

# STABLE DIFFUSION

MARCUS BJÖRKLUND & NIKLAS ANTONČIĆ

CADEC 2023.01.19 & 2023.01.25 | CALLISTAENTERPRISE.SE

# CALLISTA

# OVERVIEW



# OVERVIEW

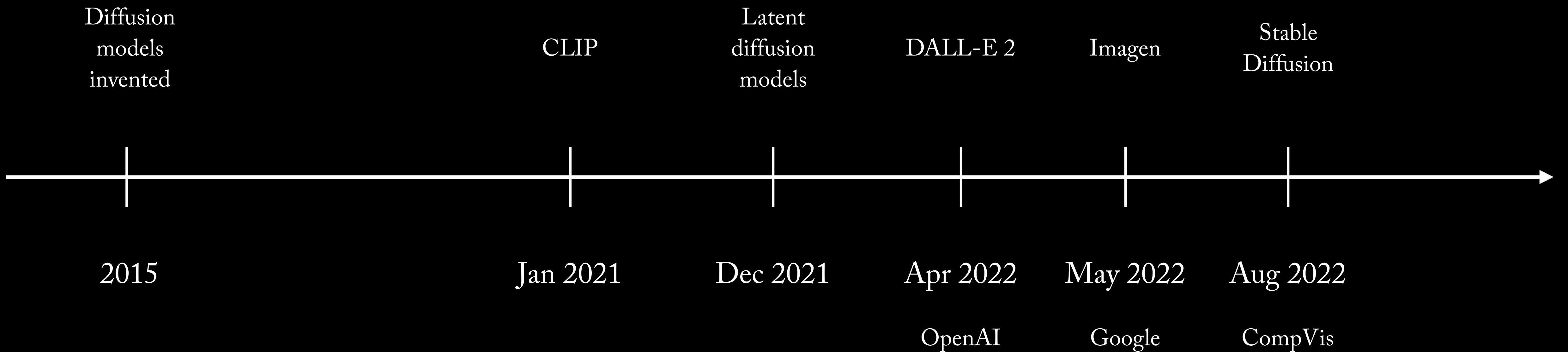
“the most beautiful panoramic landscape, oil painting, where a giant dreamy waterfall creates a river, the trees around are starting to bloom, water shining in the river, a ray of light of the sunset by greg rutkowski”

Stable  
Diffusion



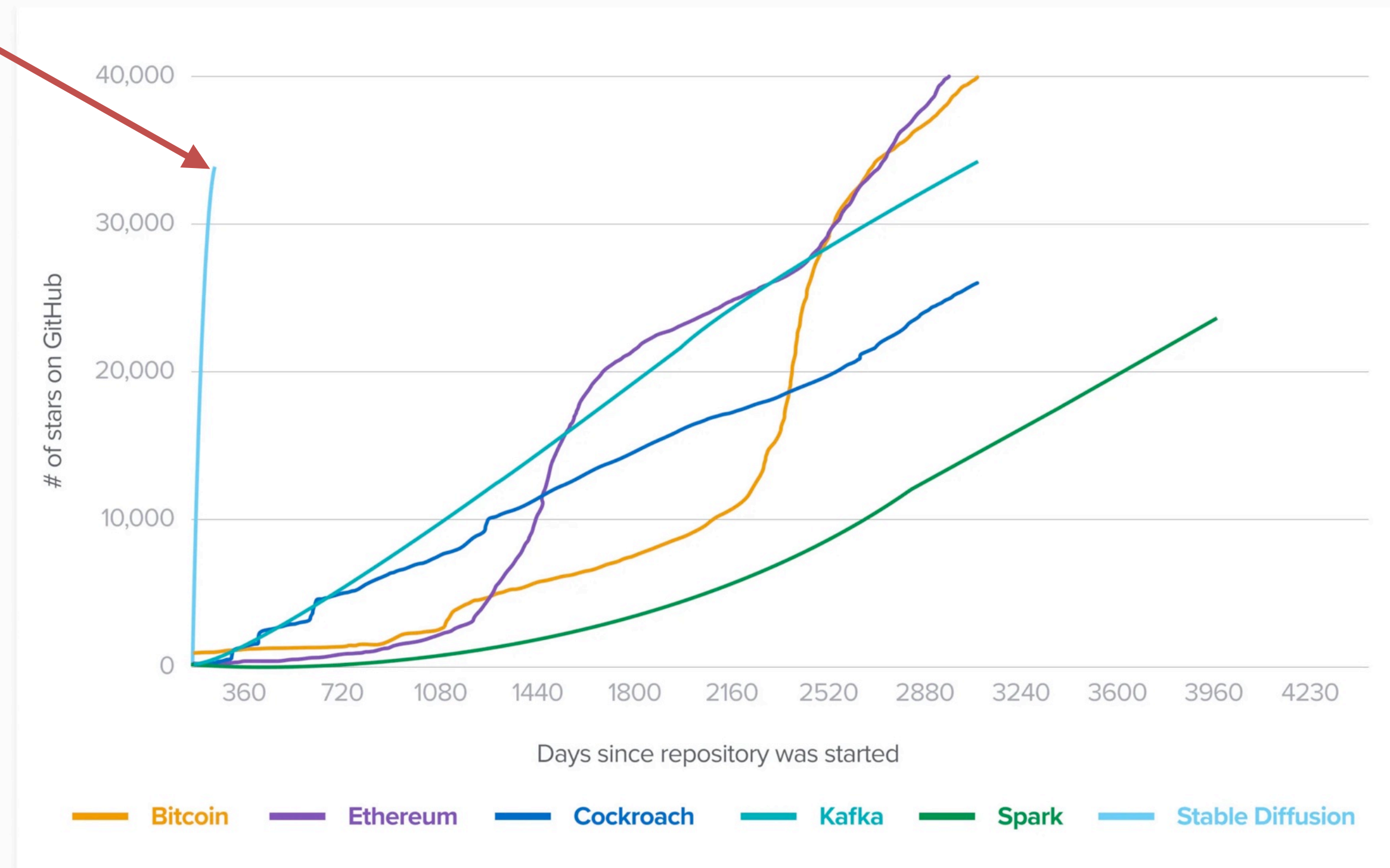
# OVERVIEW

- The evolution of **latent diffusion models**



# OVERVIEW

## Stable Diffusion Developer Adoption



Stars on GitHub for major open source infrastructure technologies. Stable Diffusion accumulated 33,600 stars in its first 90 days, a benchmark other projects achieve in years or decades.

Source: GitHub



# AGENDA

- Overview
- **Stable Diffusion**
  - OpenCLIP
  - Diffusion models
- Live demo
- Legal aspects
- Summary

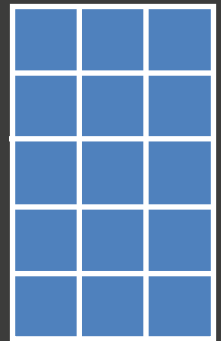
# | STABLE DIFFUSION

# STABLE DIFFUSION

Stable Diffusion model

“a photograph of  
an astronaut  
riding a horse”

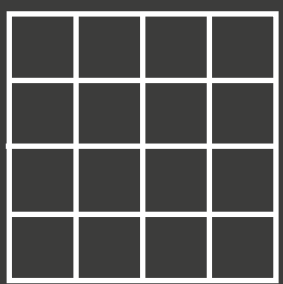
Text  
Encoder  
(OpenCLIP)



