# Advanced Image Generation with Stable Diffusion 🖼️

Eason Suen - June 1, 2023

# 📝 Agenda

1. Stable Diffusion 101
2. Generate with Style
3. Generate with Pose
4. Generate with Subject
5. Q&A

# Stable Diffusion 101

# Survey: What do you use?

**Midjourney** | **Stable Diffusion** | **DALLE 2** | **Firefly** | **Other**

Most use **Midjourney!** Others include **Leonardo.ai**

# Stable Diffusion

Fast, versatile, open-source text-to-image model.

# Timeline

**Feb, 2021**

**[Paper] "Zero-Shot Text-to-Image Generation"**

**Author:** OpenAI

**Contribution**: The predecessor of **DALLE-2 (April, 2022)** by OpenAI, first-generation text-to-image AI based on diffusion model.

**Dec, 2021**

**[Paper] "High-Resolution Image Synthesis with Latent Diffusion Models"**

**Author:** Ludwig Maximilian University of Munich (LMU), Runaway ML

**Contribution:** Introduce latent diffusion model, which is smaller and faster.

**Sep, 2022**

**[Model] Stable Diffusion - V1**

**Author:** Stability AI, LMU, Eleuther AI

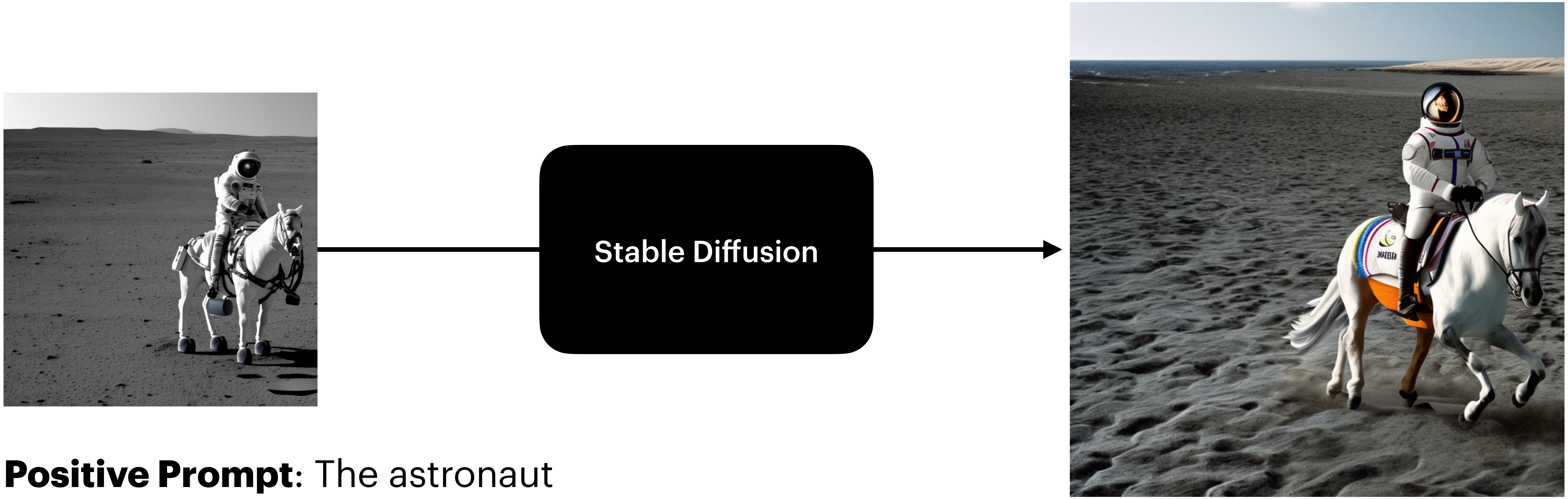
**Contribution**: Large open-source text-to-image diffusion model comparable to DALLE-2.

# Text

✍️"a photo of an astronaut riding a horse on mars"

