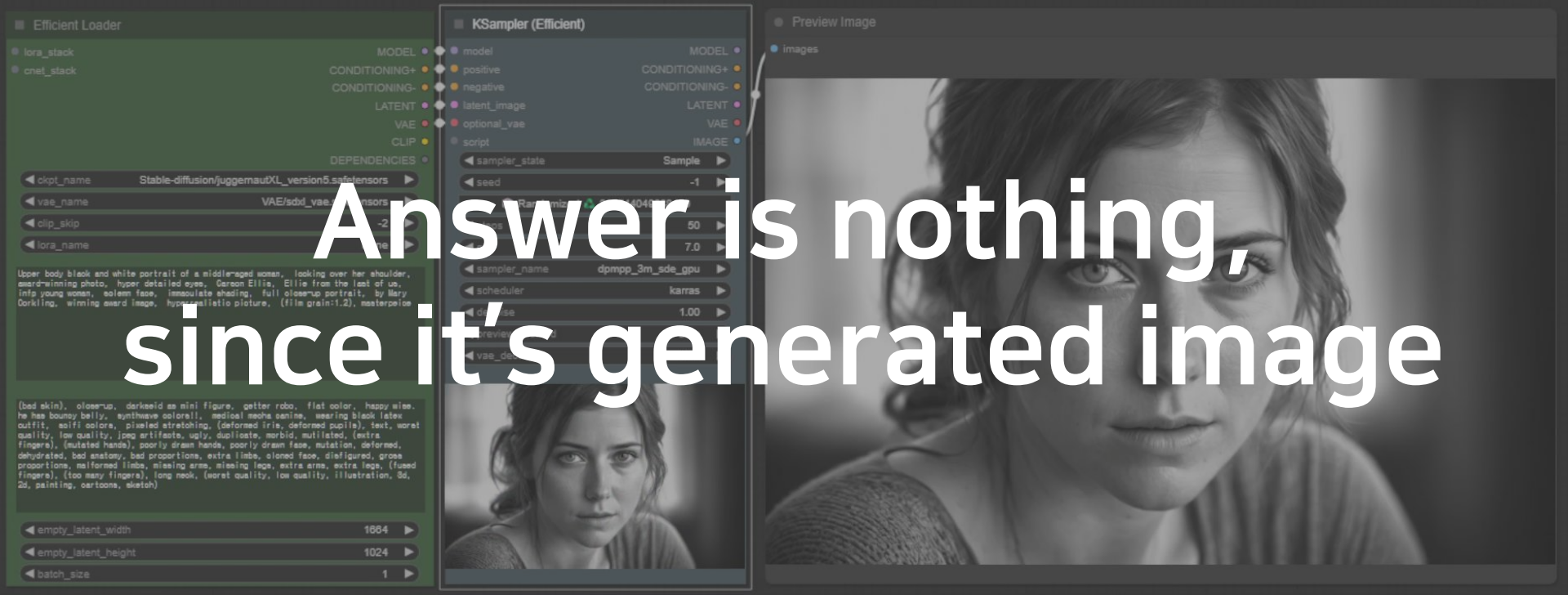
# Designing Interfaces for Text-To-Image Prompt Engineering Using Stable Diffusion Models: A Human-AI Interaction Approach

Kim, Seonuk; Ko, Taeyoung; Kwon, Yousang; Lee, Kyungho


Who is this woman?




# Answer is nothing, since it's generated image

# Evolution of AI-generative Image

**Prompt;** tag bundle that user want to generate

Answer is nothing.

**Negative Prompt;** tag bundle that user don't want to generate

# Evolution of AI-generative Image

**Prompt;** tag bundle that user want to generate

Upper body black and white portrait of a middle-aged woman, looking over her shoulder, award-winning photo, hyper detailed eyes, Carson Ellis, Ellie from the last of us, infp young woman, solemn face, immaculate shading, full close-up portrait, by Mary Corkling, winning award image, hyperrealistic picture, (film grain:1.6)

**Negative Prompt;** tag bundle that user don't want to generate

(bad skin), close-up, darkseid as mini figure, getter robo, flat color, happy wise. he has bouncy belly, synthwave colors!!, medical mecha canine, wearing black latex outfit, scifi colors, pixeled stretching

## Efficient Loader

lora_stack
cnet_stack

MODEL
CONDITIONING+
CONDITIONING-
LATENT
VAE
CLIP
DEPENDENCIES

ckpt_name: Stable-diffusion/juggernautXL_version5.safetensors
vae_name: VAE/sdxl_vae.safetensors
clip_skip: -2
lora_name: None

Upper body black and white portrait of a middle-aged woman, looking over her shoulder, award-winning photo, hyper detailed eyes, Carson Ellis, Ellie from the last of us, infp young woman, solemn face, immaculate shading, full close-up portrait, by Mary Corkling, winning award image, hyperrealistic picture, (film grain:1.2), masterpeice

(bad skin), close-up, darkseid as mini figure, getter robo, flat color, happy wise. he has bouncy belly, synthwave coloroll, medical mecha canine, wearing black latex outfit, sci-fi colors, pixelsd stretching, (deformed iris, deformed pupils), text, worst quality, low quality, jpeg artifacts, ugly, duplicate, morbid, mutilated, (extra fingers), (mutated hands), poorly drawn hands, poorly drawn face, mutation, deformed, dehydrated, bad anatomy, bad proportions, extra limbs, cloned face, disfigured, gross proportions, malformed limbs, missing arms, missing legs, extra arms, extra legs, (fused fingers), (too many fingers), long neck, (worst quality, low quality, illustration, 3d, 2d, painting, cartoons, sketch)