Token  
embeddings

1

2

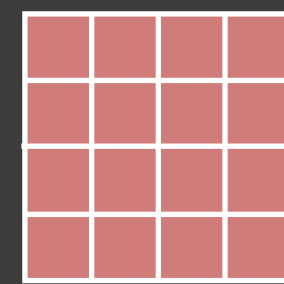


Random image  
information tensor

Image Information Creator

Diffusion

3



Processed image  
information tensor

Image  
Decoder



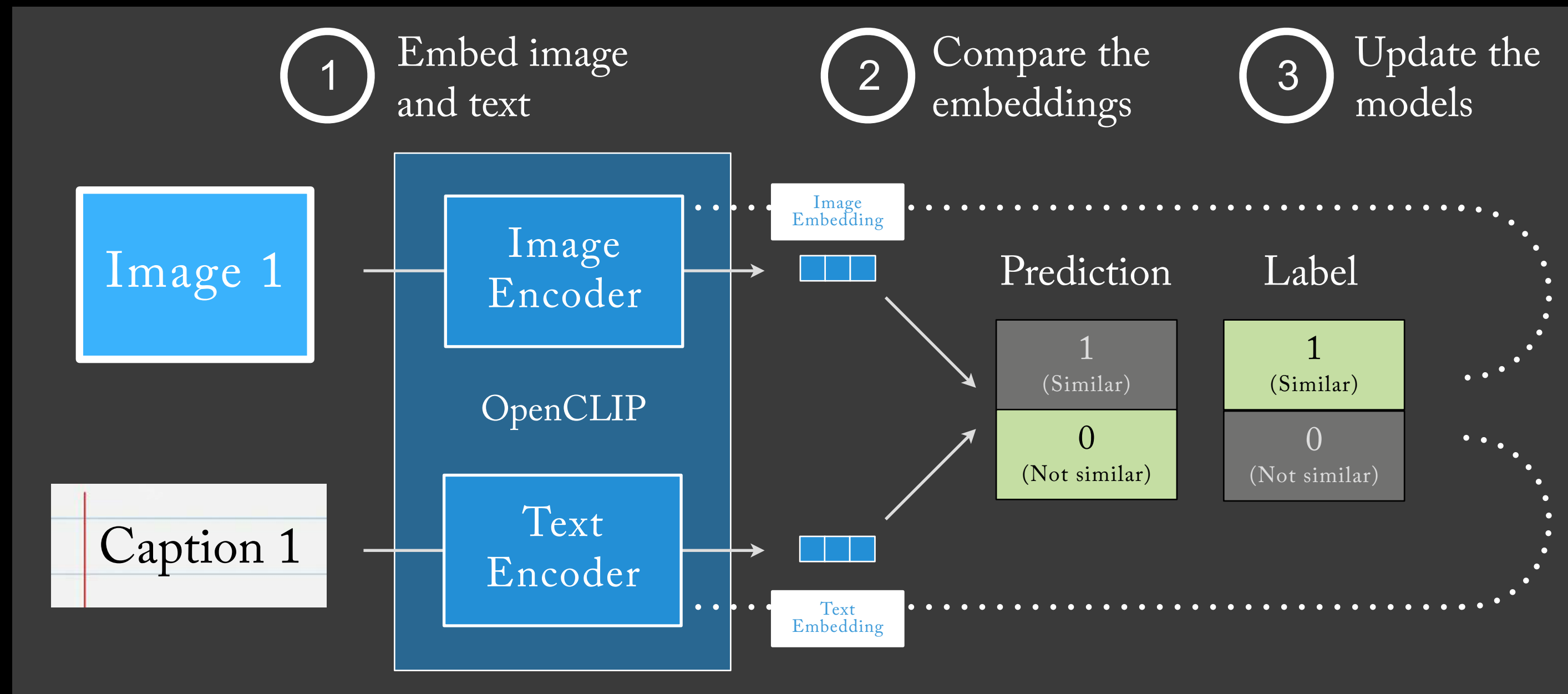


**OPENCLIP**

CALLISTA

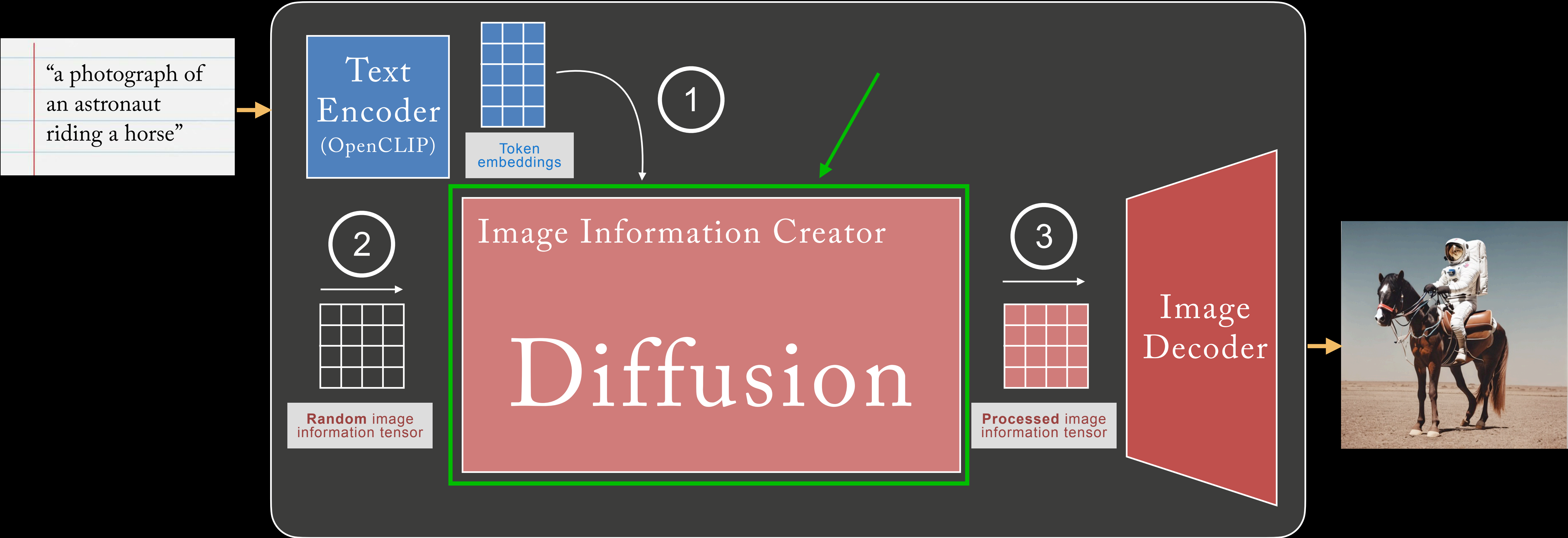
# OPENCLIP

- Was trained on LAION-2B (2.3 billion pictures fetched from Internet)
- OpenCLIP is a combination of an image encoder and a text encoder
- During training, images and captions are encoded into two numerical embeddings and the model is continuously updated until the two embeddings are similar



# STABLE DIFFUSION

## Stable Diffusion model

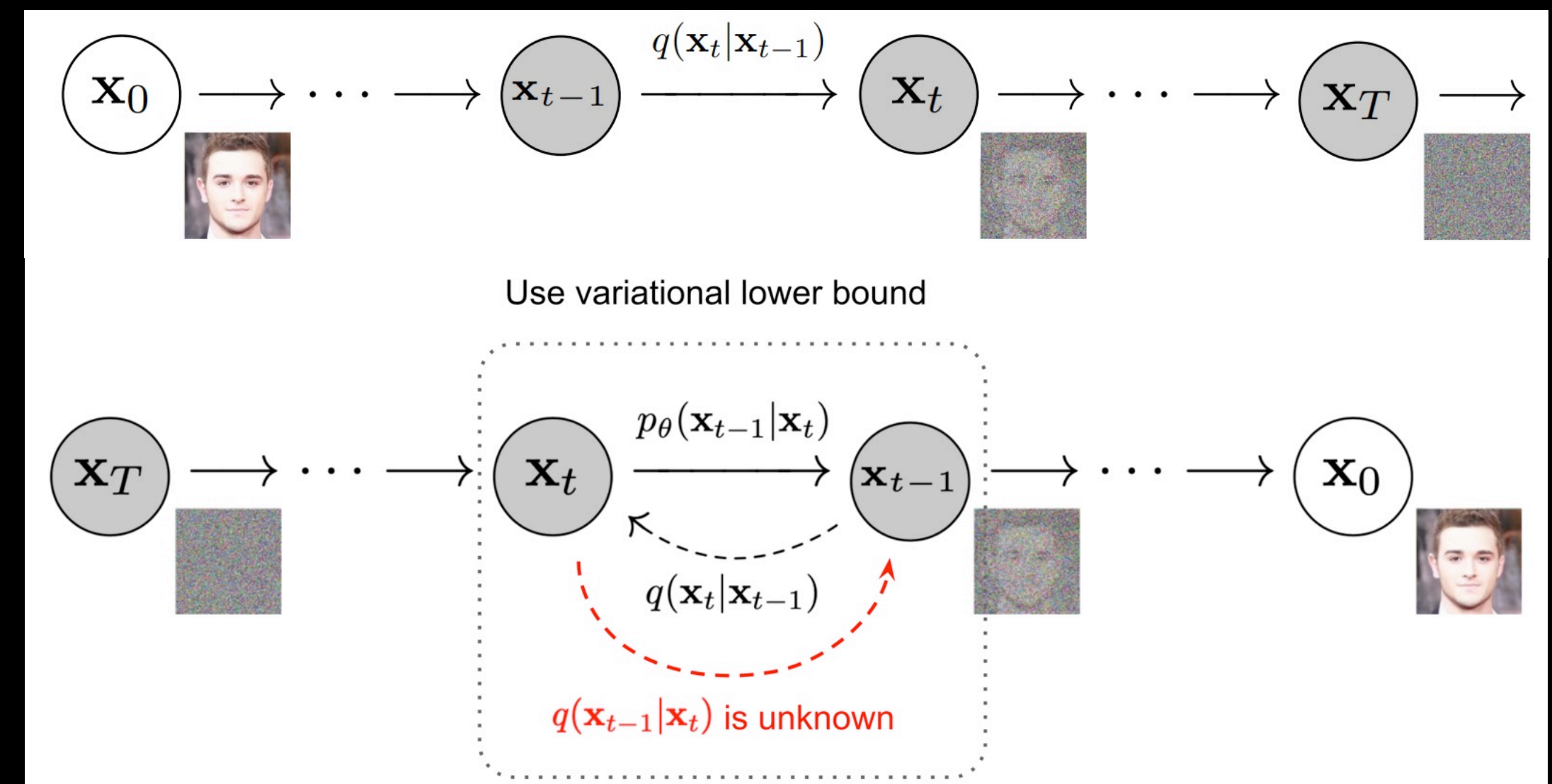


# DIFFUSION MODELS - PROCESS



# DIFFUSION MODELS - PROCESS

- A Markov chain of diffusion steps
- Forward diffusion:
  - add Gaussian noise to image through a series of  $T$  steps
- Reverse diffusion:
  - trained to recover original image by removing noise