Stable Diffusion

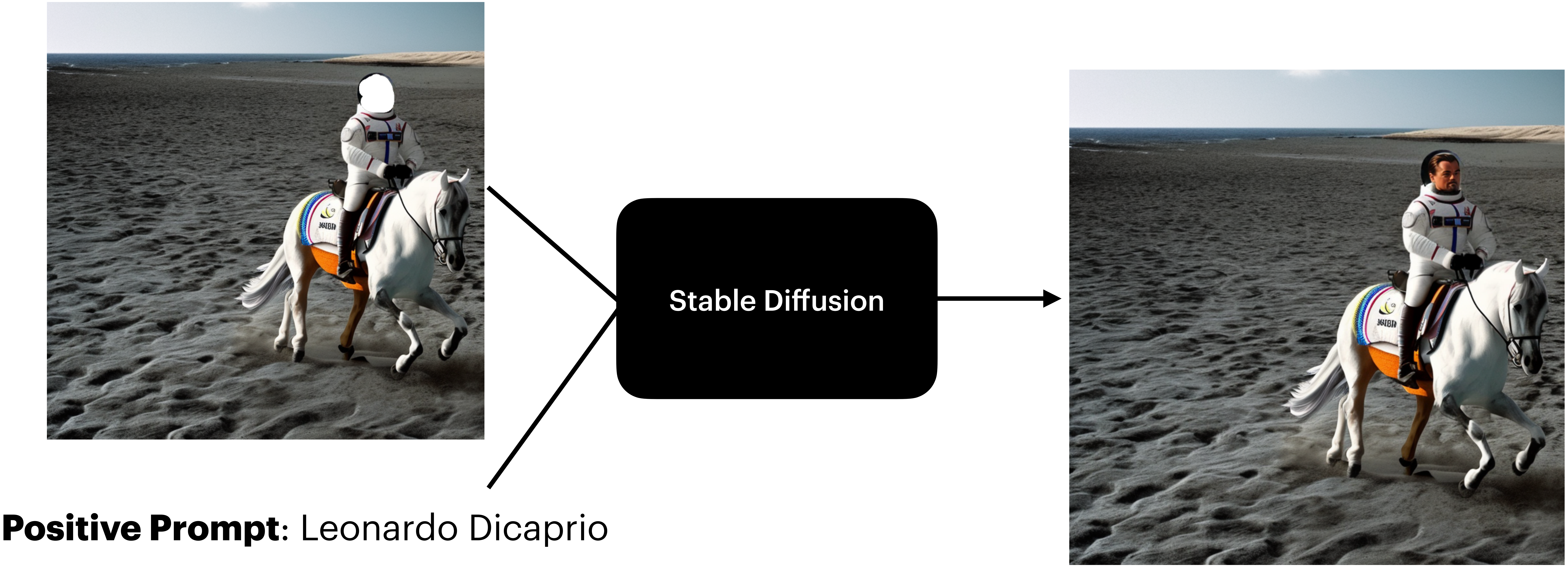
# Text + Image



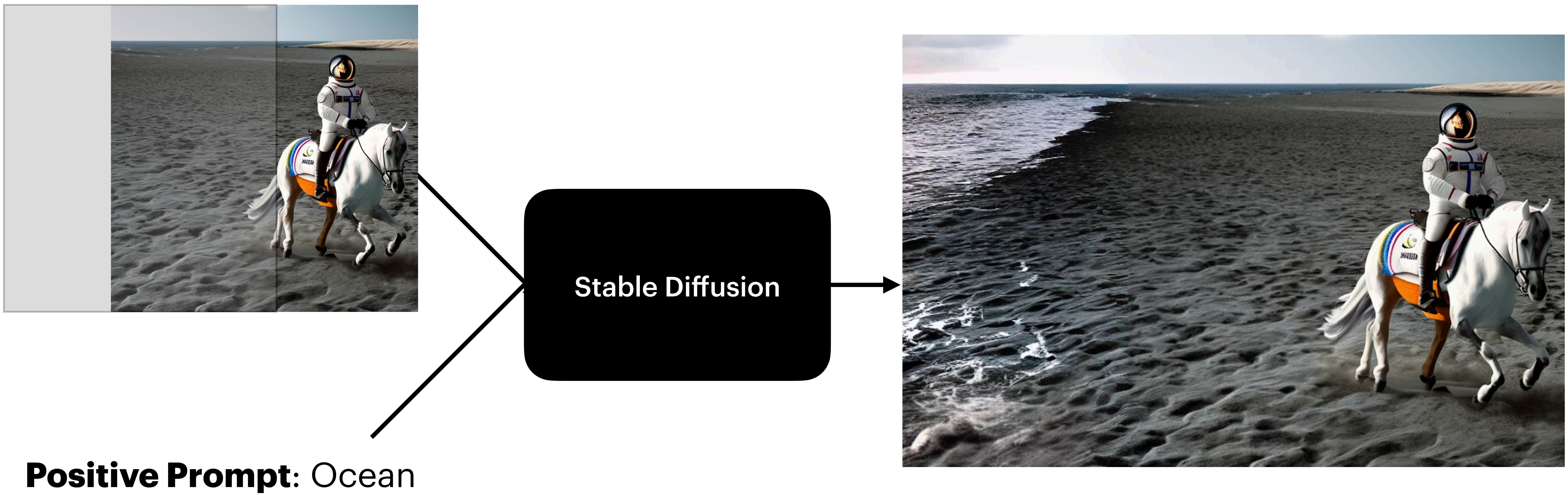

**Positive Prompt**: The astronaut riding a horse on a beach by the ocean, colorful

**Negative Prompt**: grey

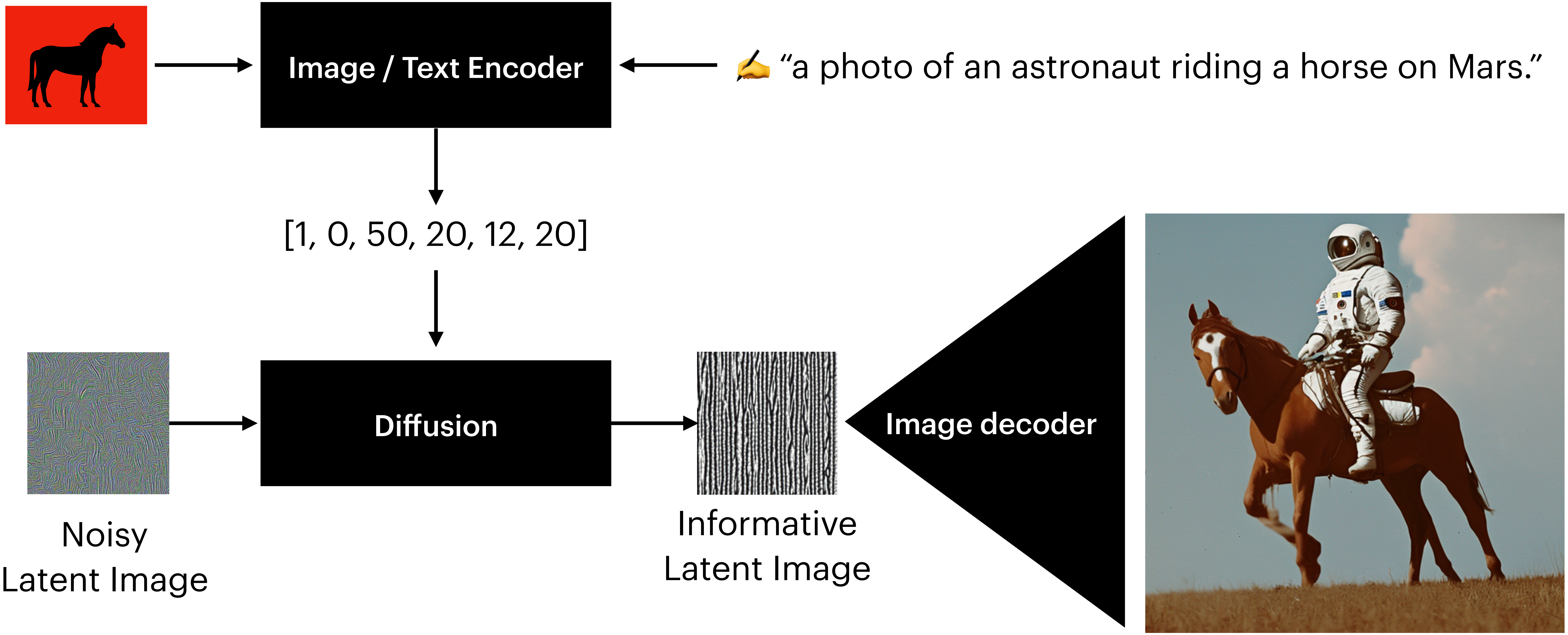
# In-Painting



# Out-Painting



# How does it works?



# Stable Diffusion: How To Use

**Use a Web App**: DreamStudio (StabilityAI), Getimg.ai

**Use API end-point**: StabilityAI, Runaway, Replicate, HuggingFace, etc

**Host the model**:

- Storage: 10 GB +

- Compute:

  - A GPU with a least 6GB of RAM
  - Generation run on Nvidia A100 (40GB) GPU is typically within 4 seconds.
  - Output Image Size: 512 x 512 | 768 x 768

# Why Stable Diffusion?

It is good and **Open Source!**

- Build product on top of it and own it 100%.

- Accelerate creative applications.

- Promote transparency in the development of foundational technology.