empty_latent_width: 1664
empty_latent_height: 1024
batch_size: 1

## KSampler (Efficient)

model
positive
negative
latent_image
optional_vae
script

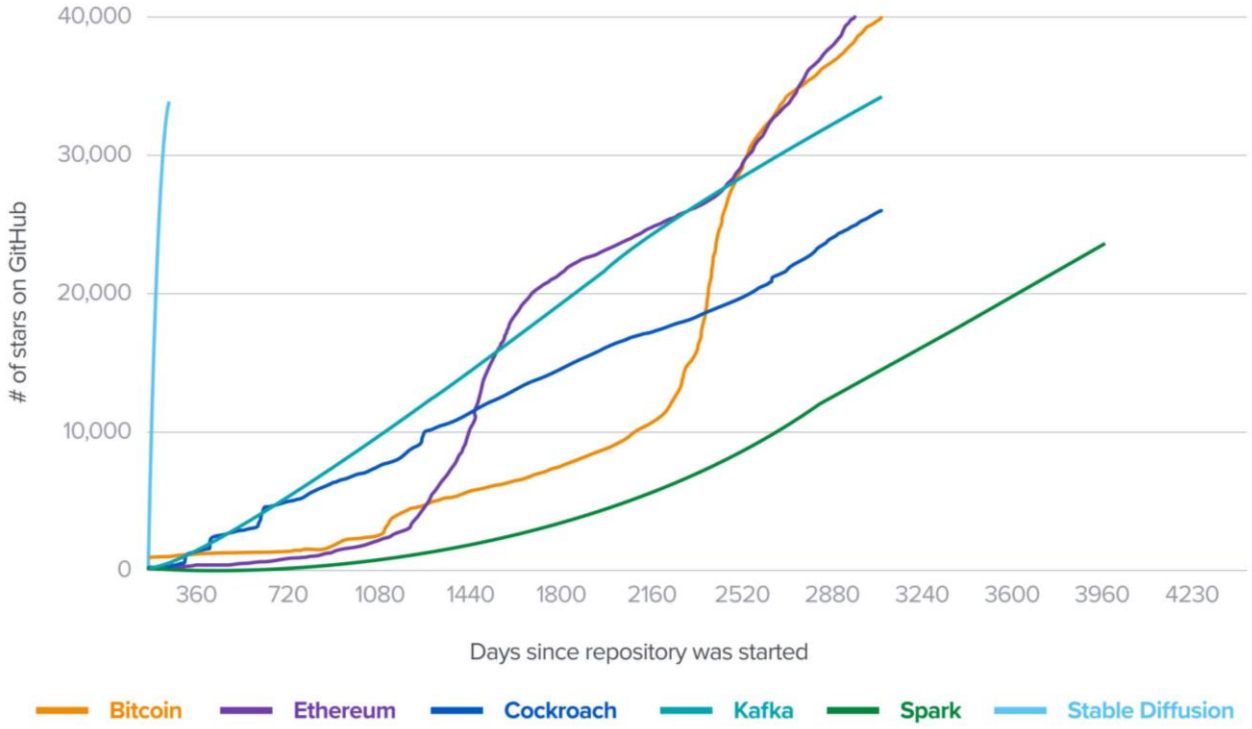
MODEL
CONDITIONING+
CONDITIONING-
LATENT
VAE
IMAGE

sampler_state: Sample
seed: -1
Randomize / ♻ 212334913849004
steps: 50
cfg: 7.0
sampler_name: dpmpp_3m_sde_gpu
scheduler: karras
denoise: 1.00
preview_method: auto
vae_decode: true