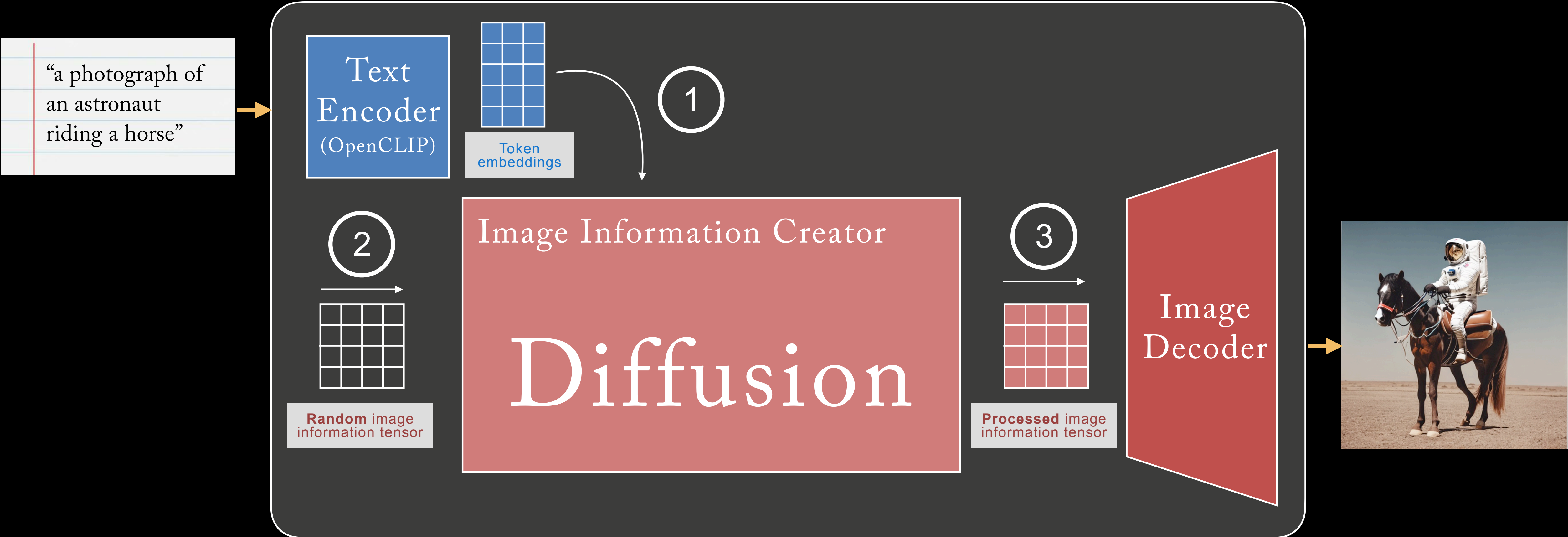


The Markov chain of forward (reverse) diffusion process of generating a sample by slowly adding (removing) noise.