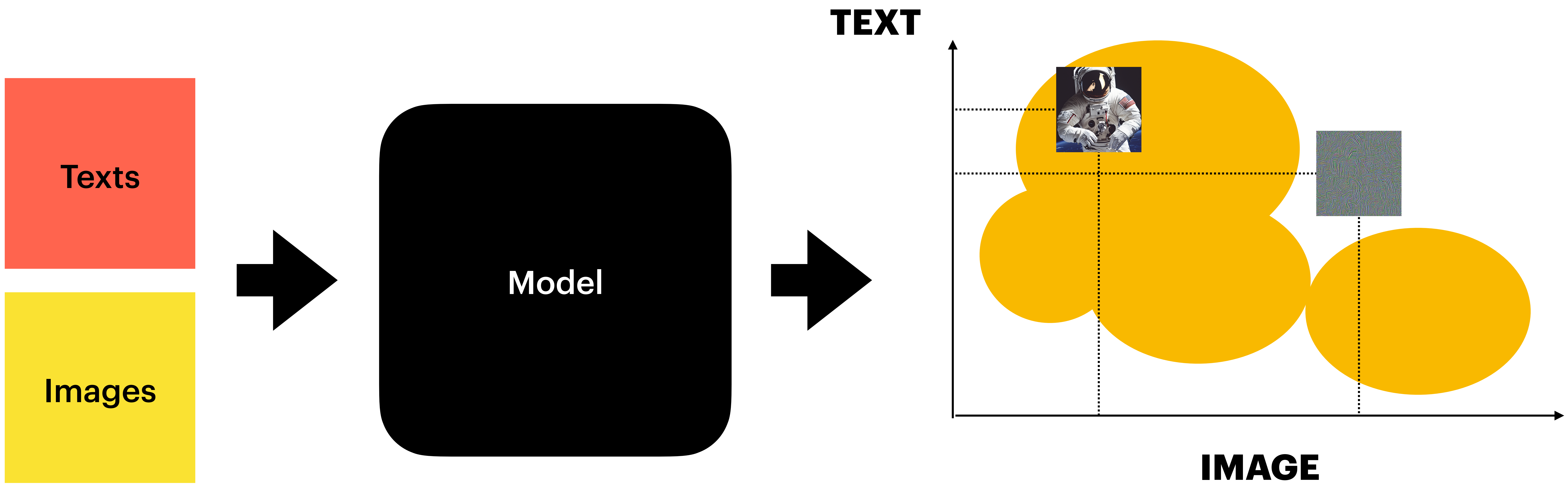
# Style-dependent Generation

# Objective

Adapt the **style** of a concept to another **subject**.

# Prompt Engineering => Search Algorithm

Foundation Model will **learn** a multivariate distribution, you **search** for the result!




**Concept = Subject + Style + Pose + Context**

# Style: Prompt is all you need

**Input [Image]**:

**Prompt [Text]**: Extract the **Subject** from the image, add the **Style** you want.

# Style Transfer

**Input Image**

+ Prompt A

+ Prompt B

Prompt A: **A photo of Amsterdam** in the style of **"Starry Night" by Van Gogh**

Prompt B: **Amsterdam** in the style of **"Starry Night" by Van Gogh**

# Pose-dependent Generation

# Objective

🎯 Generate with the same **pose**, but different **subject / style**

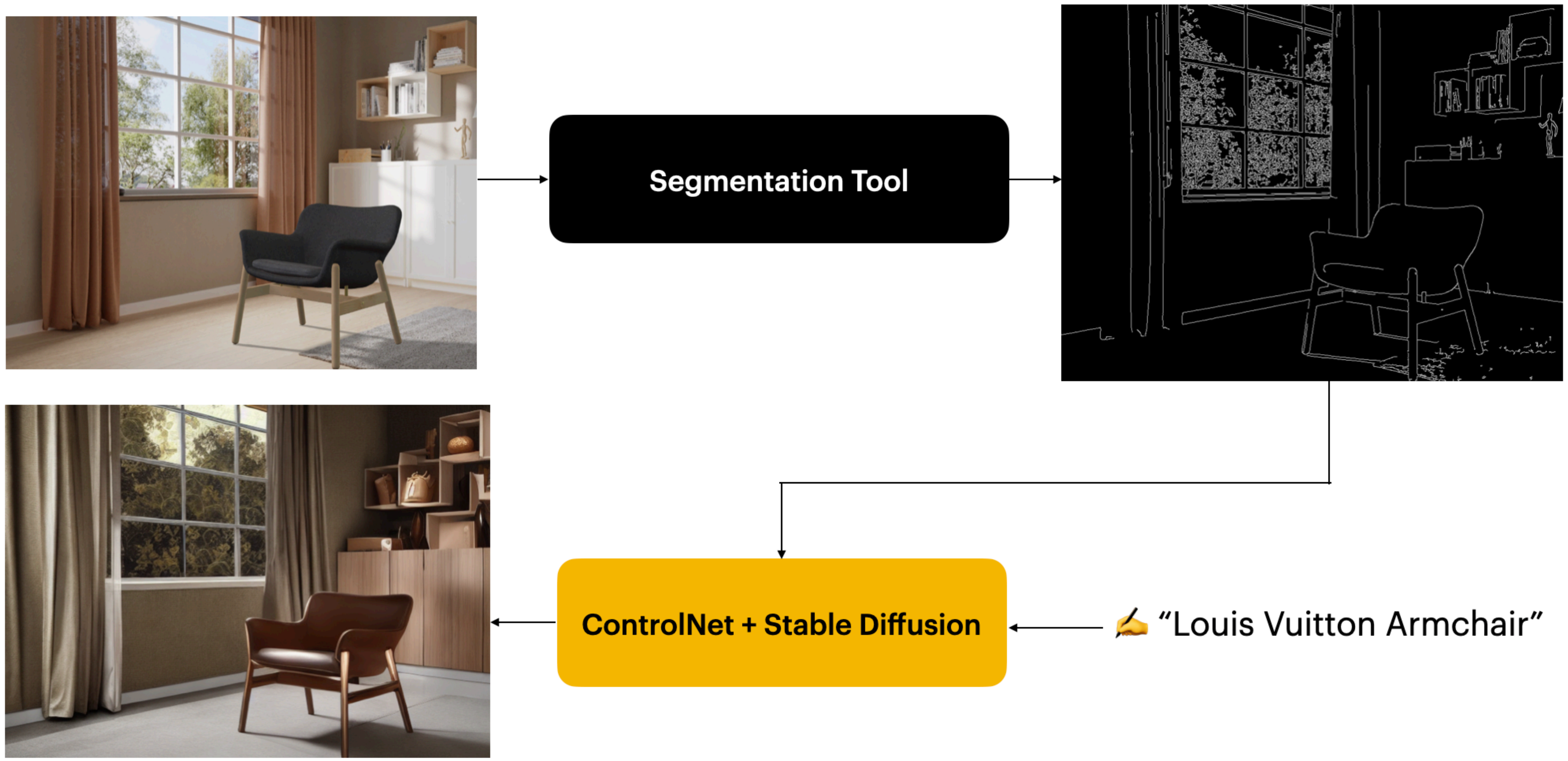
**Concept = Subject + Style + Pose + Context**