## Preview Image

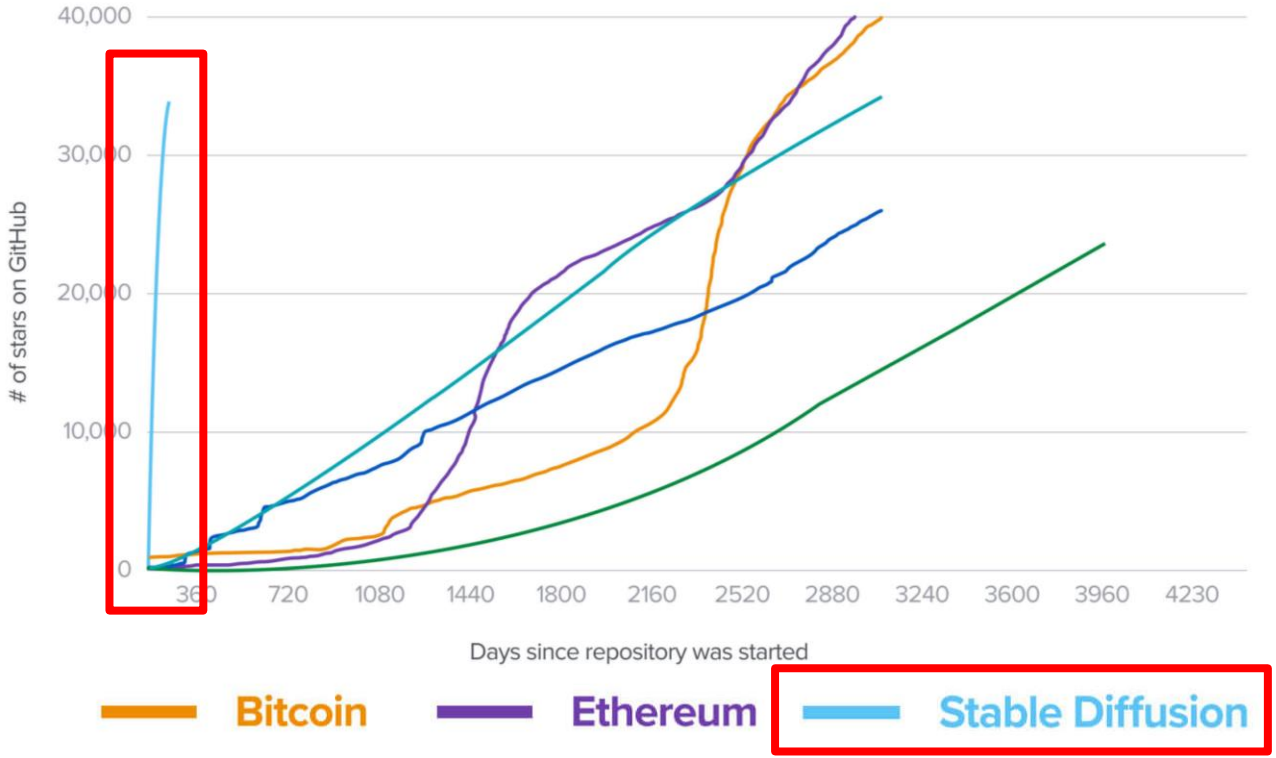
images

# Generative AI



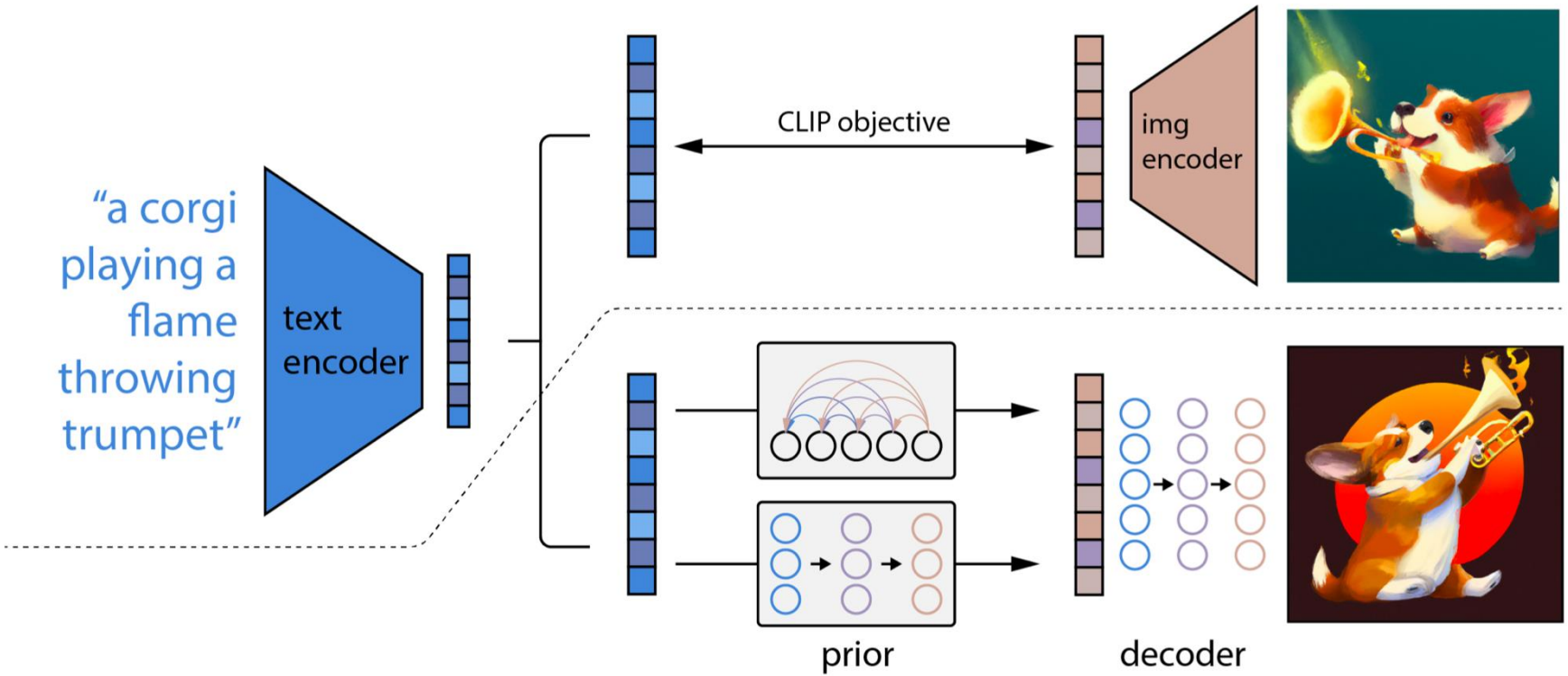

Source: A16z and Github

# Generative AI, lots of focus and improving speed!



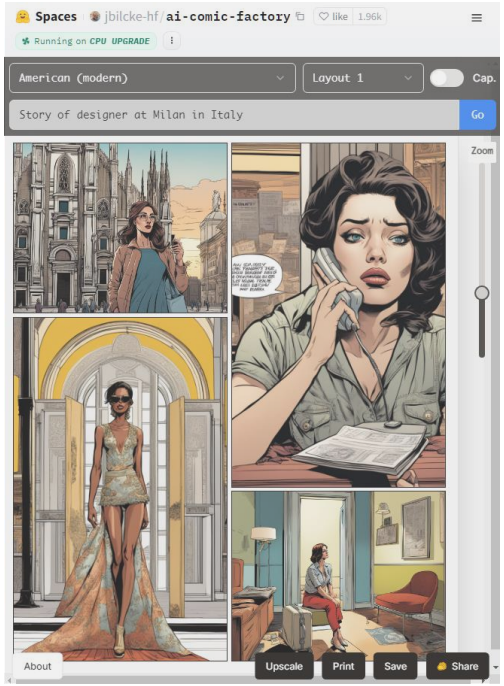

Source: A16z and Github

# Generative AI: Principles



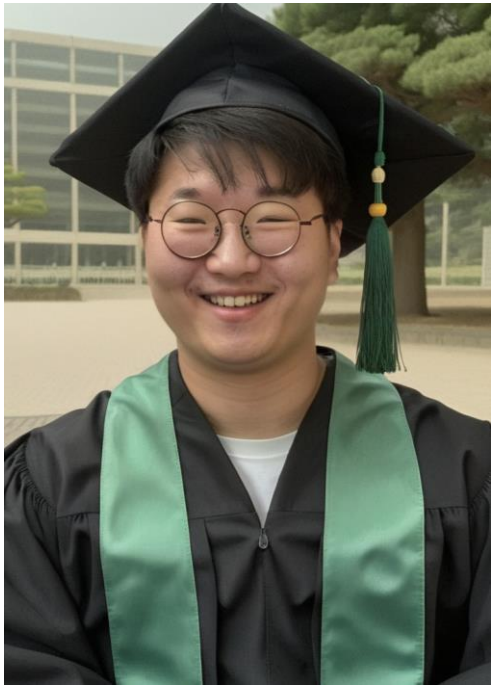

Source: Ramesh et al. 2022, Hierarchical Text-Conditional Image Generation with CLIP Latents

# Possibilities of AI-based Image for Content Creation




Create contents

Idea-to-Result



Make wish come true(?)