Image source: *Ho et al. 2020*

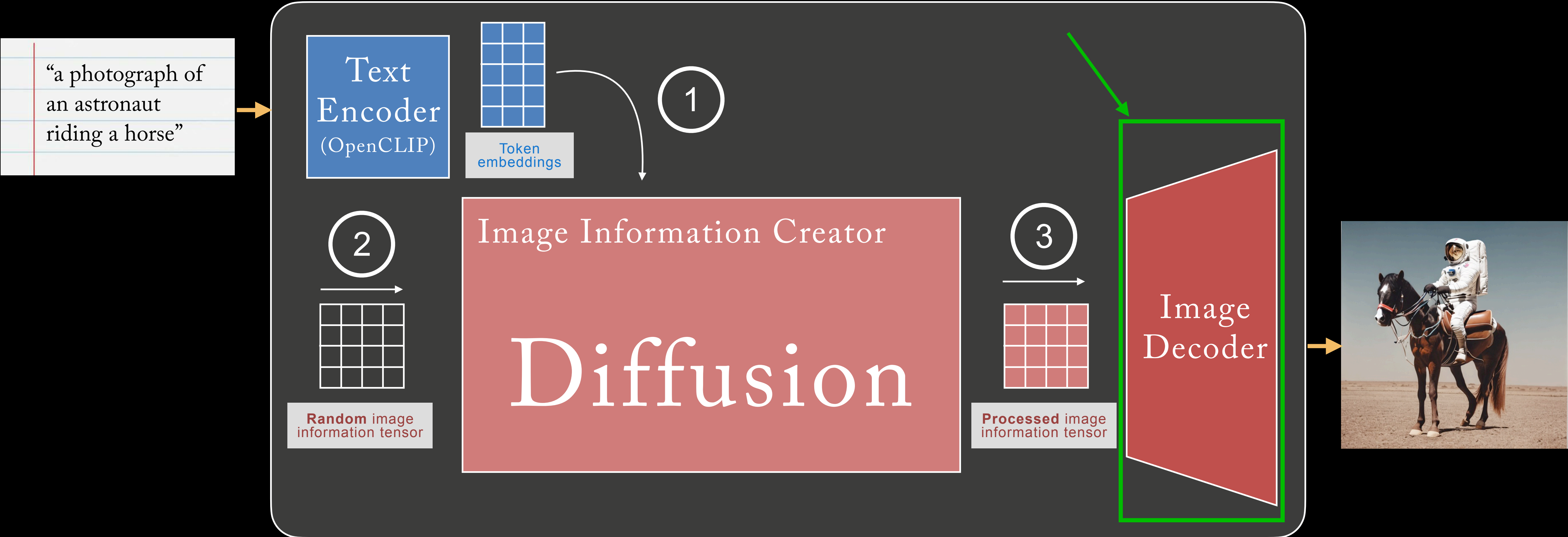
# STABLE DIFFUSION

## Stable Diffusion model



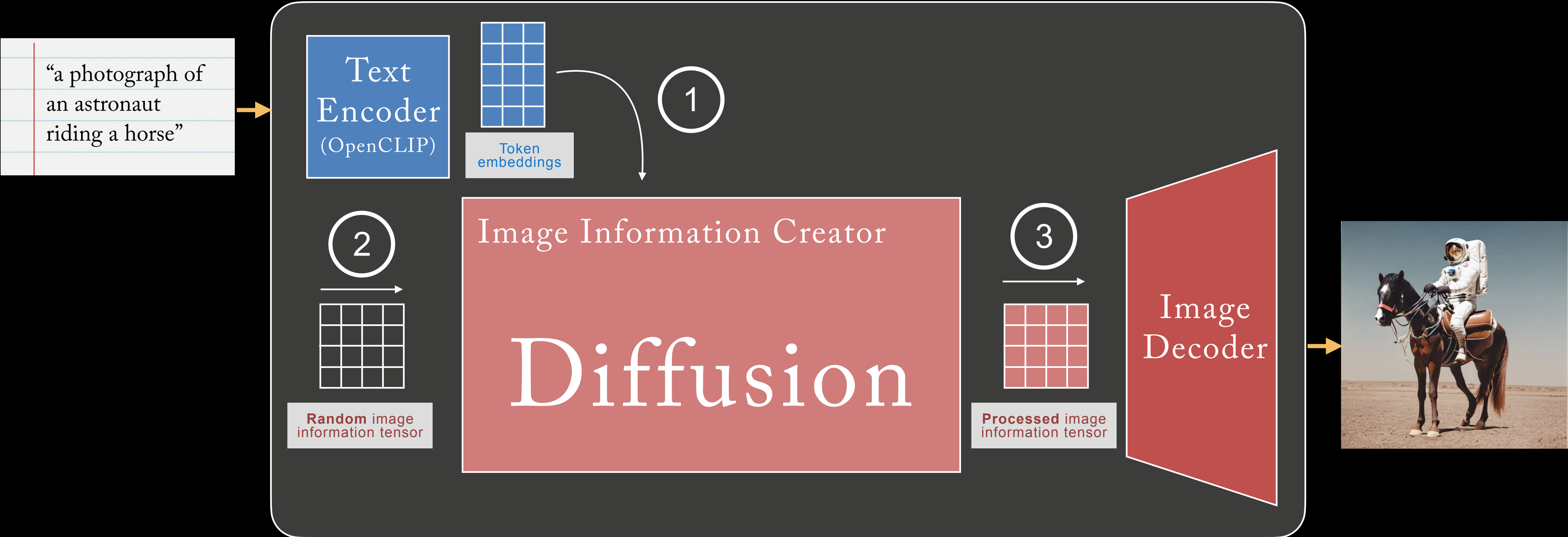
# STABLE DIFFUSION

## Stable Diffusion model



# STABLE DIFFUSION

## Stable Diffusion model

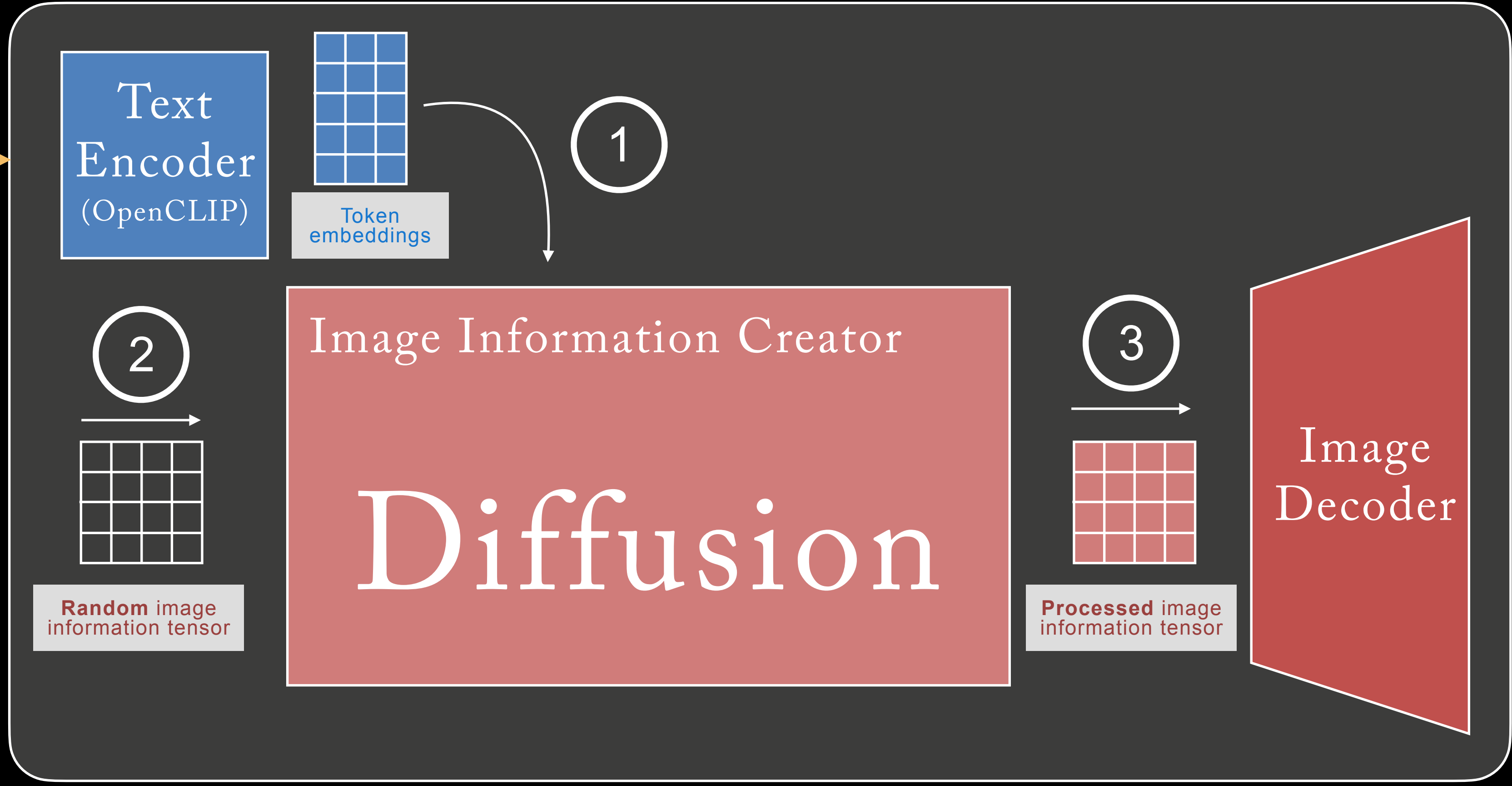




# STABLE DIFFUSION

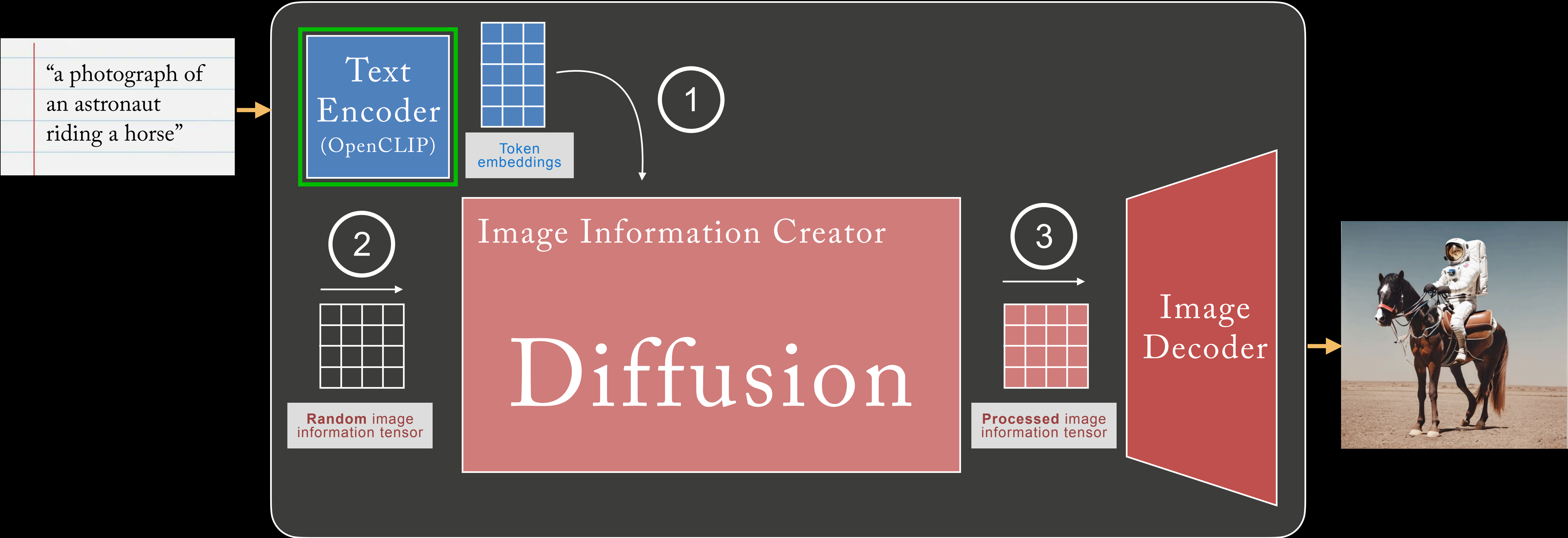
## Stable Diffusion model

“a photograph of an astronaut riding a horse”



