

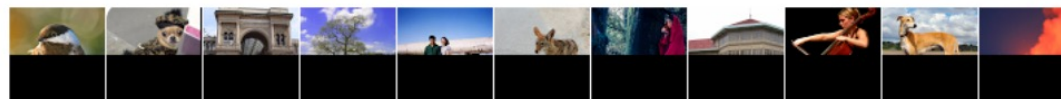
Генеративные трансформеры DALL-E

Семенова Елена БПМИ-182

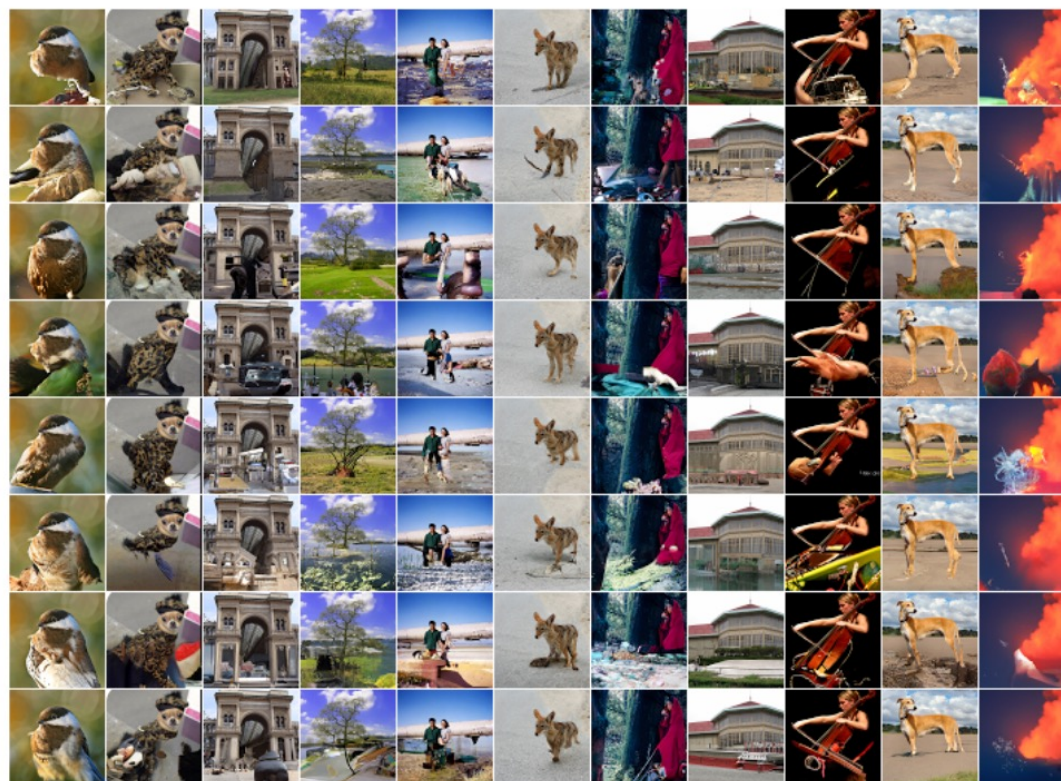
Sparse Transformers (2019)



Sparse Transformers (2019)



Prompt



Completions



Ground truth

NLP & CV

Успехи в NLP

BERT

GPT-1,2,3

Но они же работают на
последовательностях!