# ControlNet

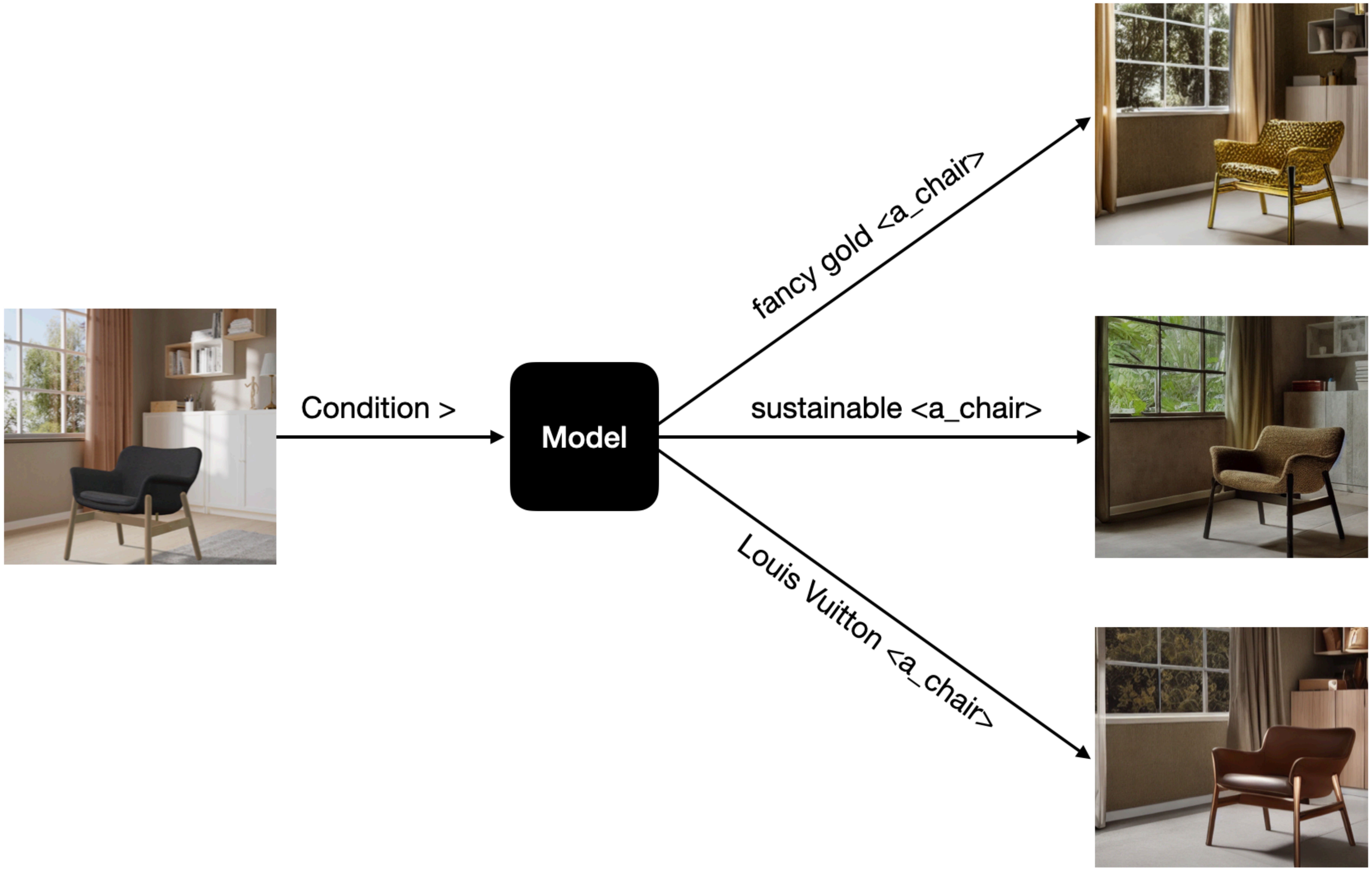
**Paper:** "Adding Conditional Control to Text-to-Image Diffusion Models" (Feb 2023)

