

DALL-E 2

Miguel Sarasa

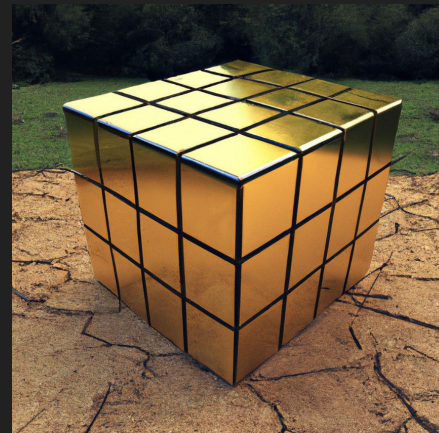
DALL-E 2



cat playing chess looking wise, rococo oil painting



A boy flying through the sky like Superman,
digital art, lens flare



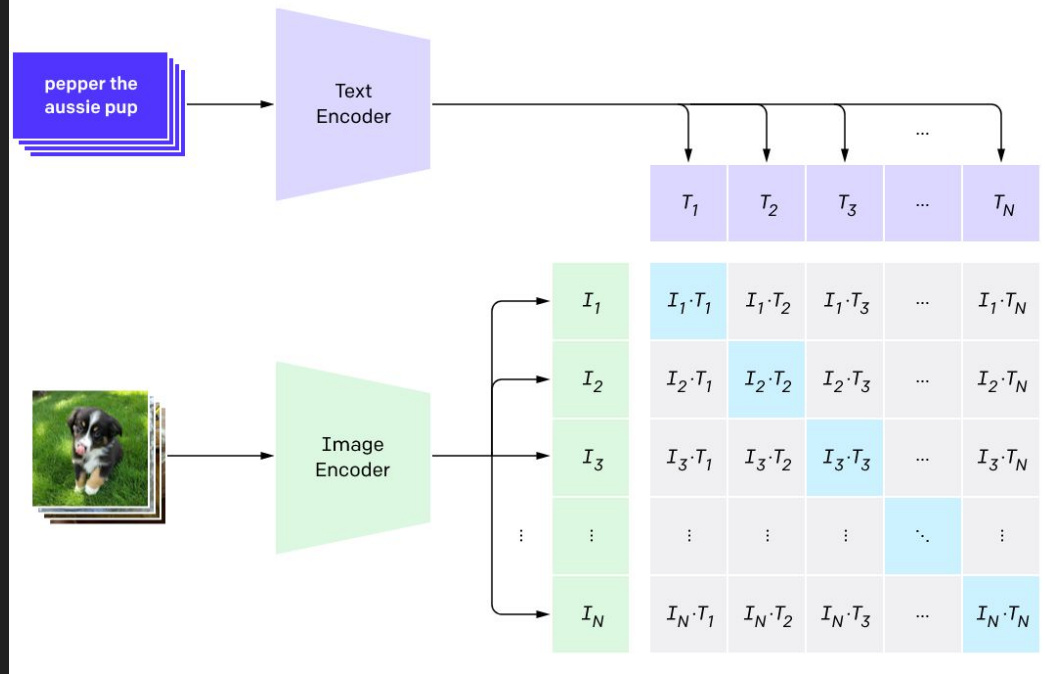
Giant gold Rubik's cube on the ground on
earth, rendering, detailed, zoomed out