Source: ai-comic-factory, civitai

# Problems of AI-based Image for Content Creation

**Input**
chicken and pizza are on the table

**Output**

1. Technical complexity of AI systems
2. Unpredictable system behaviors
3. AI is “a **new** and difficult design material”

(Yang, et al., 2020)

Source: Yang, et al., 2020, Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design

# Challenges of AI-based Image for Content Creation

**Input**

chicken and pizza are on the table

**Output**

**Improving**

- Text Prompt
- Models
- Exten-sions
- External Services
- UI

# Challenges of AI-based Image for Content Creation

**Input**

chicken and pizza are on the table

**Output**

**Improving**

- Text Prompt
- Models
- Exten-sions
- External Services
- UI

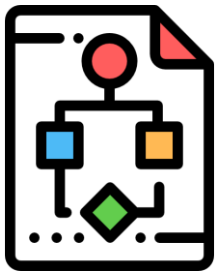
# Research Questions

RQ1: [**Observation**] What are the distinguished behavioral patterns?

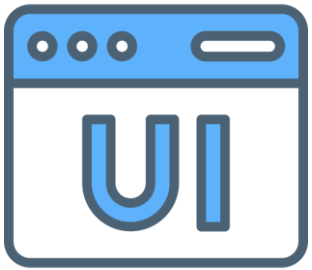
RQ2: [**Analysis**] What are relationship between behaviors and others?

RQ3: [**Enhancement**] How to enhance the creative process with design?

**RQ1: Observation**



**RQ2: Analysis**




**RQ3: Enhancement**

Icon credit: Flaticon

# Subjects

**Arcalive AI Art channel**

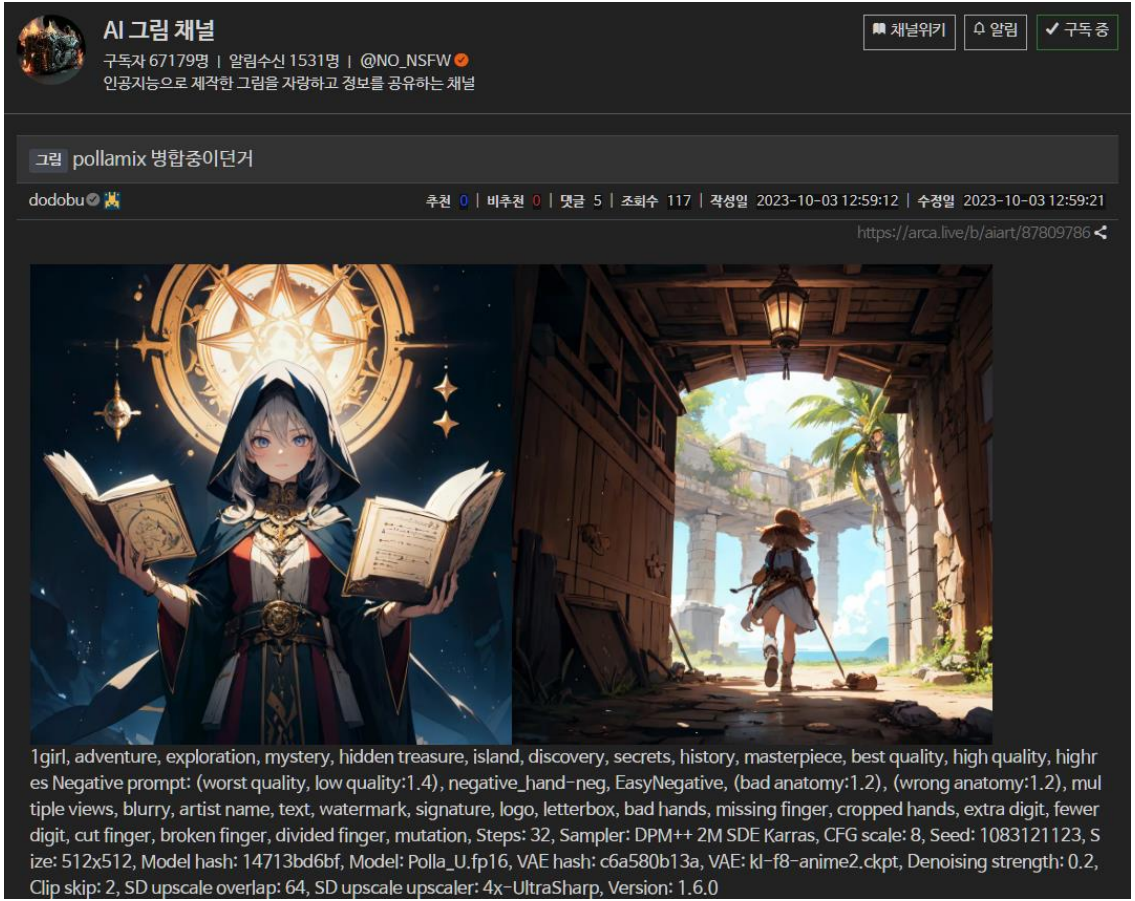
Korean internet forum

on image generative AI

Launched on Sept. 8, 2022

On Oct. 3, 2023

- 67,178 Subscribers
- 179,353 Posts

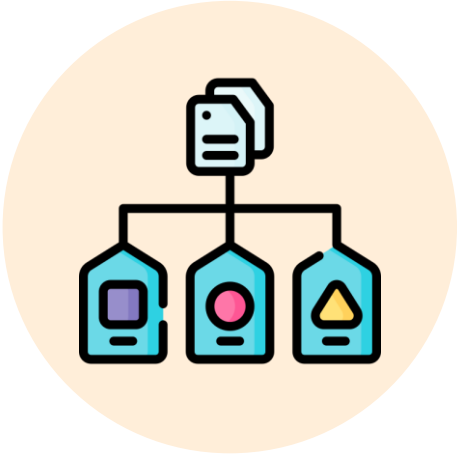



Source: Arcalive AI Art channel

# Method

We selected mixed-methods approach:

**Case Studies & Thematic Analysis & Classification**

Icon credit: Flaticon

# Method: Case Studies

**[Subjects]**
**All 1,321 posts** until Nov. 30, 2022
(10 or more recommendation)

**Review**

Read and reviewed posts one by one

**Summarization**

Summarized common and specific user behaviors by each post

**Information**

Got background and user experiences of image generative AI

Icon credit: Flaticon

# Method: Thematic Analysis & Classification

**[Subjects]**
**All 1,068 posts** until Nov. 14, 2022
(10 or more recommendation)

## Affinity Diagramming

Define each group as **behaviors**
(based on summaries of each post)