Применяем к CV

Image GPT

DALL-E

Image GPT (2020)

- Подаются замаскированные изображения
- GPT-2 (до 1.5 млрд параметров) – 2019
- Низкие разрешения – 32x32, 48x48, 64x64

Image GPT (2020)

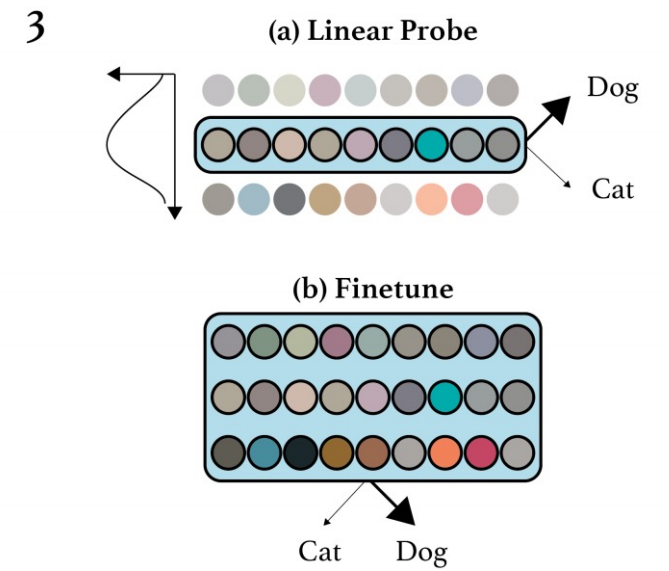
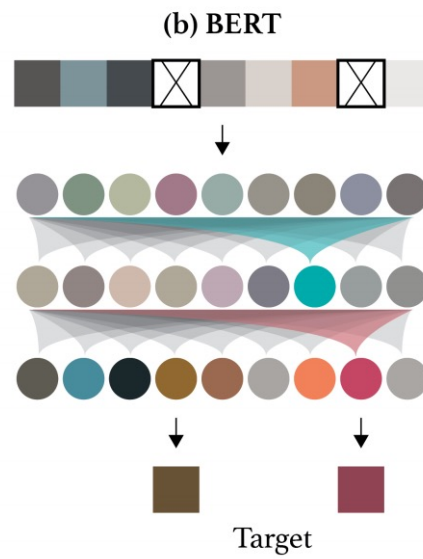
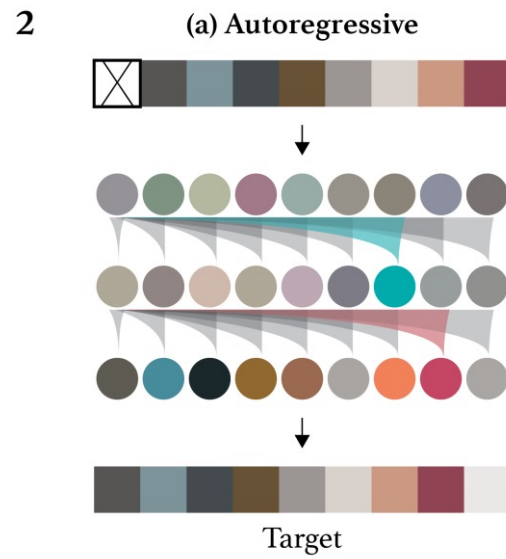
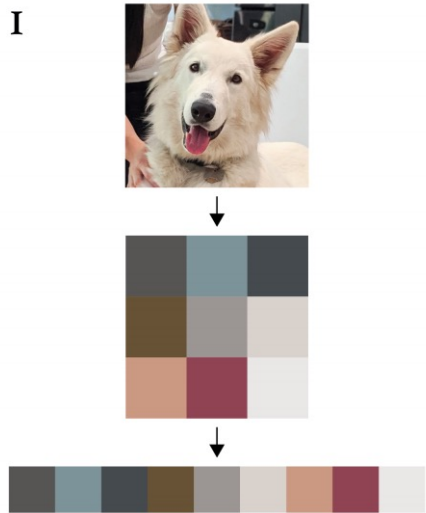


Image GPT (2020)