Gliederung

- OpenAI CLIP
- OpenAI GLIDE
- Kombination zu DALL-E 2
- Alternativen
- Ethik

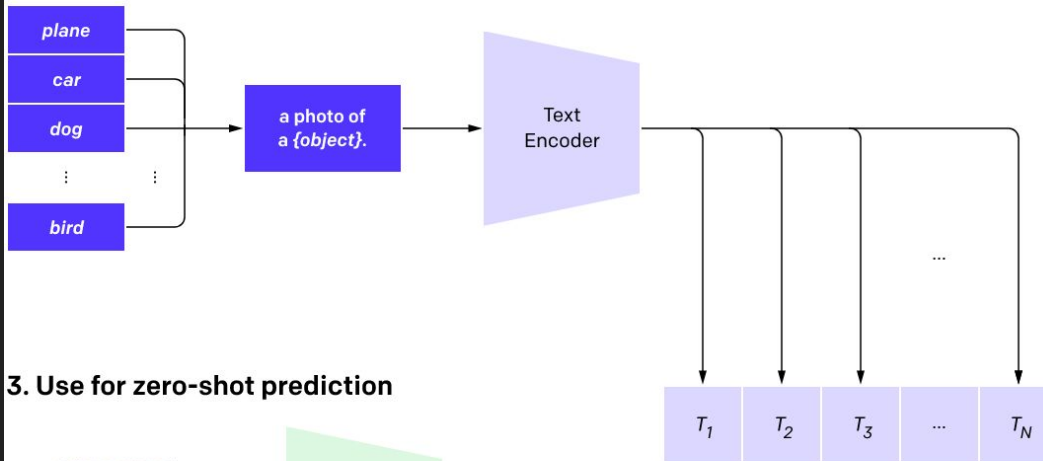
OpenAI CLIP

1. Contrastive pre-training

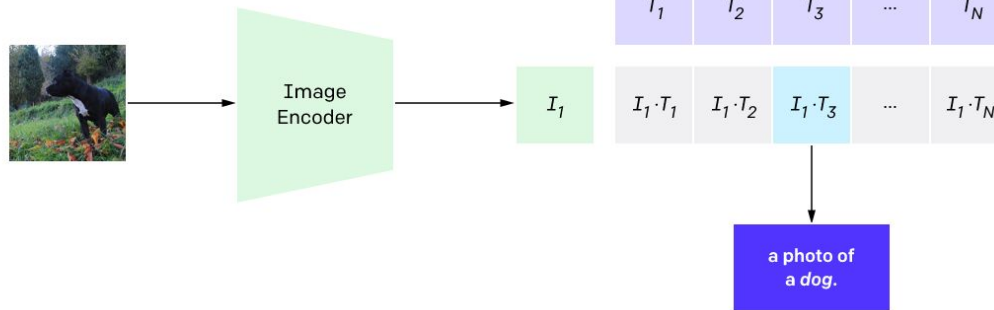


OpenAI CLIP

2. Create dataset classifier from label text



3. Use for zero-shot prediction



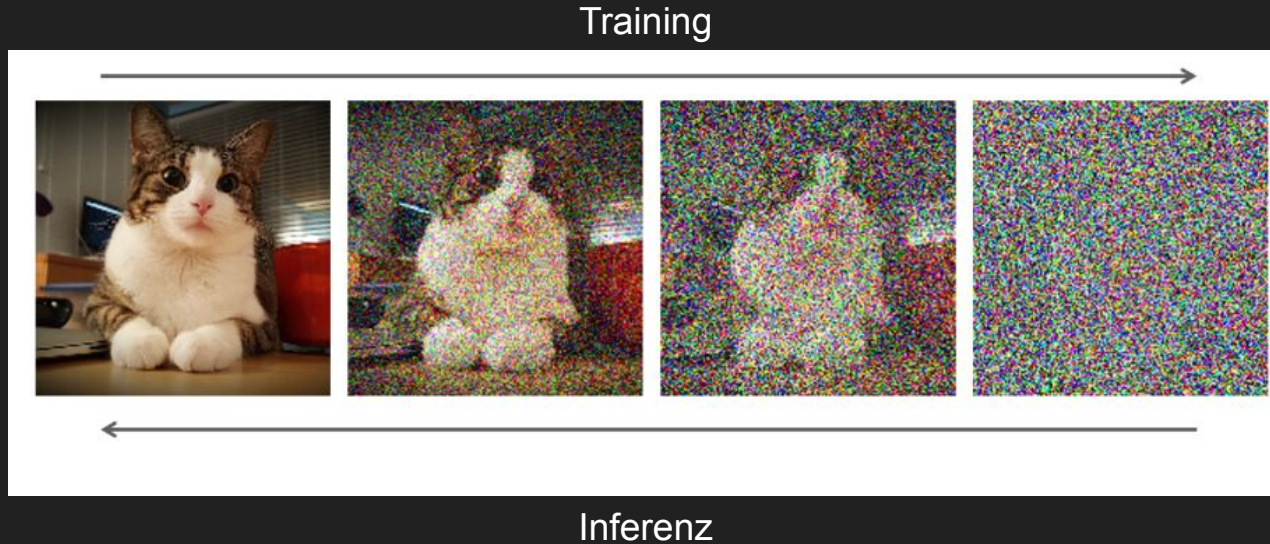
OpenAI CLIP als Suchmaschine

<https://colab.research.google.com/github/haltakov/natural-language-image-search/blob/main/colab/unsplash-image-search.ipynb#scrollTo=ujCKerTnFBk4>