**What it does:** Allow Stable Diffusion to accept information specific to pose.

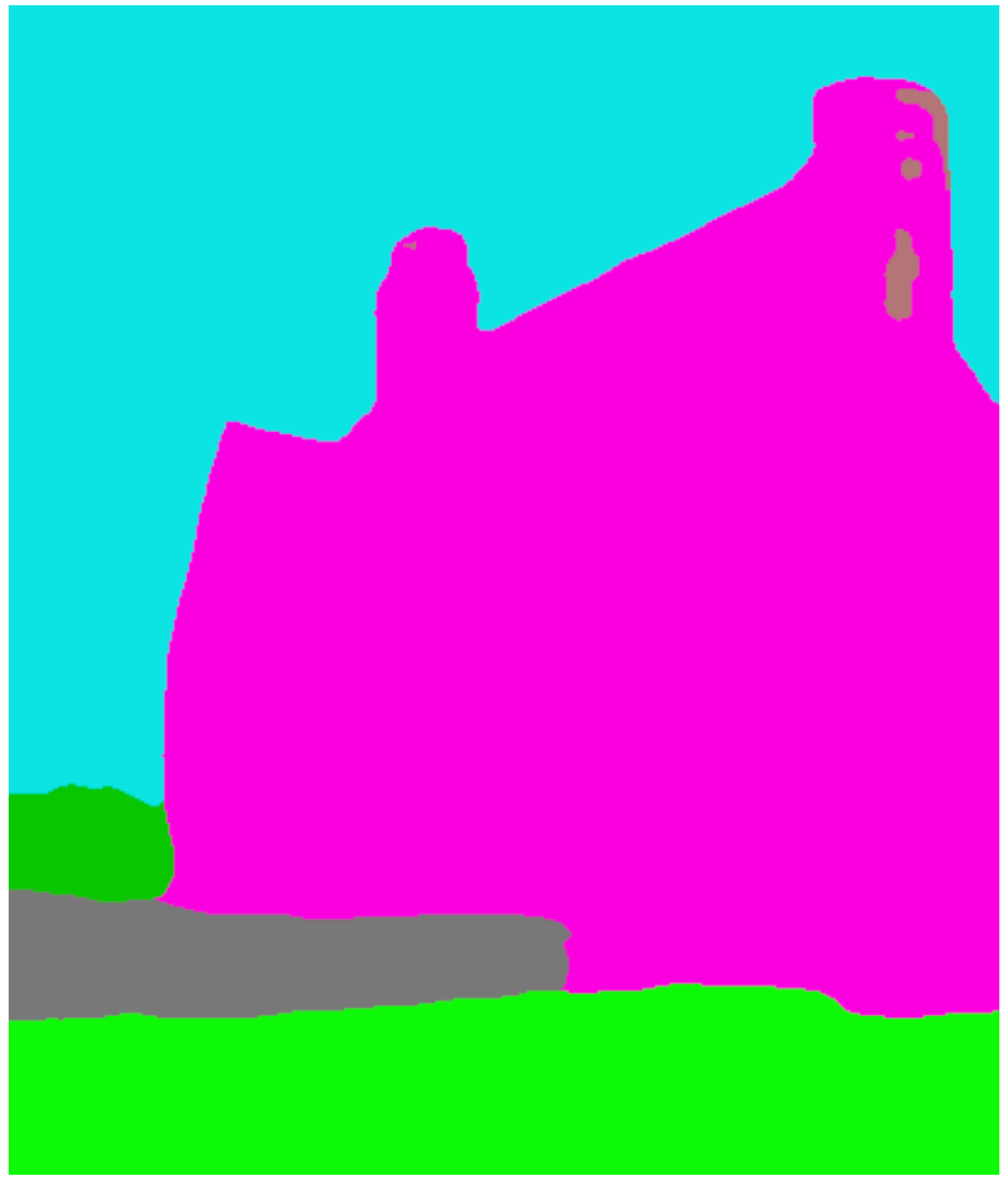
**How it works:**

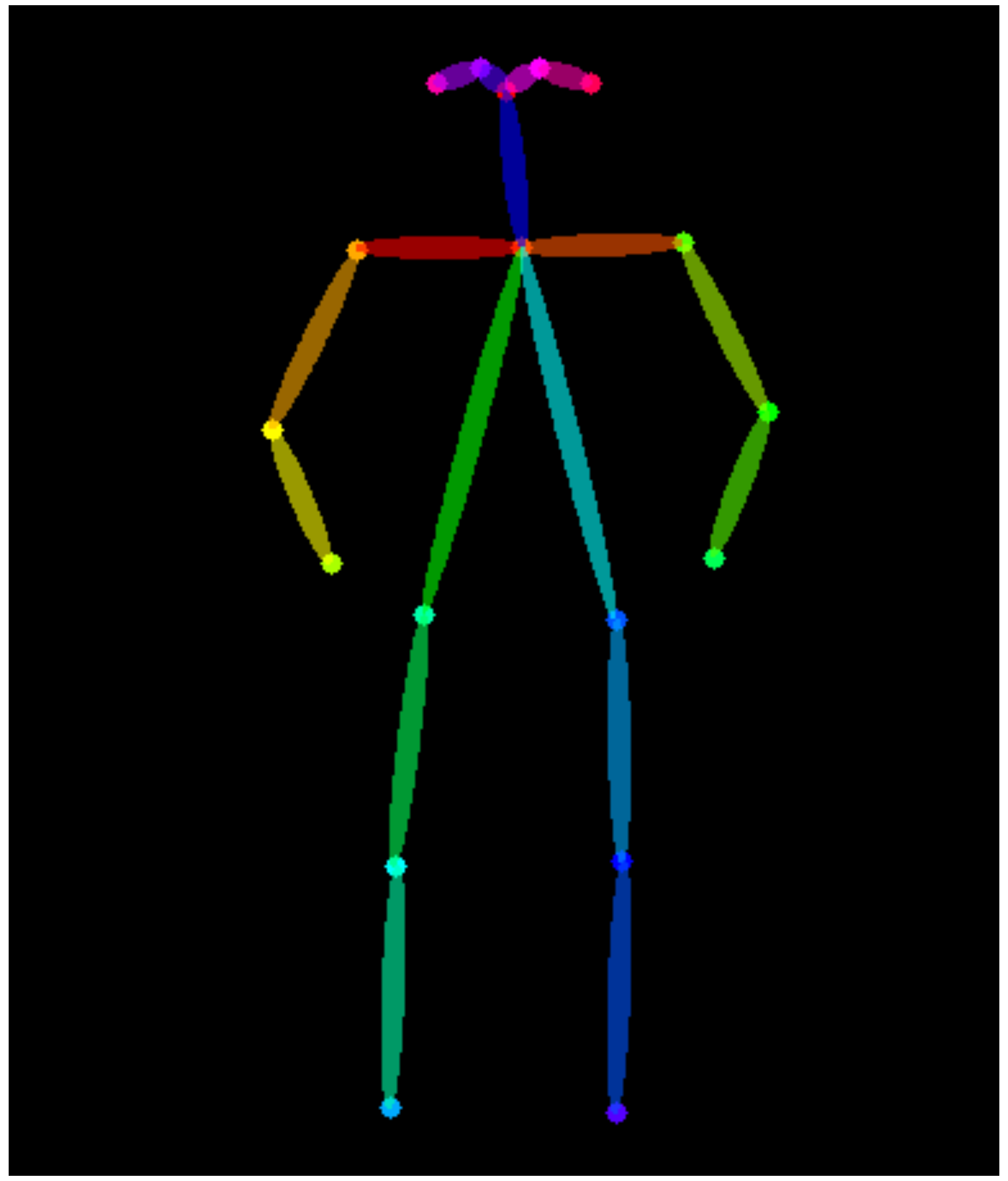


# Demo


Condition >
Model
fancy gold <a_chair>
sustainable <a_chair>
Louis Vuitton <a_chair>

# ControlNet: Other Segmentation Inputs





# Subject-dependent Generation

# Objective

🎯 Generate **subject** in different **style, pose** or **context**

**Concept = Subject + Style + Pose + Context**

# Naive: Image Prompt

The **Midjourney** way: Just use many image as prompts!

**Prompt**: Armchair on a beach

Stable Diffusion

?