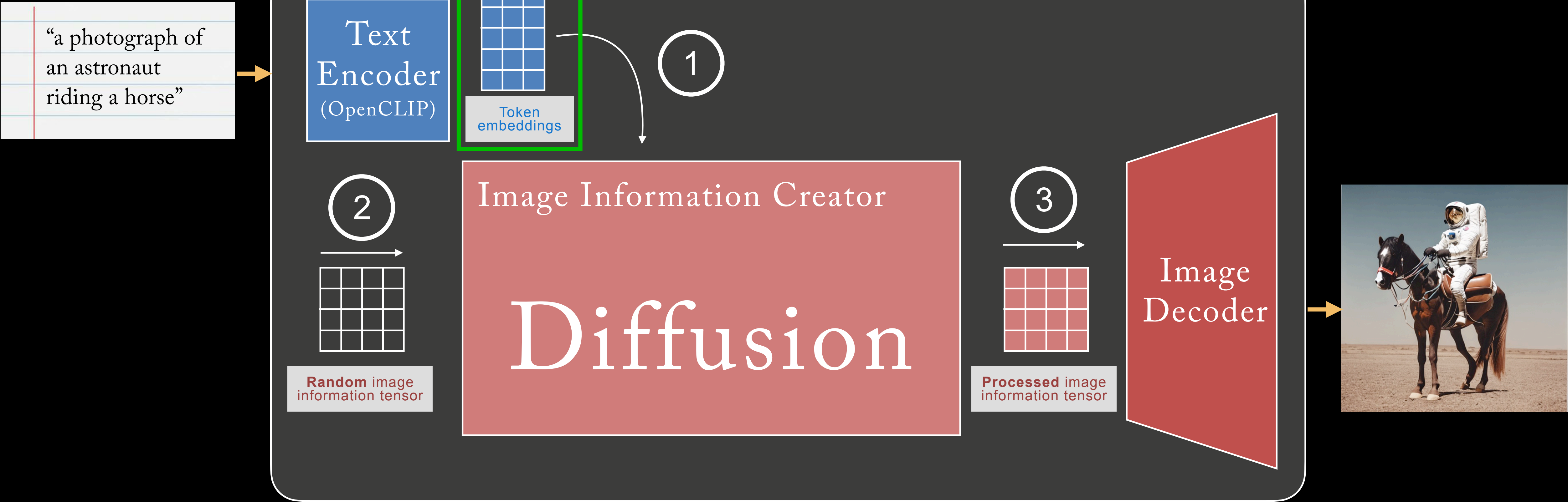
# STABLE DIFFUSION

## Stable Diffusion model



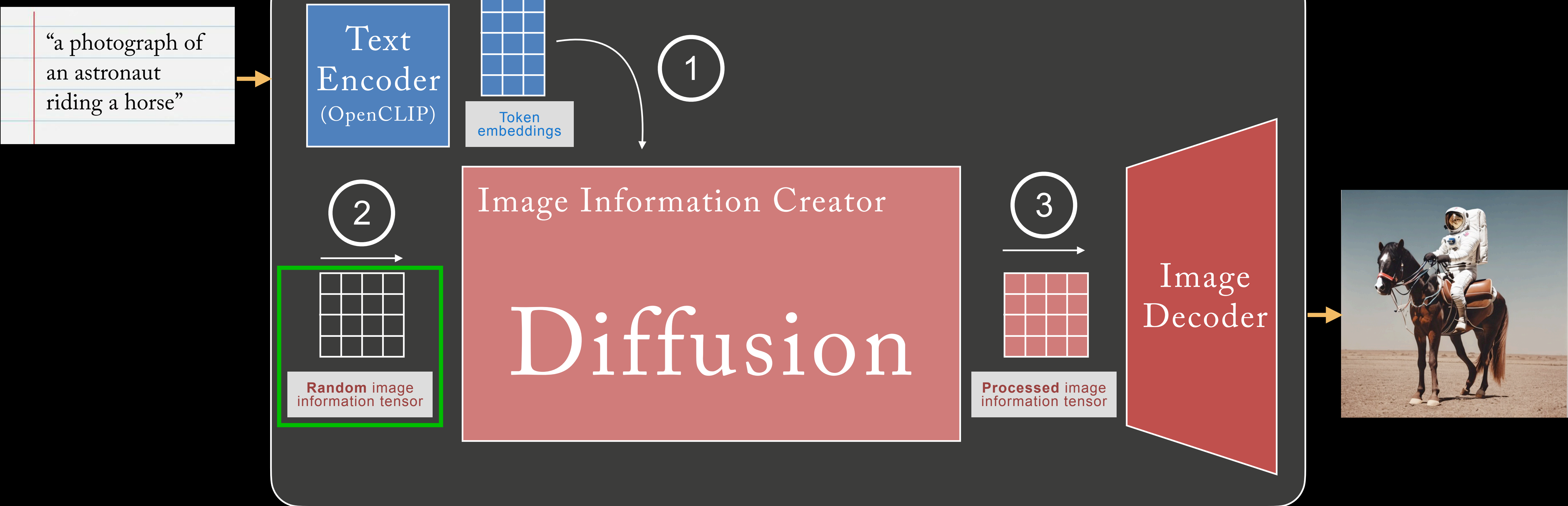
# STABLE DIFFUSION

## Stable Diffusion model



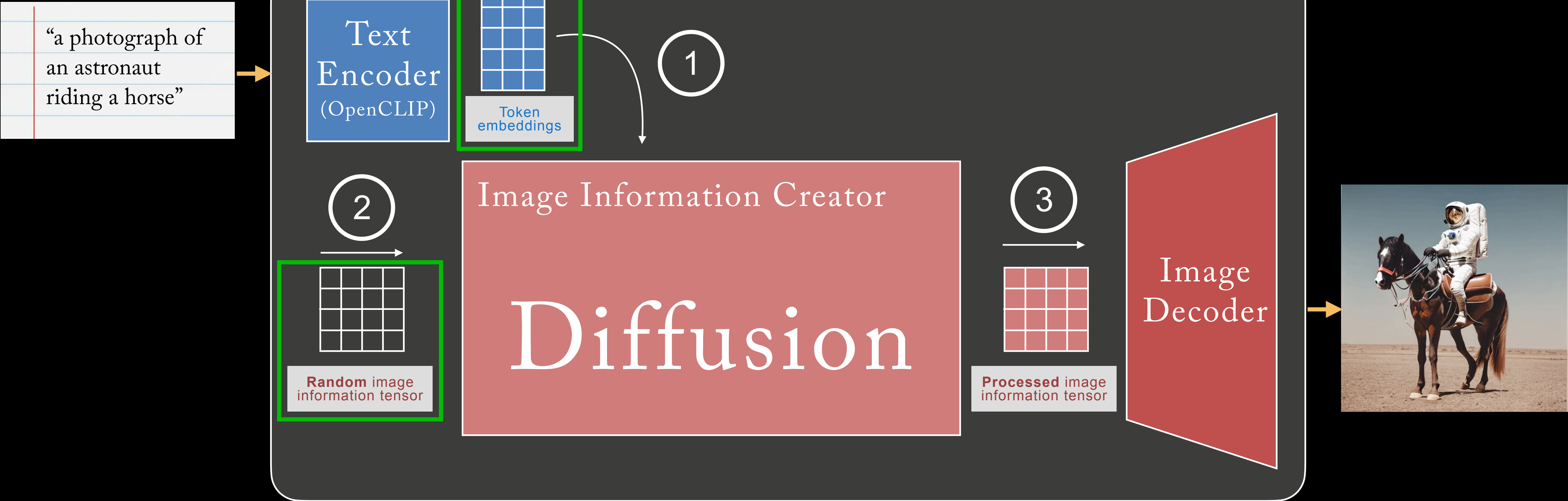
# STABLE DIFFUSION

## Stable Diffusion model



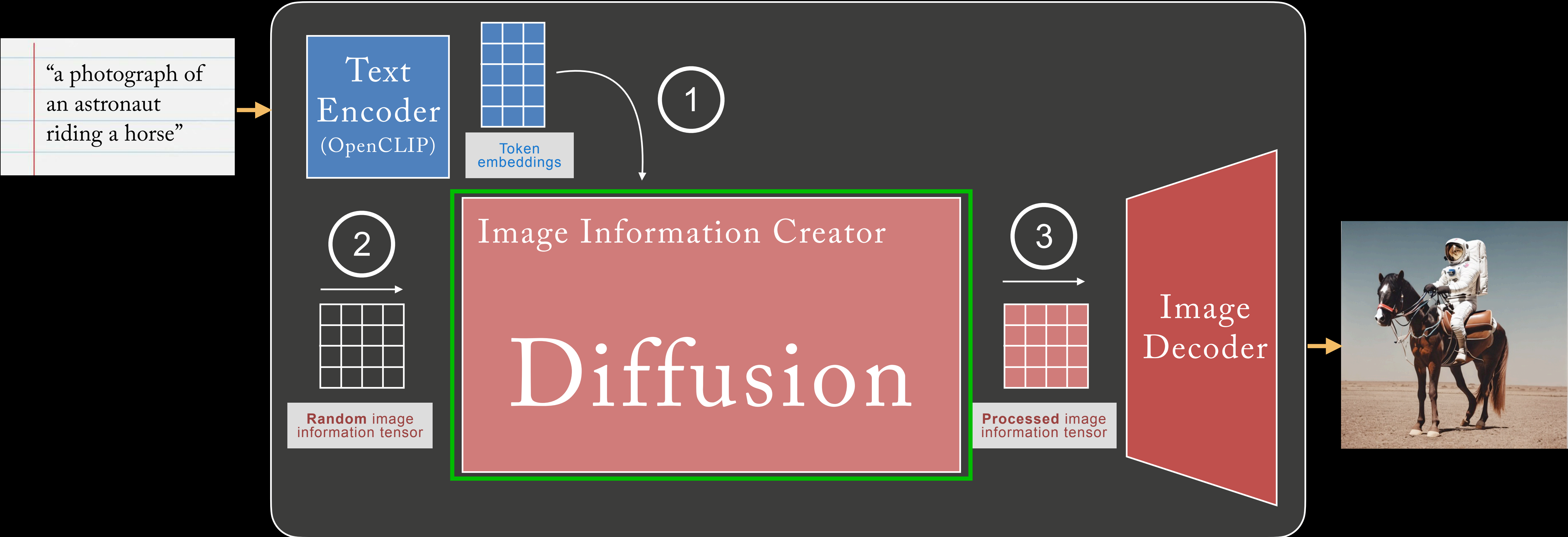
# STABLE DIFFUSION

## Stable Diffusion model



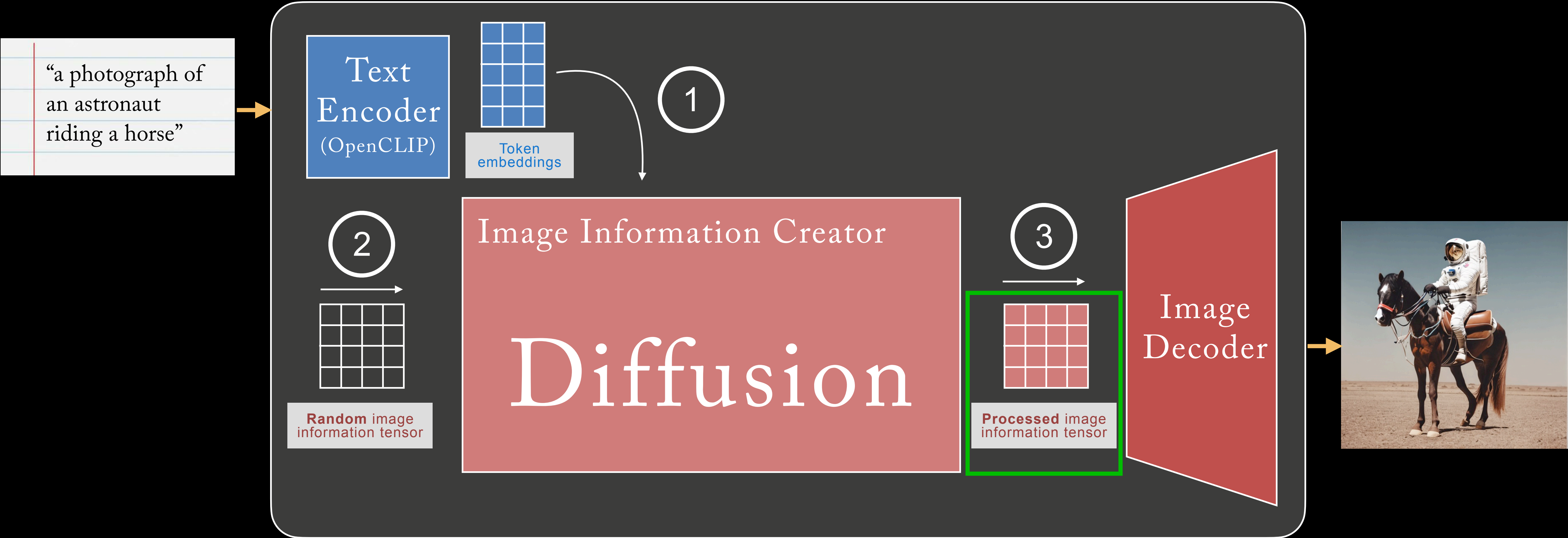
# STABLE DIFFUSION

## Stable Diffusion model



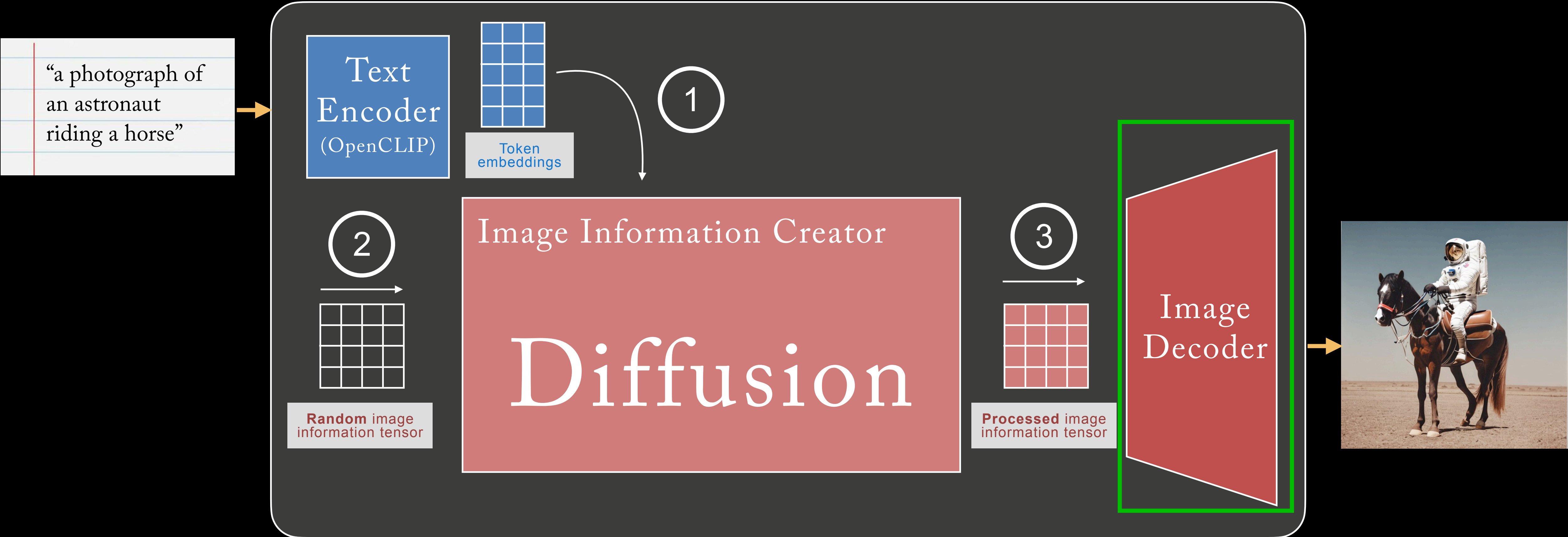
# STABLE DIFFUSION

## Stable Diffusion model



# STABLE DIFFUSION

## Stable Diffusion model



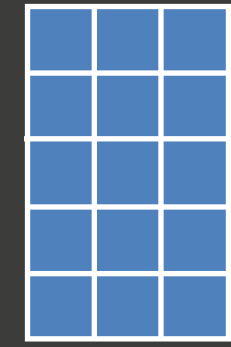


# STABLE DIFFUSION

## Stable Diffusion model

“a photograph of  
an astronaut  
riding a horse”

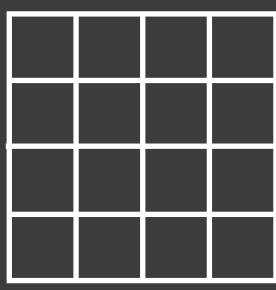
Text  
Encoder  
(OpenCLIP)



Token  
embeddings

1

2

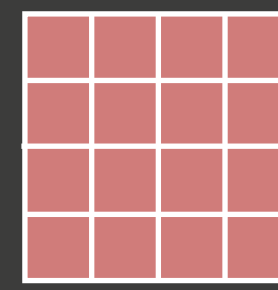


Random image  
information tensor

Image Information Creator

Diffusion

3



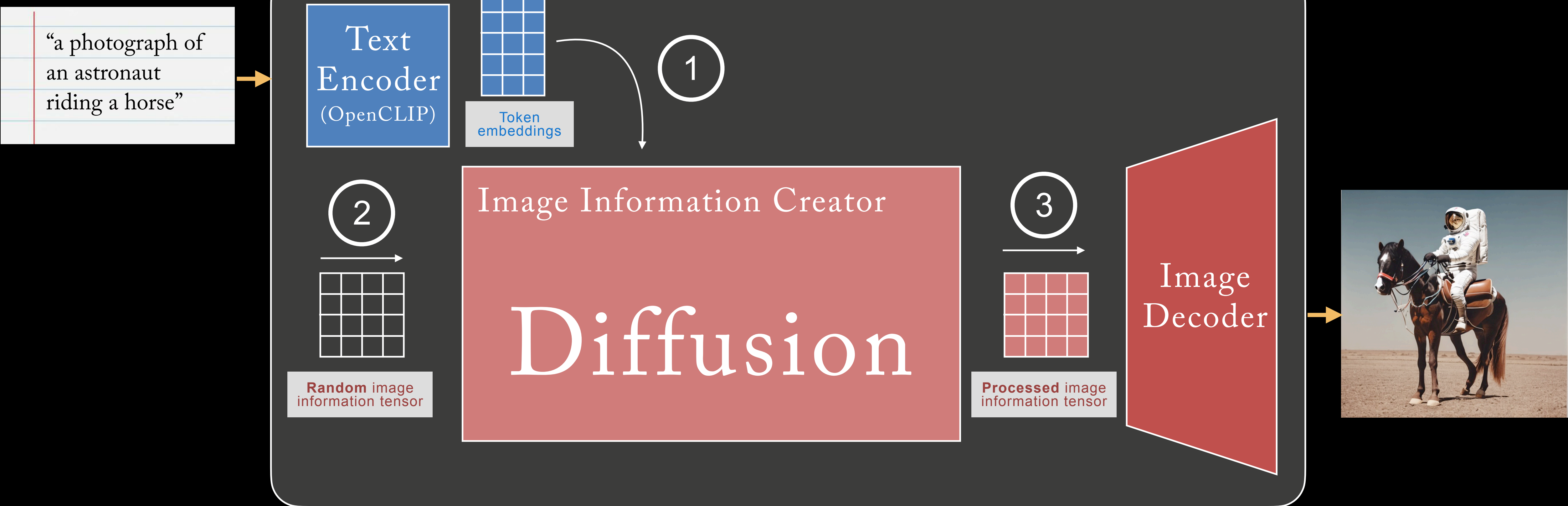
Processed image  
information tensor

Image  
Decoder



# STABLE DIFFUSION

## Stable Diffusion model



# AGENDA

- Overview
- Stable Diffusion
  - OpenCLIP
  - Diffusion models
- **Live demo**
- Legal aspects
- Summary

## AGENDA STABLE DIFFUSION DEMO

- Hardware
- SD Implementations
- Text to Image
- Image to Image
- Finetuning
- Upscaling

## **HARDWARE - THE BEAST**

PC from 2017

CPU Intel X86

GPU Nvidia 1080ti



## | HARDWARE - WHY NOT RUN ON MY MAC?

Stable Diffusion uses PyTorch framework for ML



- It can run on both CPU and GPU
- CPU 5-10 times slower than GPU
- PyTorch uses CUDA framework for GPU integration with Nvidia GPU
- Recently also support for Apples Metal Performance Shaders for GPU on Apple Silicon



# HARDWARE - BENCHMARK

Generating two images with same prompt and seed