OpenAI GLIDE

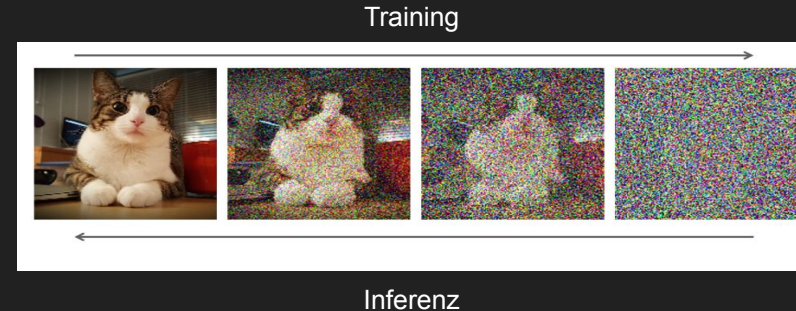
- Modell, das Bilder generieren kann
- Basiert auf einem Diffusionsmodell
- Schwierig, den Output zu steuern
 - Ansatz 1: Textdaten bei Training und Inferenz hinzufügen
 - Ansatz 2: Während der Inferenz mit einem Klassifikator lenken

OpenAI GLIDE: Diffusionsmodelle



OpenAI Glide: Diffusionsmodelle

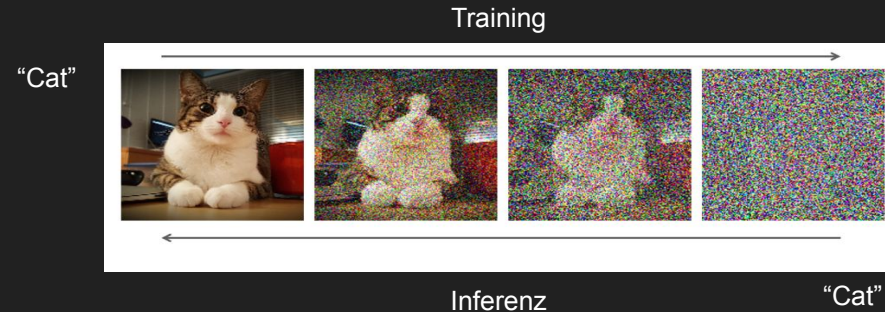
- Training:
 - Schrittweise
 - Dem Bild wird zufälliges Rauschen hinzugefügt
 - Das neuronale Netz lernt dieses Rauschen bei jedem Schritt
 - Am Ende besteht das Bild nur noch aus Rauschen
- Inferenz
 - Schrittweise
 - Das Bild startet als reines Rauschen
 - Das neuronale Netz sagt vorher, welches Rauschen entfernt werden soll
 - Am Ende steht ein rauschfreies Bild



<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

OpenAI GLIDE: Ansatz 1 zur Steuerung

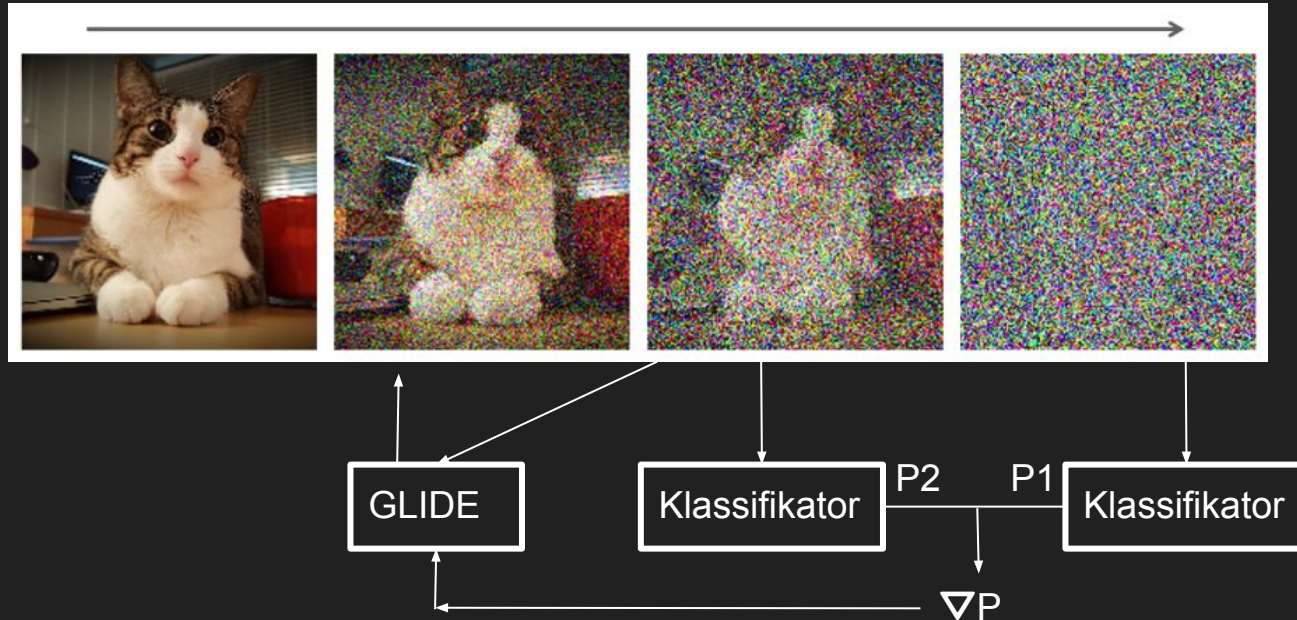
- Beim Training wird neben dem jeweiligen Bild auch ein Text mitgegeben
- Die Inferenz startet mit Rauschen und einem Text
- So soll das Modell Bilder generieren, die zum Text passen
- Problem: GLIDE ignoriert die Texte einfach