# Advanced: Fine-tune Stable Diffusion

**Dreambooth**: A method to personalize text-to-image models like Stable Diffusion given just a few (3-5) images of a subject.

**Steps:**

A. Collect target data (e.g., your faces in different angles, 3-10 images, the more the better)

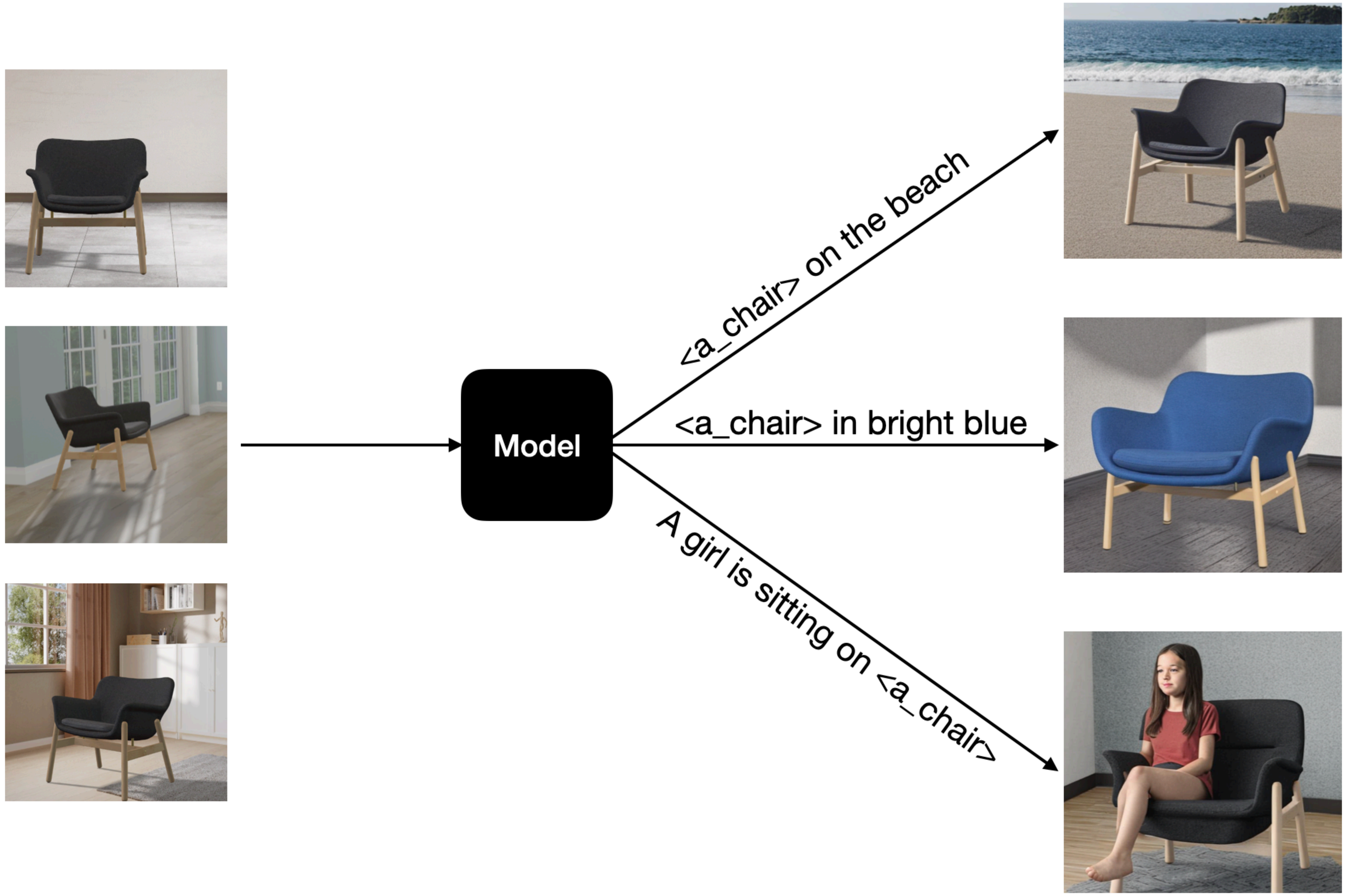
B. Generate sample class data (to preserve model's class-specific knowledge)

C. Fine-tune stable diffusion with [A], [B] and a rare token like **<a_chair>**

D. Generate with the special token, e.g., **<a_chair> on the beach**

# Demo: Generate Furniture

Data Source: Ikea 🙂



# Demo: Generate Furniture

😩 **Main Challenges:**

1. Fine-tuning foundational model gets expensive.

   **-> LoRA**

2. Fine-tuned Model is not good at multi-concept, yet.

   **-> Custom Diffusion**

# Q & A