**[Subjects]**
**All 1,321 posts** until Nov. 30, 2022
(10 or more recommendation)

## Thematic Analysis

1. Group **behaviors** into **types**
2. Classified posts into types and got **statistics**
3. Analyze quantitative and qualitative

## Classification

1. Classified extensions and services into types
2. Analyze quantitative and qualitative

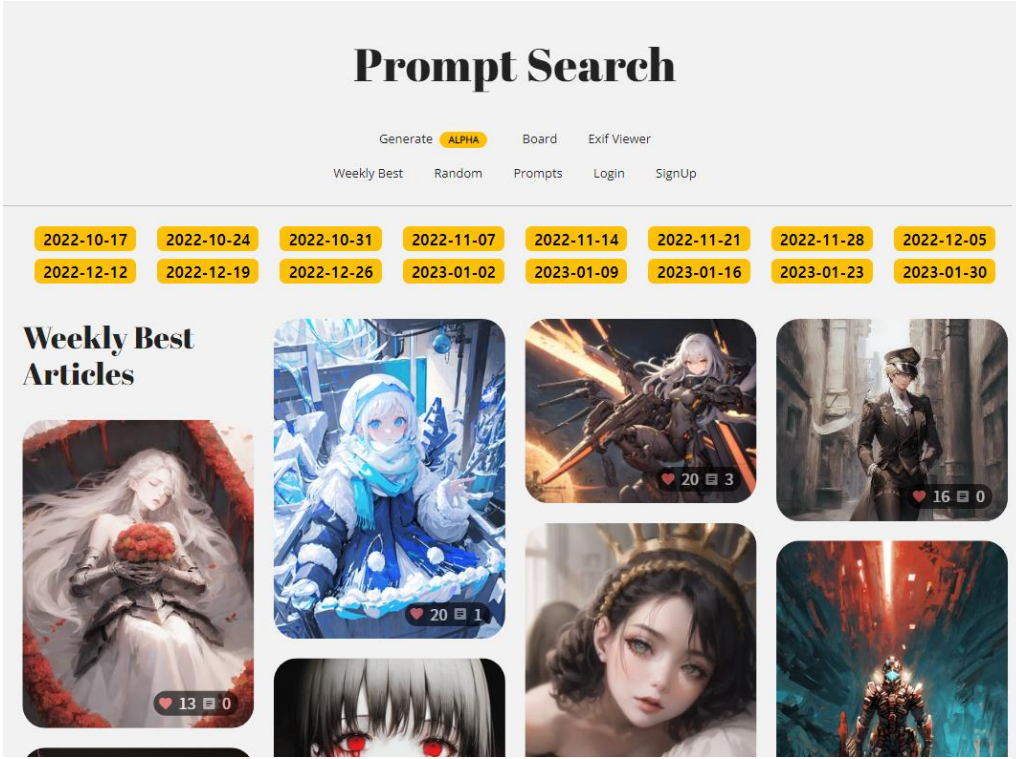
Icon credit: Flaticon

# What are the distinguished behavioral patterns?

| 8 Types | | 25 Behaviors* |
|---|---|---|
| Consumption | | 3 behaviors (856 posts, 64.8%) |
| Exploration | Semantic | 4 behaviors (473 posts, 35.8%) |
| | Symbolic | 3 behaviors (107 posts, 8.1%) |
| Production | | 3 behaviors (892 posts, 67.5%) |
| Summarization | | 2 behaviors (160 posts, 12.1%) |
| Development | | 2 behaviors (266 posts, 20.1%) |
| Hypothesization | | 2 behaviors (22 posts, 1.7%) |
| Validation | | 4 behaviors (137 posts, 10.4%) |
| Innovation | | 2 behaviors (60 posts, 4.5%) |

*More detailed results of the thematic analysis are available in the Appendix

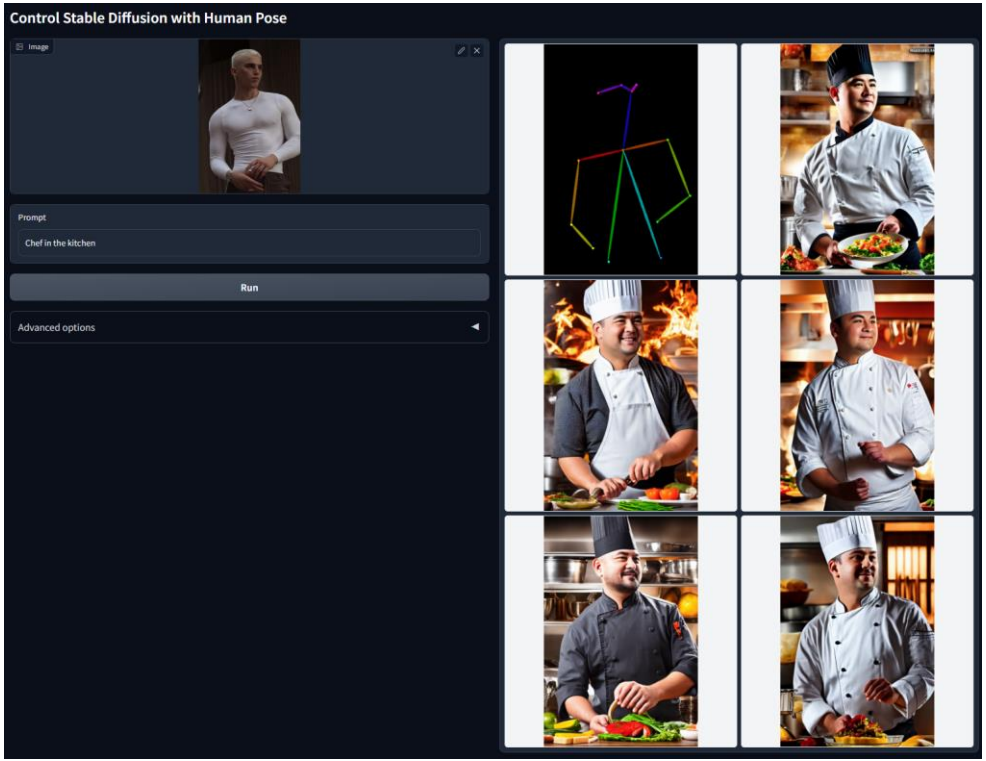
# What are relationship between behaviors and others?