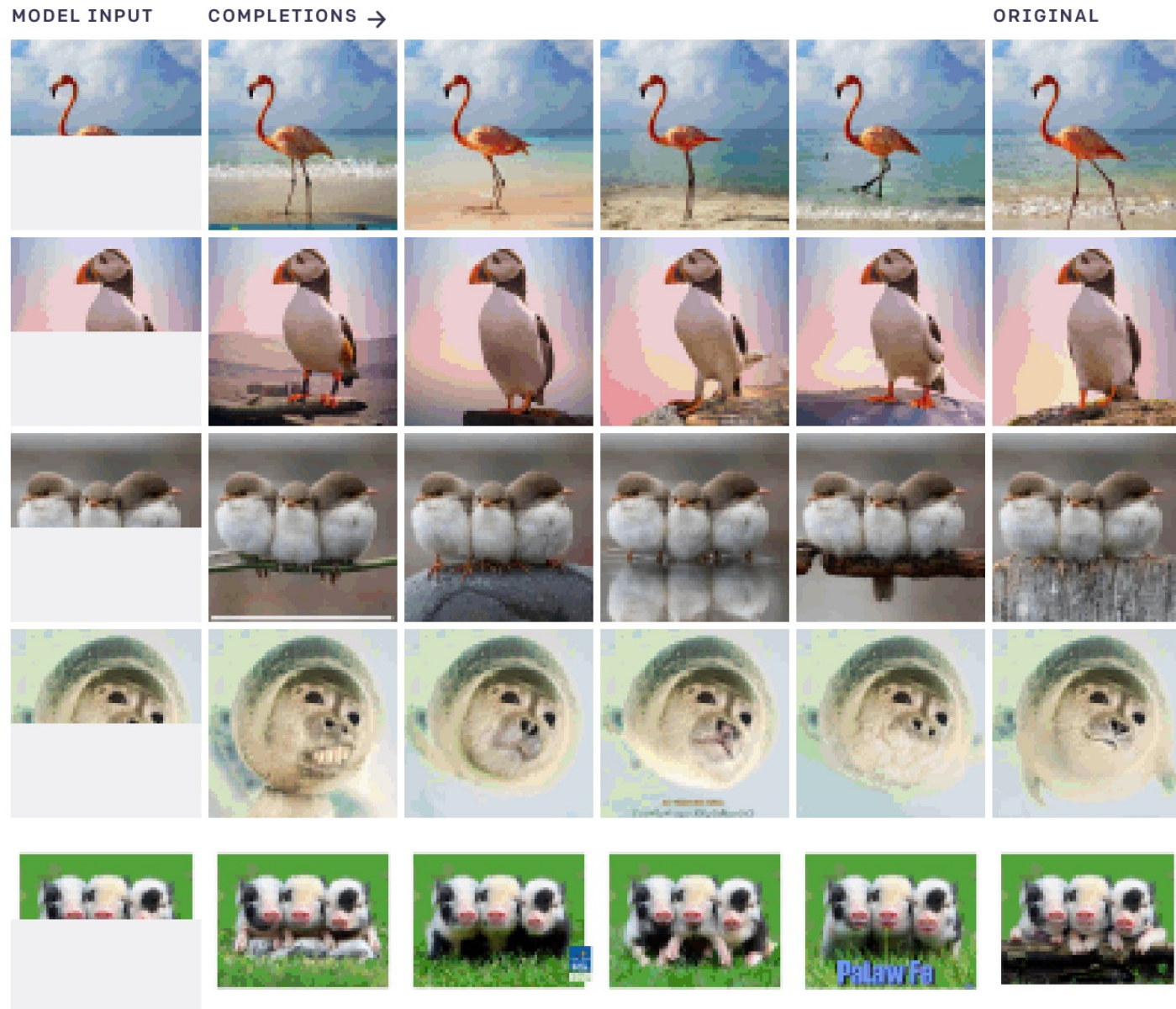