- Macbook M1 Pro on CPU 20 min
- Macbook M1 Pro on GPU 5 min
- The Beast on 1080ti 1 min



## STABLE DIFFUSION IMPLEMENTATIONS

- Original reference implementation by CompVis (1.x)
- Latest reference implementation by Stability AI (2.x)

```
niklas@satchmo: ~/code/stable-diffusion
niklas@satchmo:~/code/stable-diffusion$ python scripts/txt2img.py --prompt "a photograph of an astronaut riding a horse"
```

- Text to Image
- Image to Image
- Inpainting
- Upscaling x4
- Depth to Image





# STABLE DIFFUSION WEBUI

- Created by AUTOMATIC1111
- Uses reference implementation
- Lots of added functionality
- REST API in Python
- Web UI in Python
- Extensions

The screenshot displays the Stable Diffusion WebUI interface. At the top, the checkpoint is set to 'protogenX34OfficialR\_1.ckpt [60fe2f34]'. The prompt is 'green sapling rowing out of ground, mud, dirt, grass, high quality, photorealistic, sharp focus, depth of field'. The negative prompt is empty. The 'Generate' button is highlighted in orange. Below the prompt, there are settings for 'Sampling method' (Euler a), 'Sampling steps' (20), 'Width' (512), 'Height' (512), 'CFG Scale' (12), 'Seed' (1441787169), 'Batch count' (4), and 'Batch size' (1). There are also checkboxes for 'Restore faces', 'Tiling', and 'Hires. fix'. The 'Script' is set to 'None'. A generated image of a green sapling growing out of the ground is shown in a preview window. Below the image, there are buttons for 'Save', 'Zip', 'Send to img2img', 'Send to inpaint', and 'Send to extras'. The prompt and generation details are repeated below the image. At the bottom, there are links for 'API', 'Github', 'Gradio', and 'Reload UI', along with system information: 'python: 3.10.6 • torch: 1.13.1+cu117 • xformers: N/A • gradio: 3.15.0 • commit: 983167e6'.