| | |
|---|---|
| parameters | masterpiece, best quality, illustration, fantasy, dark basement indoors, bookshelf, potion, gem, bottle, animal skull, plant, candle, yuri, 2girls, alchemist, skinny, grey hair, yellow eyes, hooded cloak, robe NAI WEBUI Original |
| negative_prompt | (worst quality, low quality, normal quality, low resolution, lowres:1.4), 3d, blurry, text, artist name, signature, multiple views, pointy ears NAI WEBUI Original |
| steps | 28 |
| sampler | DPM++ SDE Karras |

Source: Prompt Search

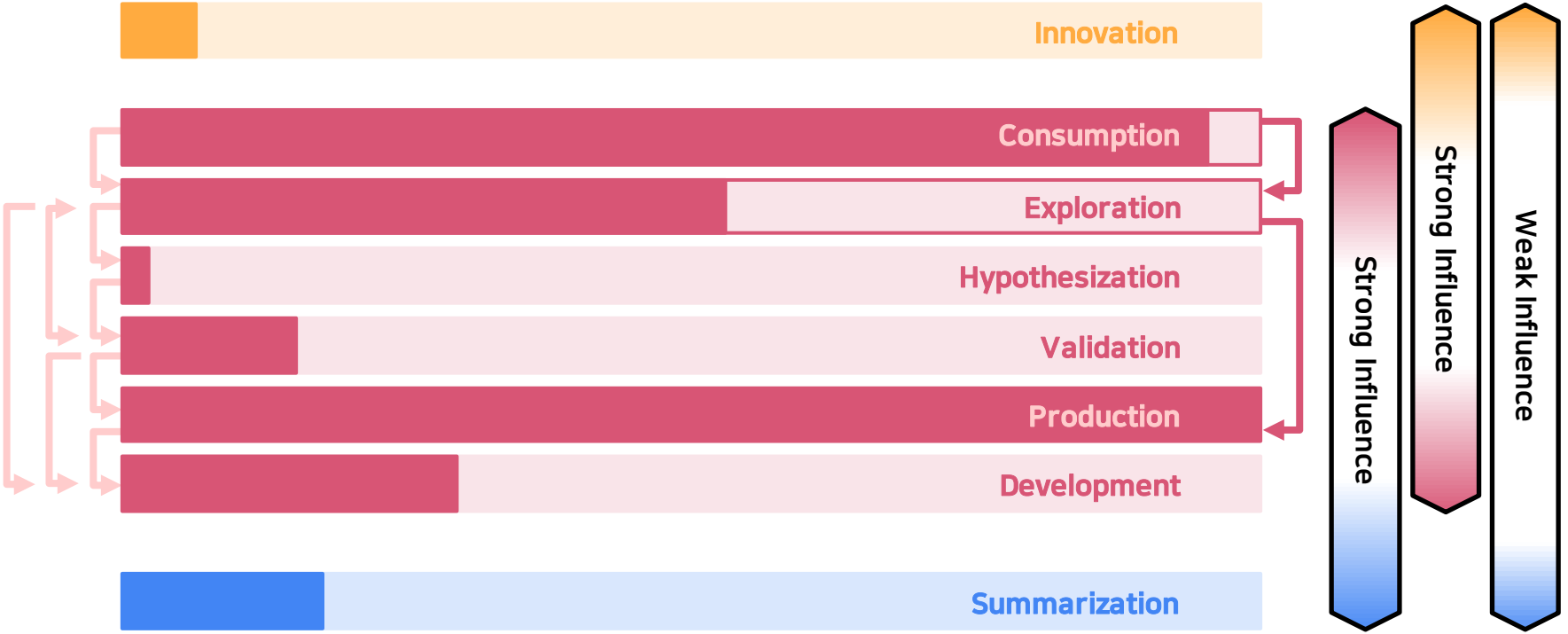
# What are relationship between behaviors and others?




| 8 Types | | Extensions | Services |
|---|---|---|---|
| Consumption | | 7/7 | 7/11 |
| Exploration | Semantic | 7/7 | 8/11 |
| | Symbolic | 3/7 | 6/11 |
| Production | | 4/7 | 9/11 |
| Summarization | | 0/7 | 5/11 |
| Development | | 5/7 | 2/11 |
| Hypothesization | | 7/7 | 5/11 |
| Validation | | 7/7 | 5/11 |
| Innovation | | 4/7 | 0/11 |

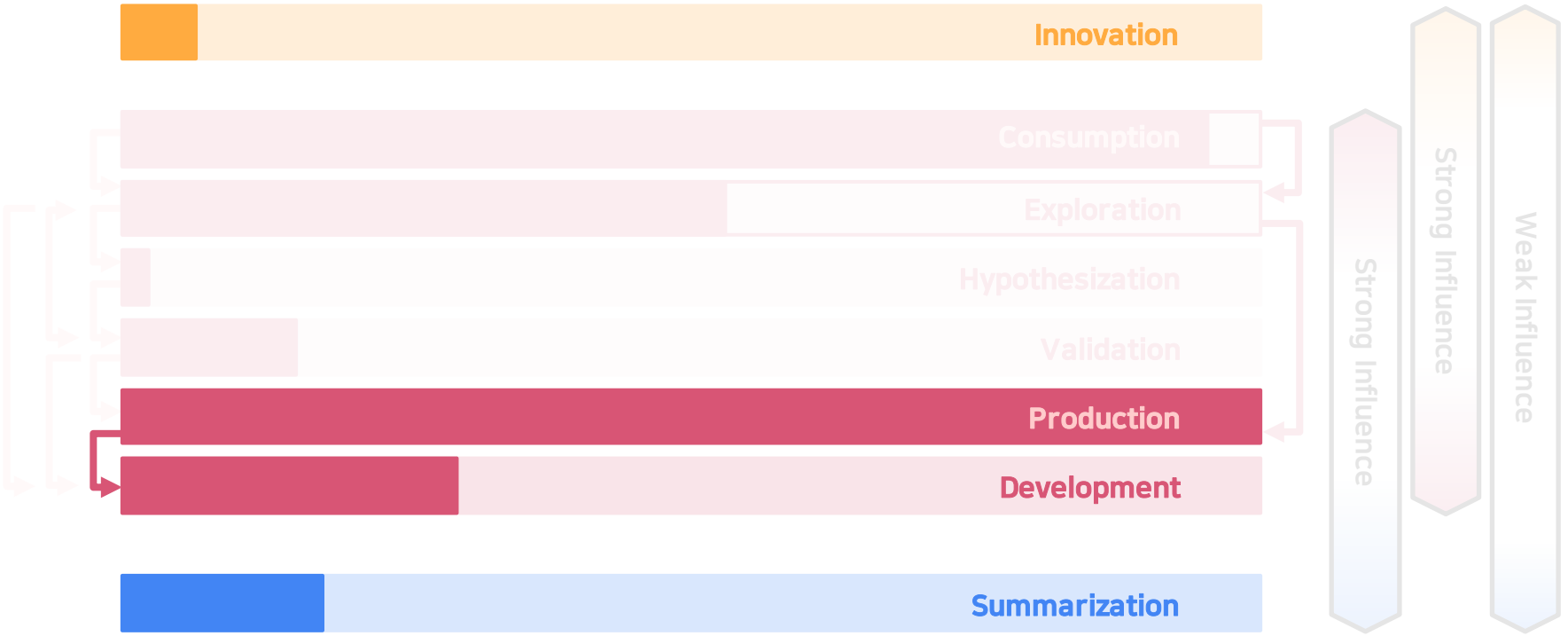
Source: Zhang. et al., 2023, Adding Conditional Control to Text-to-Image Diffusion Models