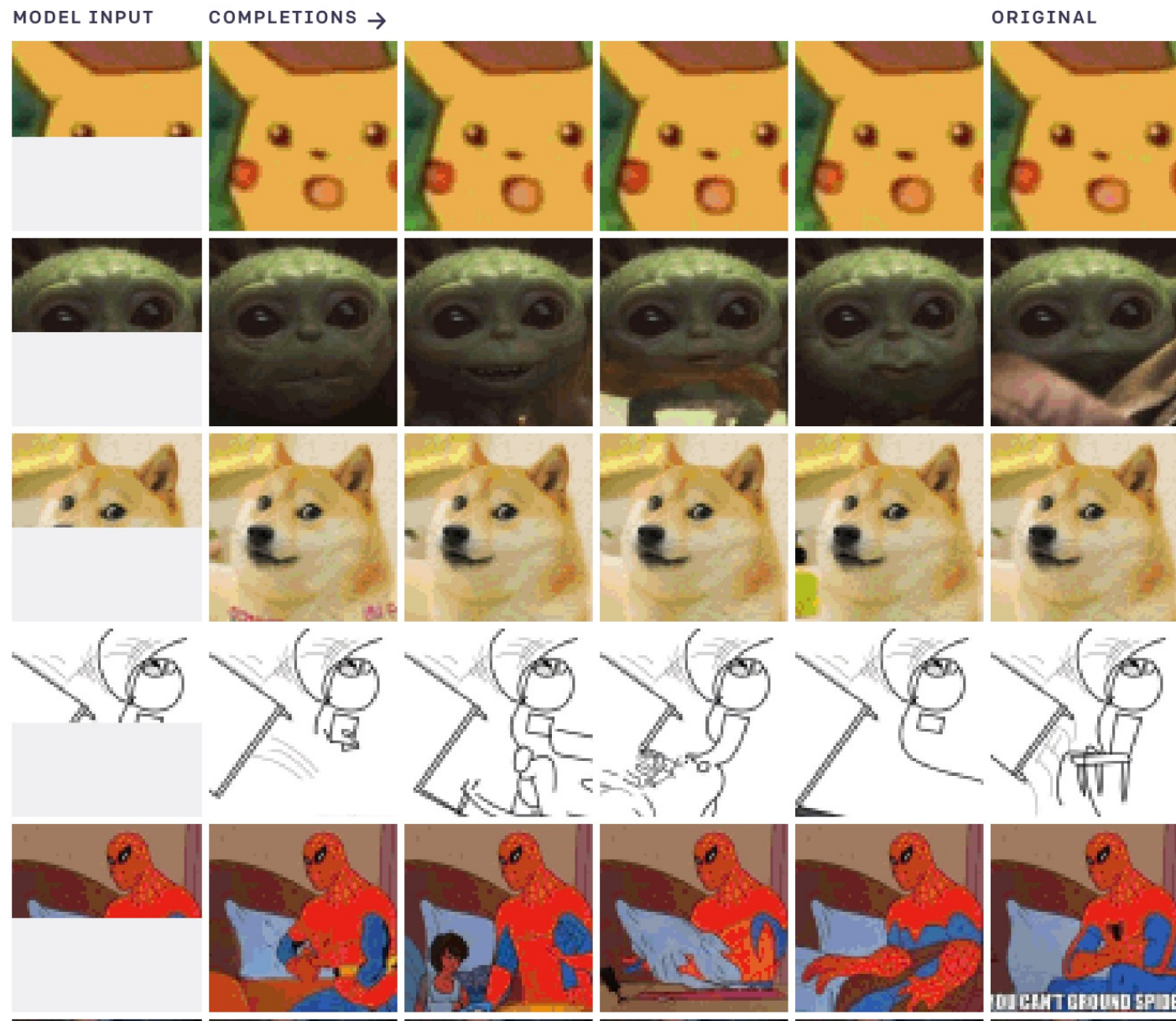