<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

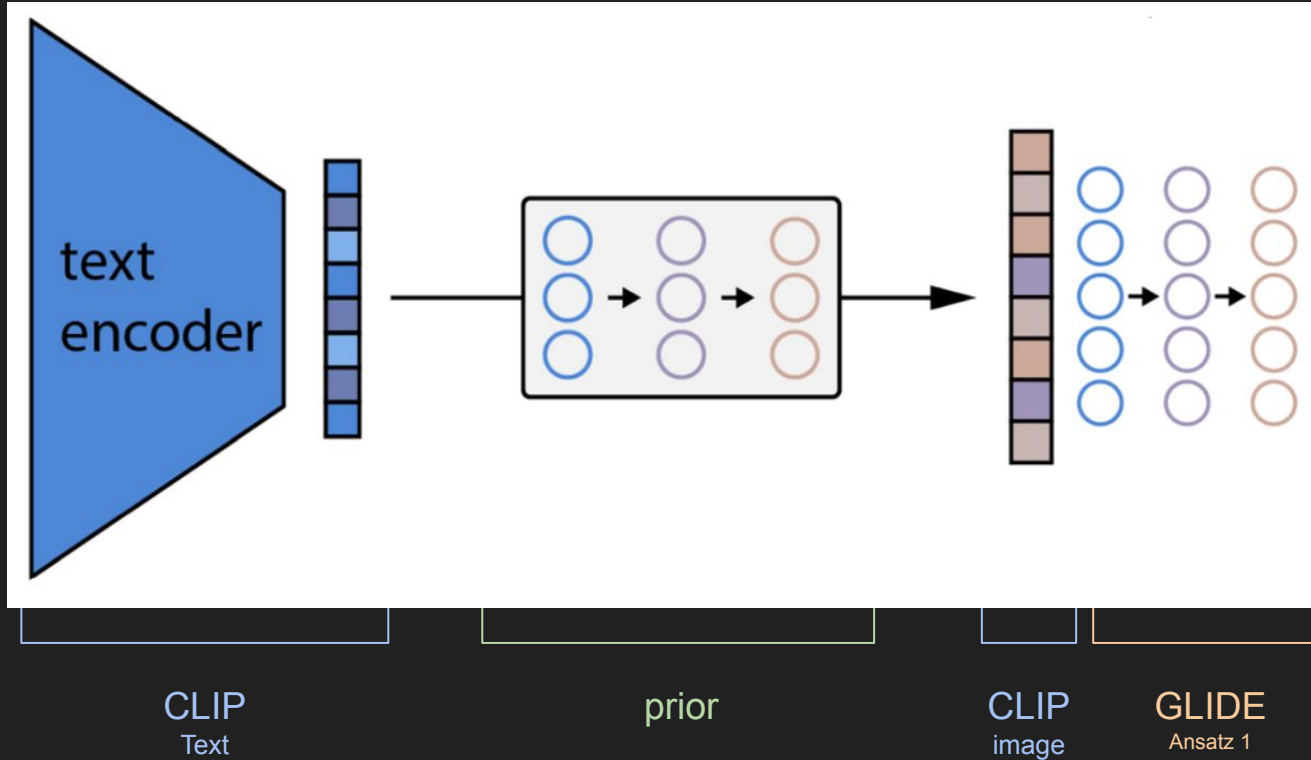
OpenAI GLIDE: Ansatz 2 zur Steuerung

Training



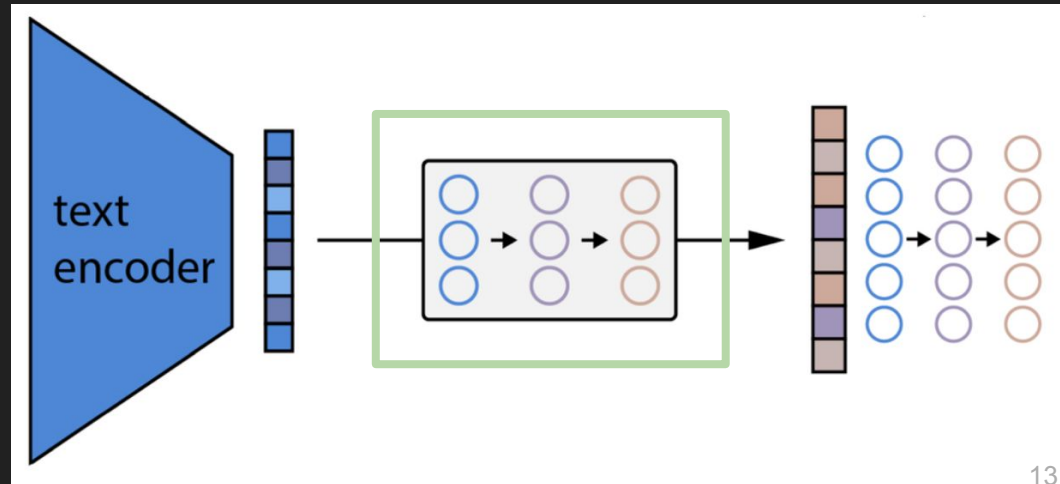
<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

DALL-E 2



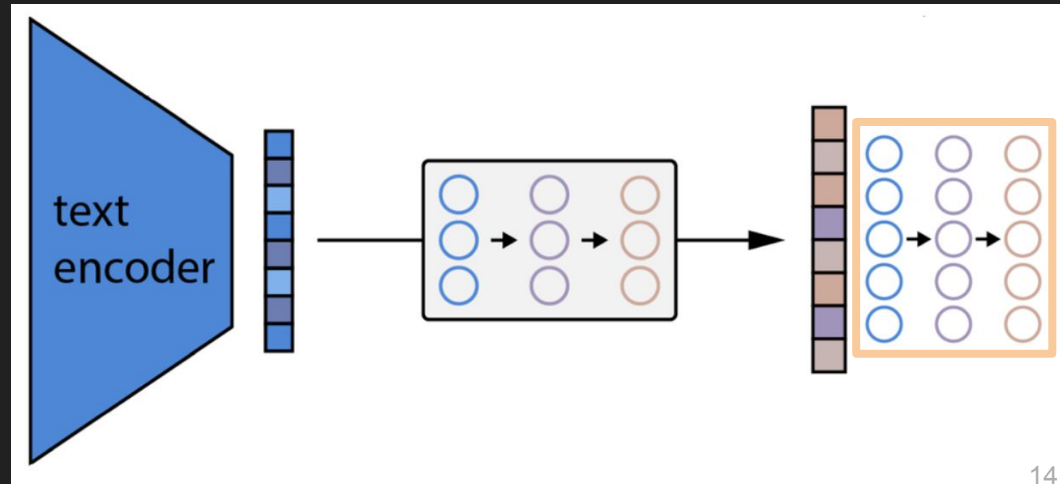
DALL-E 2: prior

- Neuronales Netz
- Übersetzt CLIP Text embeddings zu CLIP Image embeddings



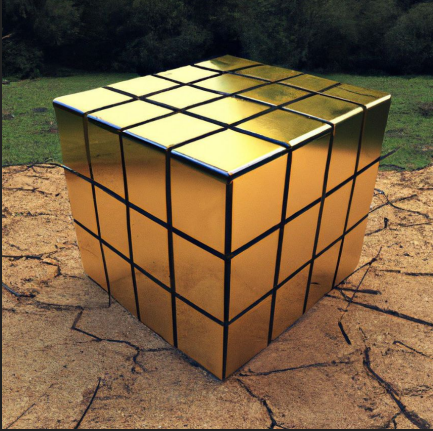
DALL-E 2: Decoder

- Modifizierter GLIDE mit Ansatz 1 zur Steuerung
 - GLIDE Paper: Direkte Steuerung durch Textdaten
 - DALL-E 2 Paper: Steuerung durch CLIP image Embeddings
- GLIDE ignoriert die CLIP Image embeddings nicht mehr