# What are relationship between behaviors and others?

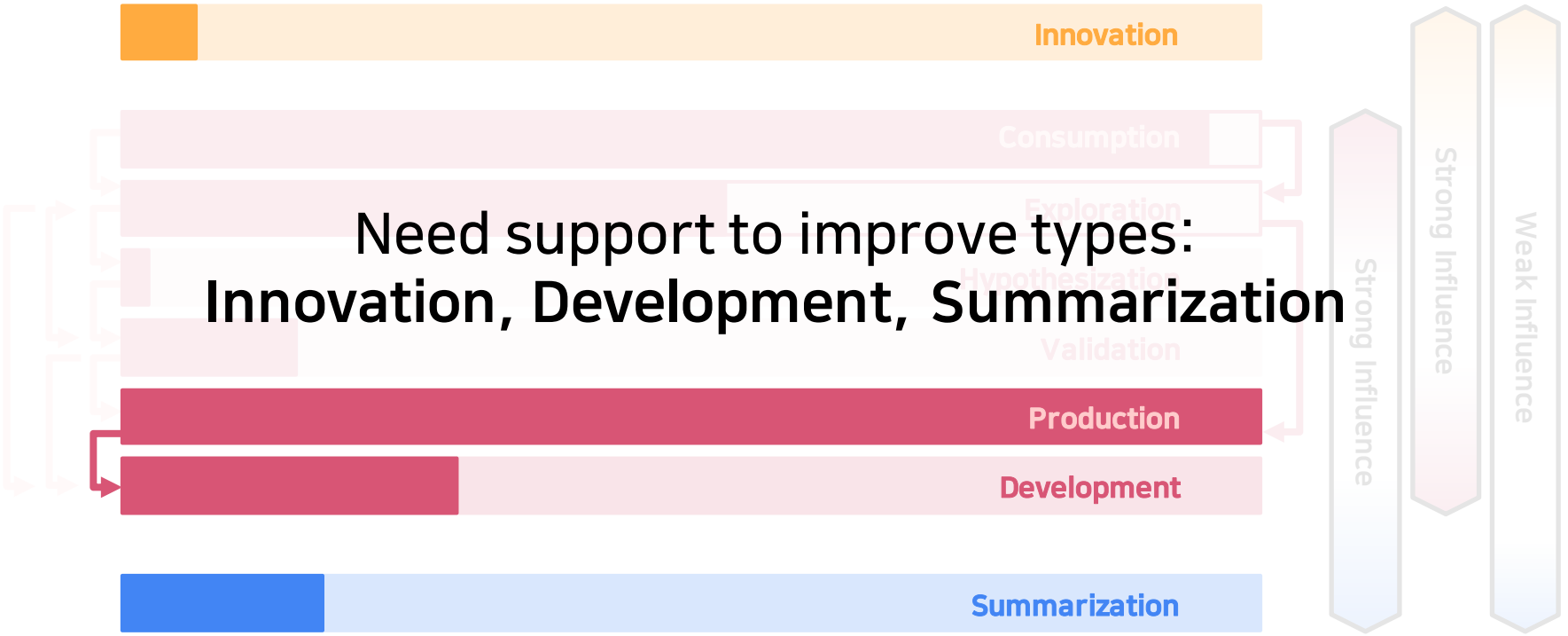


# How to enhance the creative process with design?

# How to enhance the creative process with design?



# How to enhance the creative process with design?




Need support to improve types:
**Innovation, Development, Summarization**

# Discussion: Suggest design considerations

#1 User interface with **social networks (SNS)**

#2 Build interaction-based **recommendation systems**

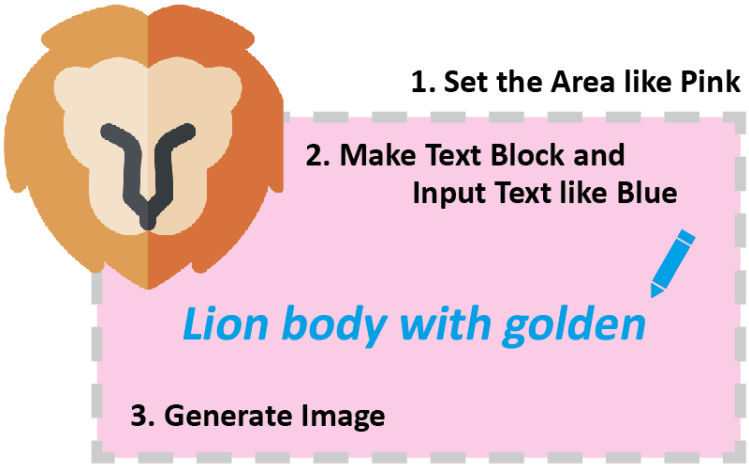
#3 Improve the **digital drawing interface**