Image GPT (2020)



Еще идеи?

А давайте еще больше NLP!

DALL-E? (2021)

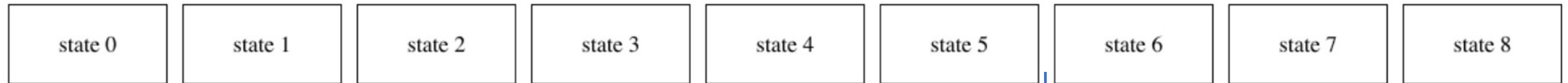
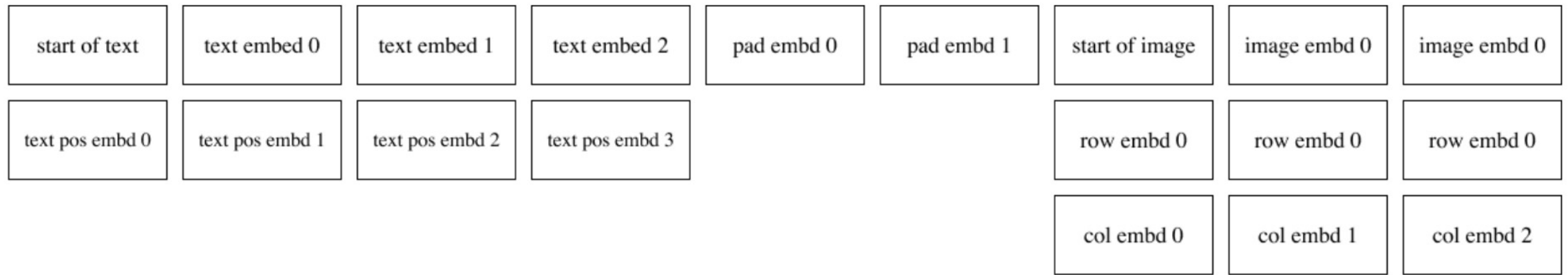


DALL-E (2021)

- Подаются текстовые описания (или не только)
- GPT-3 (до 175 млрд параметров) – 2020
- Используется DVAE (Discrete VAE)
- Более высокие разрешения – 256x256

DALL-E (2021)