## STABLE DIFFUSION WEBUI - SETUP

- Graphics card driver
- CUDA Toolkit
- Python
- Clone the project repository
- xFormers
- model weights (20 ish GB)

## DEMO - TEXT TO IMAGE

*Starwars film poster with Luke Skywalker riding a bike on a sunny day.*

Model: v2.1\_768, Model hash: ad2a33c361, Sampler: Euler a, Steps: 20, CFG scale: 7, Size: 768x768, Seed: 2782

- Model - actual trained model file
- Prompt - text
- Sampling method - Disc Diff Eq solver algorithm
- Classifier-Free Guidance Scale - Prompt “volume control”
- Dimensions - One side should be as trained
- Seed - random default

## DEMO - IMAGE TO IMAGE

- Applying style to an existing image guided by text
- De-noising strength is “volume control” for Img2Img

Creating a superhero out of an IT-consultant

# DEMO - FINETUNING WITH DREAMBOOTH

- DreamBooth
- Re-training part of the model with some images to create a new category

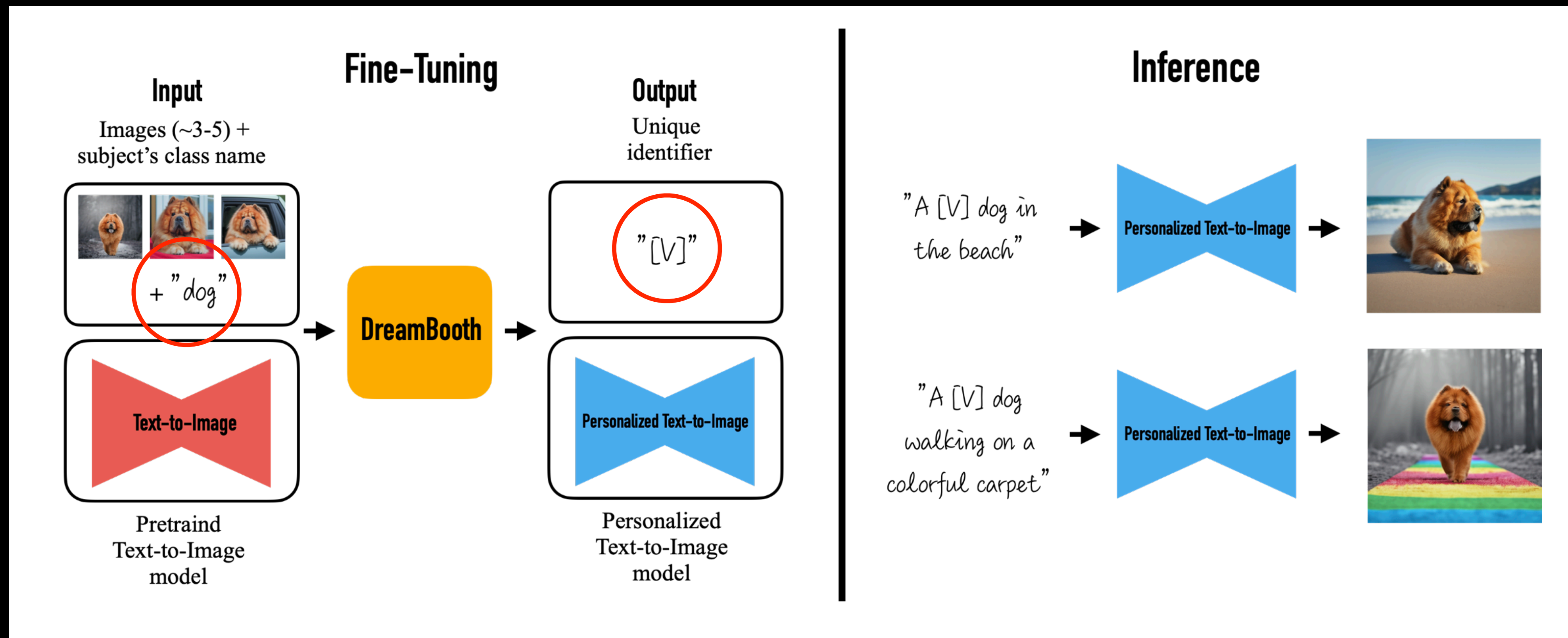
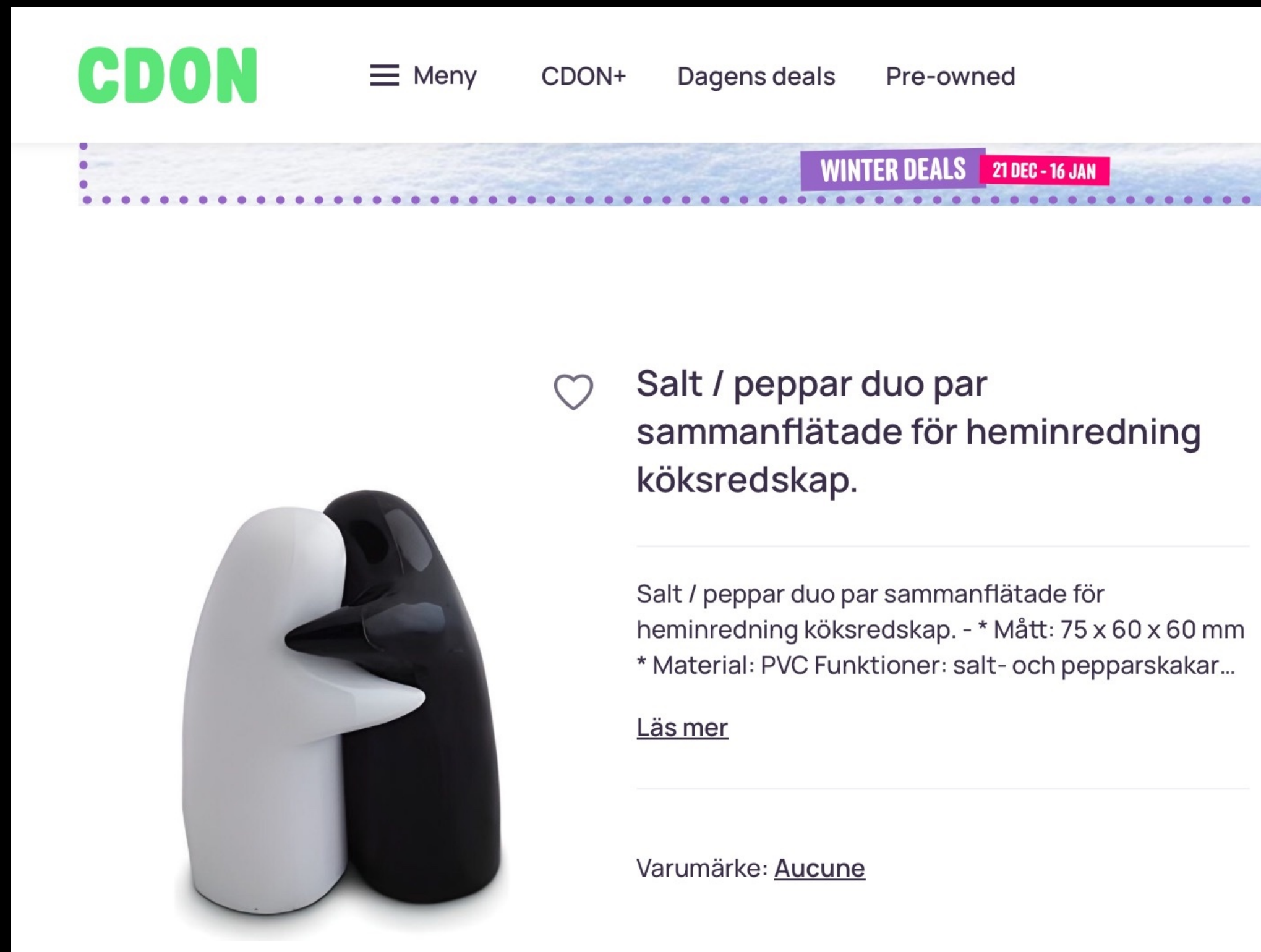


Image from Google paper <https://arxiv.org/abs/2208.12242>

# DEMO - FINETUNING WITH DREAMBOOTH




The screenshot shows a product listing on the CDON website. At the top left is the CDON logo in green. To its right are navigation links: a menu icon, 'Meny', 'CDON+', 'Dagens deals', and 'Pre-owned'. A purple banner across the top reads 'WINTER DEALS 21 DEC - 16 JAN'. The product is a 'Salt / peppar duo par sammanflätade för heminredning köksredskap.' The image shows two curved, interlocking pieces, one white and one black. The text below the image provides details: 'Salt / peppar duo par sammanflätade för heminredning köksredskap. - \* Mått: 75 x 60 x 60 mm \* Material: PVC Funktioner: salt- och pepparskakar...'. There is a 'Läs mer' link and a brand name 'Aucune' at the bottom.

**CDON**    ☰ Meny    CDON+    Dagens deals    Pre-owned

WINTER DEALS 21 DEC - 16 JAN

♡ Salt / peppar duo par sammanflätade för heminredning köksredskap.



Salt / peppar duo par sammanflätade för heminredning köksredskap. - \* Mått: 75 x 60 x 60 mm  
\* Material: PVC Funktioner: salt- och pepparskakar...