**#4 Design intuitive style-based and object-based interface**

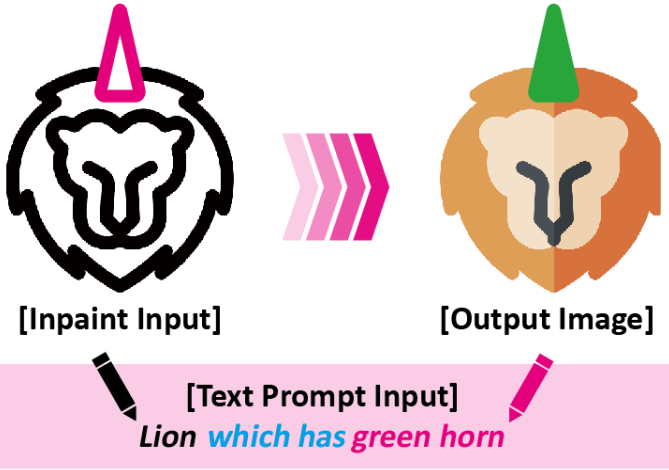
# Discussion: Suggest design considerations

## #4 Design intuitive style-based and object-based interface

### Text-on-Image



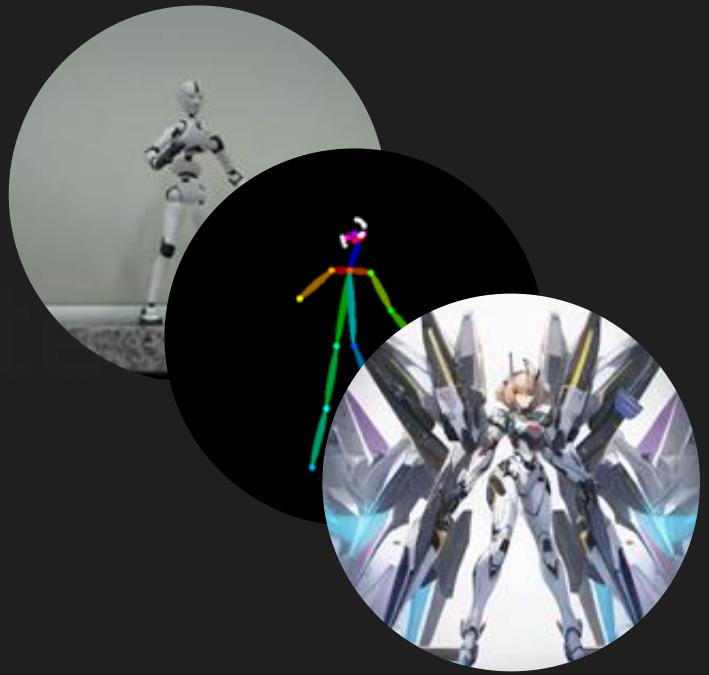

### Color-Coupling



# Limitations and Future Works

Limited attempts to develop and control experiment interface to improve UX/UI

w/ Tangible User Interface



# Thank you.


We hope our research can help more creators have
more effective and desired image generation processes

2023 IASDR Congress

Designing Interfaces for Text-To-Image Prompt Engineering Using Stable Diffusion Models: A Human-AI Interaction Approach

Seonuk Kim, d02reams@unist.ac.kr
Presenter, Expressive Computing Lab., UNIST

Taeyoung Ko, tyk0506@unist.ac.kr
Expressive Computing Lab., UNIST

Yousang Kwon, yk7244@unist.ac.kr
Expressive Computing Lab., UNIST

Kyungho Lee, kyungho@unist.ac.kr
Assistant Professor, Expressive Computing Lab., UNIST

DESIGN DEPT. | UNIST

# Appendix. Results of the thematic analysis

| Types | Approach | | Behaviors | # (%) of posts |
|---|---|---|---|---|
| Consumption | | 1 | Mimic the prompts. | 856 (64.8 %) |
| | | 2 | Ask questions. | |
| | | 3 | Search for information. | |
| Exploration | Semantic | 4 | Explore appropriate tag names to input in terms of the semantic approach. | 473 (35.8 %) |
| | | 5 | Determine if combinations of tags are good or bad for image generation (e.g., some components were not generated by stronger semantic tags.) | |
| | | 6 | Input substituted tags that share characteristics to generate intended components. | |
| | | 7 | Input tags with narrow semantics to generate predictable components. | |
| | Symbolic | 8 | Input tags in terms of symbolic logical approach. (e.g., control the hair color by inputting weighted tags at the negative prompt.) | 107 (8.1 %) |
| | | 9 | Increase tag influence with repetitive input. | |
| | | 10 | Set weights to increase or decrease tag influence. | |
| Production | | 11 | Share prompts with generating images. | 892 (67.5 %) |
| | | 12 | Suggest general tag bunches to improve quality. | |
| | | 13 | Search and feed information outside of the forum. | |

# Appendix. Results of the thematic analysis

| Types | | Behaviors | # (%) of posts |
|---|---|---|---|
| Summarization | 14 | Summarize prompt-related information. | 160 (12.1 %) |
| | 15 | Define the main components of images and format of prompts: tag ordering. | |
| Development | 16 | Tagging symbols with user-specified semantics via textual inversion; also called embedding in the forum. | 266 (20.1 %) |
| | 17 | Develop to provide insights for exploring prompts. (e.g., train new models, tune hypernetworks, and develop support services.) | |
| Hypothesization | 18 | Approach AI as works similarly to human thought processes. | 22 (1.7 %) |
| | 19 | Explore features of the model, including train data, to suggest suitable prompts. | |
| Validation | 20 | Validate hypotheses raised in the forum. | 137 (10.4 %) |
| | 21 | Determine whether combinations of tags are good or bad based on cases. | |
| | 22 | Screen prompts via controlled experiments. | |
| | 23 | Heuristic exploration such as trial and error. | |
| Innovation | 24 | Reorganize or redefine the main components of images. (e.g., input alpha transparency as a tag into the prompt.) | 60 (4.5 %) |
| | 25 | Input nonverbal symbols as new tags. (e.g., special symbols such as Unicode, emoji, etc.) | |