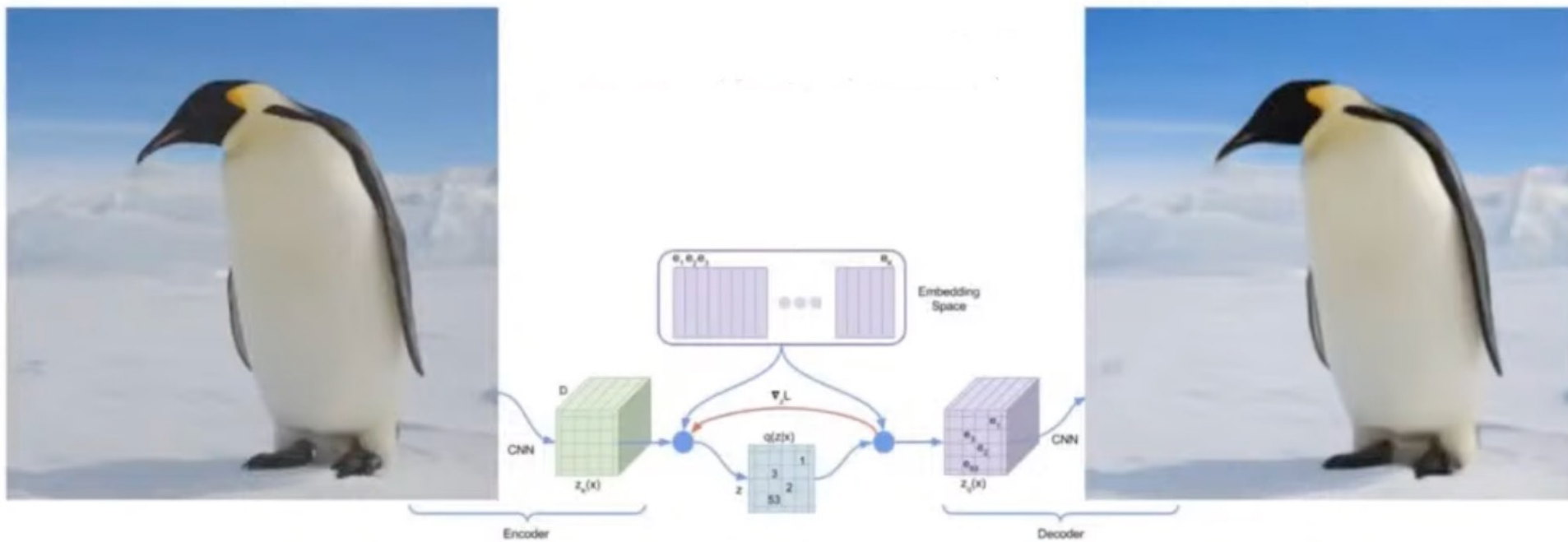
+



256 токенов для текста

256x256 токенов для
изображения? - TOO MUCH

DVAE



256x256



32x32

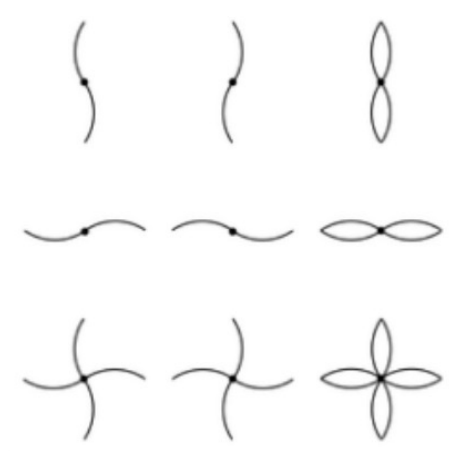
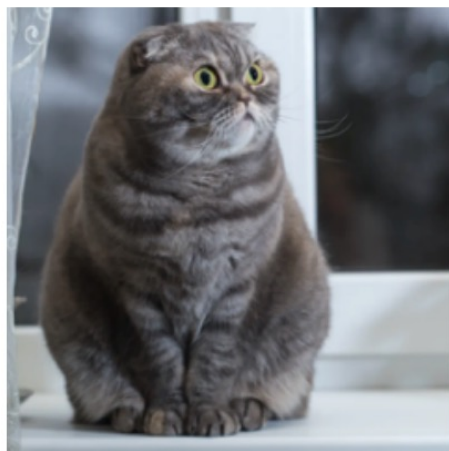


256x256

Подаем на вход

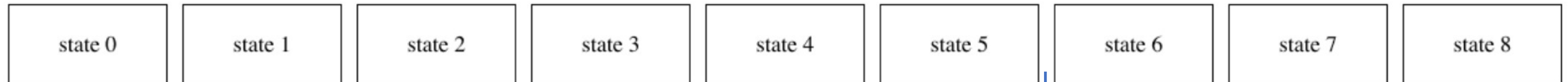
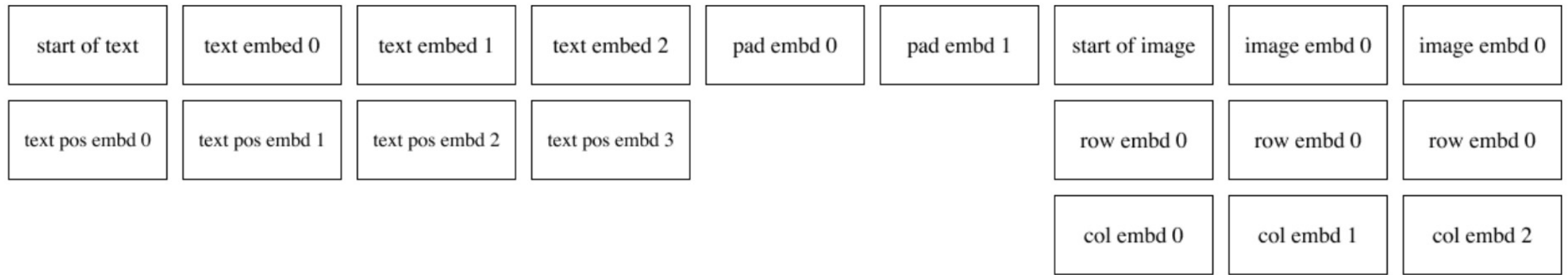
DVAE