Alternativen

Giant gold Rubik's cube on the ground on earth, rendering, detailed, zoomed out



DALL-E 2

- Keine Pornographie
- Kein Gore
- Keine realistischen Gesichter



Midjourney

- Keine Pornographie
- Kein Gore



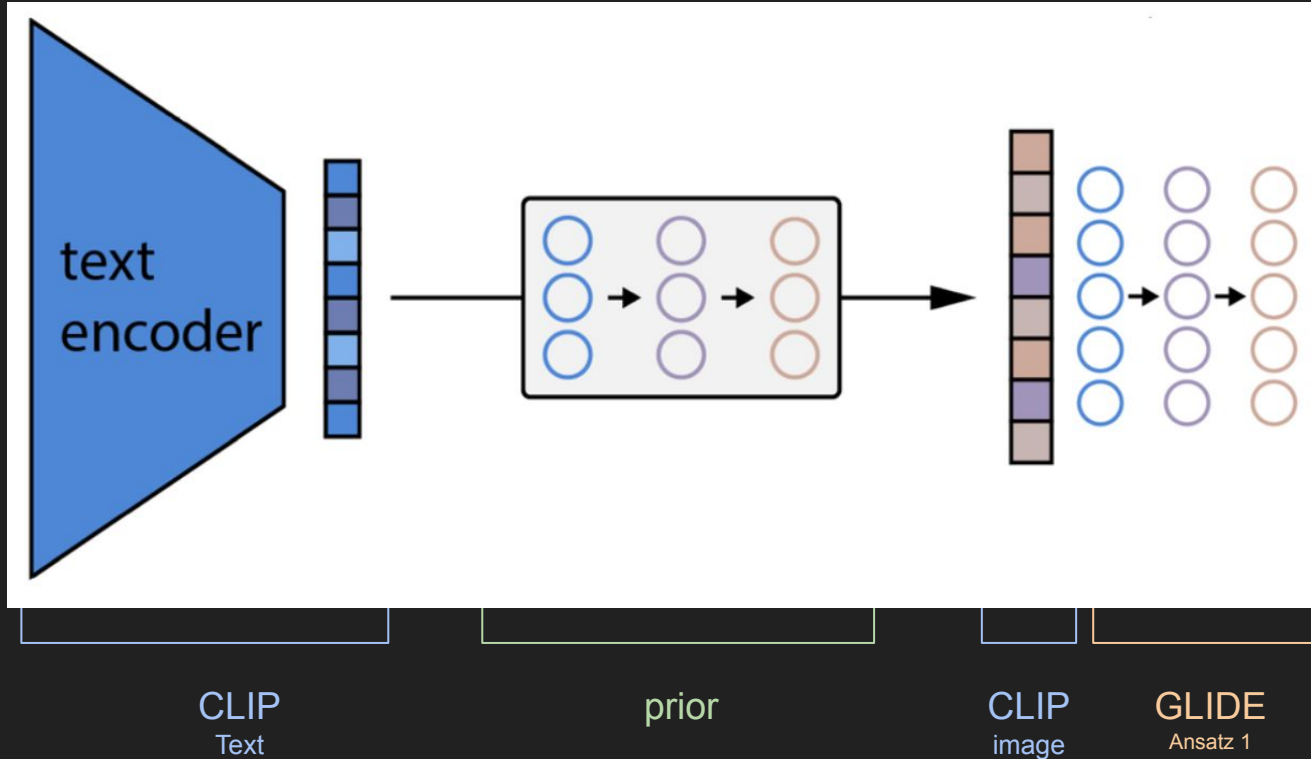
Craiyon (DALL-E Mini)

- Keine Regeln

Ethik

- Fake News
- Anschwärzen
- Bildmaterial als Beweis wird unzuverlässig
- Kinderpornographie

Recap DALL-E 2



Quellen

[The best examples of DALL-E 2's strange, beautiful AI art](#)

[Hierarchical Text-Conditional Image Generation with CLIP Latents](#)

[lucidrains/DALLE2-pytorch](#)

[Diffusion Models Beat GANs on Image Synthesis](#)

[Unsplash Image Search](#)

[Zero-Shot Text Classification with Hugging Face](#)

[How does DALL-E 2 actually work?](#)

[DALL-E Content Policy](#)

[DALL-E 2 Pretraining mitigations](#)