[Läs mer](#)

Varumärke: [Aucune](#)

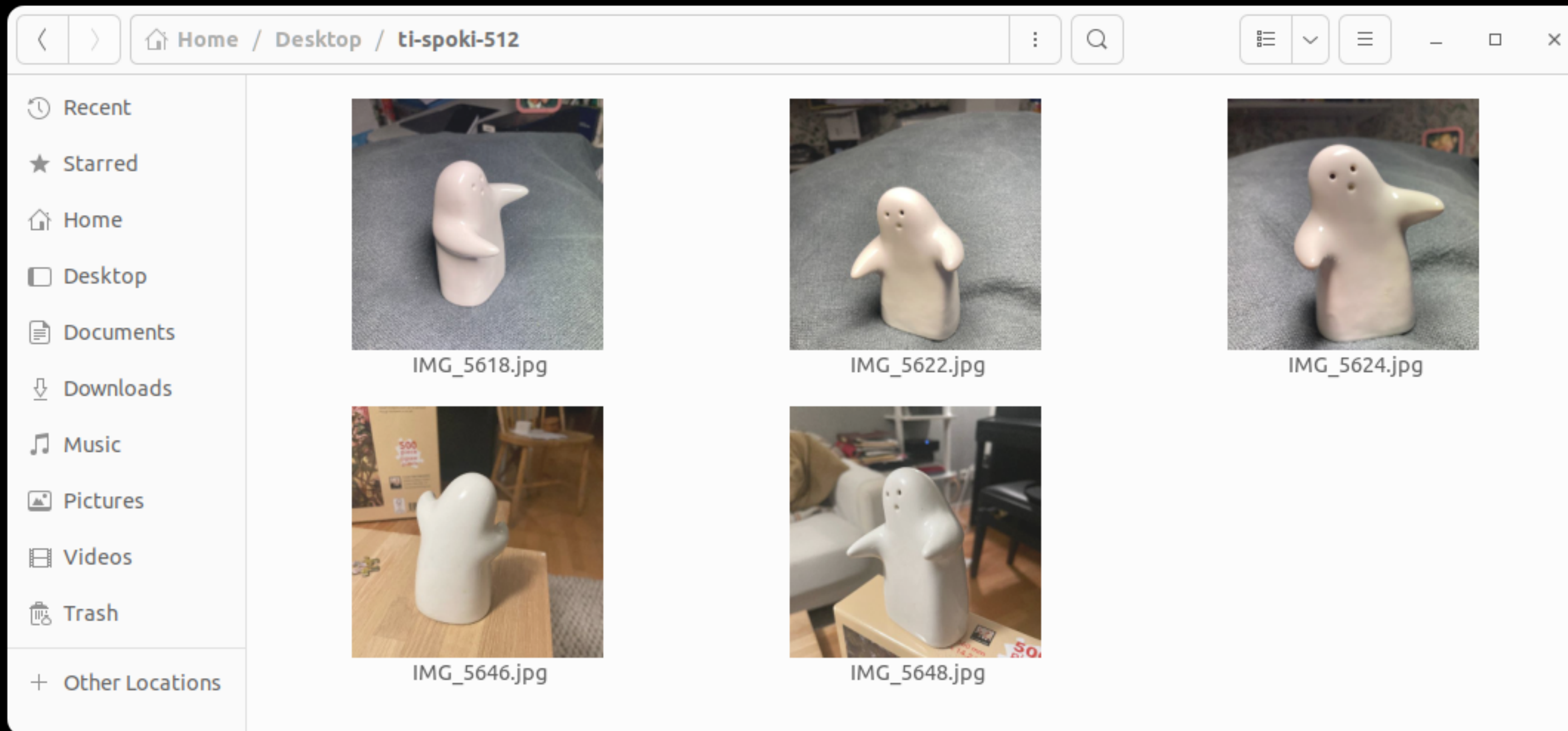
# DEMO - FINETUNING WITH DREAMBOOTH

- Spoki



# DEMO - FINETUNING WITH DREAMBOOTH

- Class prompt: A photo of a **figure**
- Instance prompt: A photo of a **spoki** figure





# DEMO - FINETUNING WITH DREAMBOOTH

DreamBooth\_Stable\_Diffusion\_SDA.ipynb

File Edit View Insert Runtime Tools Help Cannot save changes

RAM Disk Editing

Files

- data
  - class
  - instance
    - IMG\_5618.jpg
    - IMG\_5622.jpg
    - IMG\_5624.jpg
    - IMG\_5646.jpg
    - IMG\_5648.jpg
- drive
- sample\_data
- stable\_diffusion\_weights
  - output
    - 0
    - 800
      - samples
      - scheduler
      - text\_encoder
      - tokenizer
      - unet
      - vae
        - args.json
        - model.ckpt
        - model\_index.json
- concepts\_list.json
- convert\_diffusers\_to\_original\_stabl...
- grid.png
- train\_dreambooth.py

Disk 43.02 GB available

Example 1: photo of zwx person, photo of a person  
Example 2: photo of zwx toy, photo of a toy

```
instance_prompt: "photo of spoki figure"  
class_prompt: "photo of a figure"  
training_steps: 800
```

Convert to fp16? (takes half the space (2GB)).  
fp16:


Compile xformers (Try only if you see xformers error. Will take 1 more hour).  
complie\_xformers:

Clear log after run?  
CLEAR\_LOG:

Show code

[\*] WEIGHTS\_DIR=/content/stable\_diffusion\_weights/output/800  
Dreambooth completed successfully. It took 30.8 minutes.  
Model saved to /content/drive/MyDrive/Dreambooth\_model/model.ckpt

Image 0 Image 1 Image 2 Image 3



800

30m 59s completed at 10:28 PM

**UPSCALING X4**



200 x 200



800 x 800



200 x 200

# AGENDA

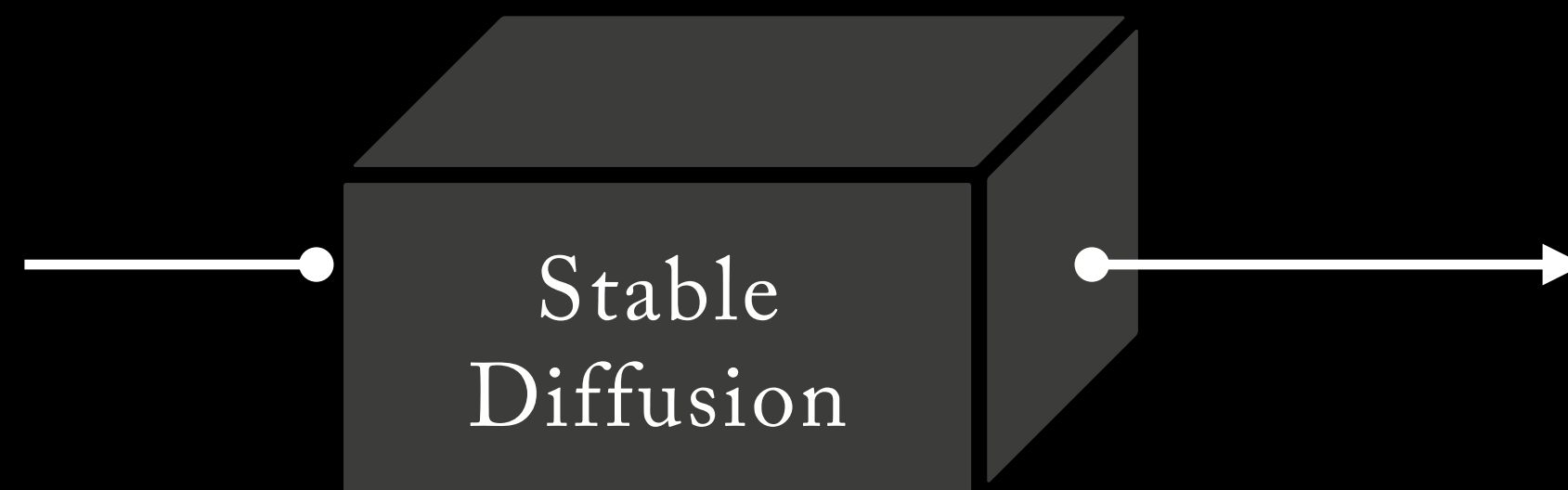
- Overview
- Stable Diffusion
  - OpenCLIP
  - Diffusion models
- Live demo
- Legal aspects
- Summary

# LEGAL ASPECTS

## LEGAL ASPECTS

- Everything looks better with Greg Rutkowski
- Is this his art, or our art?
- Latest Stable Diffusion version:
  - » Now it is harder for users to mimic specific artists

“the most beautiful
panoramic landscape,
oil painting, where a
giant dreamy waterfall
creates a river, the trees
around are starting to bloom,
water shining in the river,
a ray of light of the sunset
by greg rutkowski”



# AGENDA

- Overview
- Stable Diffusion
  - OpenCLIP
  - Diffusion models
- Live demo
- Legal aspects
- Summary

# | SUMMARY

## | SUMMARY

- Stable Diffusion is the state-of-the-art in AI image generation and it has taken the world by storm
- AI image generation points clearly to a shift in how humans create art
- This technology is mainly used in the gaming, film and advertisement industry
- The release of Stable Diffusion is a clear milestone in this development because it made a high-performance model available to the masses
- Regulation will eventually catch up with the technology
- Stable Diffusion itself will not take your job



## THANKS! - SOME USEFUL LINKS

- Model weights:
  - Huggingface: <https://huggingface.co/models>
- Code
  - Stable diffusion v1: <https://github.com/CompVis/stable-diffusion>
  - Stable diffusion v2: <https://github.com/Stability-AI/stablediffusion>
  - Stable diffusion webui: <https://github.com/AUTOMATIC1111/stable-diffusion-webui>
- Prompt Engineering and practical things
  - Stable Diffusions Art: <https://stable-diffusion-art.com/>
- Papers:
  - Stable Diffusion: <https://arxiv.org/abs/2112.10752>
  - CLIP: <https://arxiv.org/abs/2103.00020>
- Another great explanation:
  - <https://medium.com/@steinsfu/stable-diffusion-clearly-explained-ed008044e07e>

# MORE ON AI - GOTHENBURG AI ALLIANCE



## CONFERENCE

MACHINE LEARNING / DATA SCIENCE / DATA ENGINEERING

April 5, 2023 @ Svenska Mässan

CALL FOR SPEAKERS

<https://gaia.fish>