Теряется точность изображения



DALL-E (2021)

+



256 токенов для текста

32x32=1024 токенов для
изображения

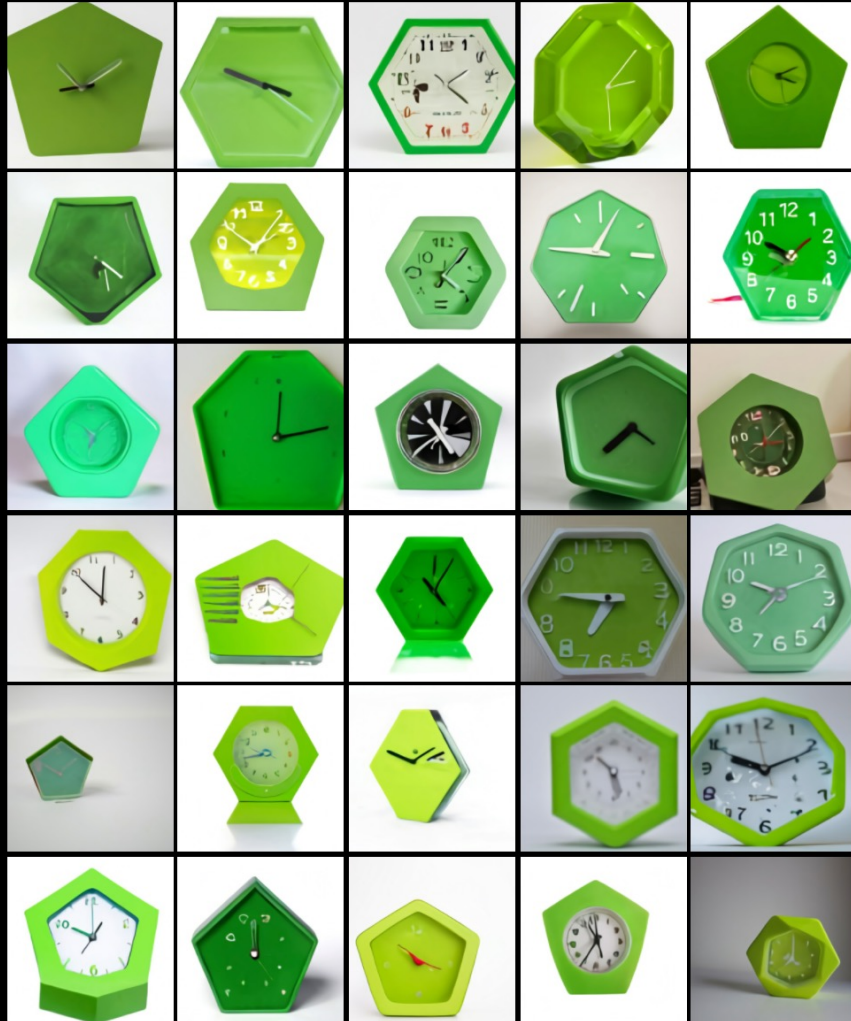
DALL-E (2021)

1. Применяется DVAE для сжатия изображения
2. Сжатое изображение превращается в одномерную последовательность
3. Текстовое описание и сжатое изображение объединяются (256 и 1024 токенов)
4. Общая последовательность подается авторегрессионному трансформеру

DALL-E (2021)

TEXT PROMPT a pentagonal green clock. a green clock in the shape of a pentagon.

AI-GENERATED
IMAGES

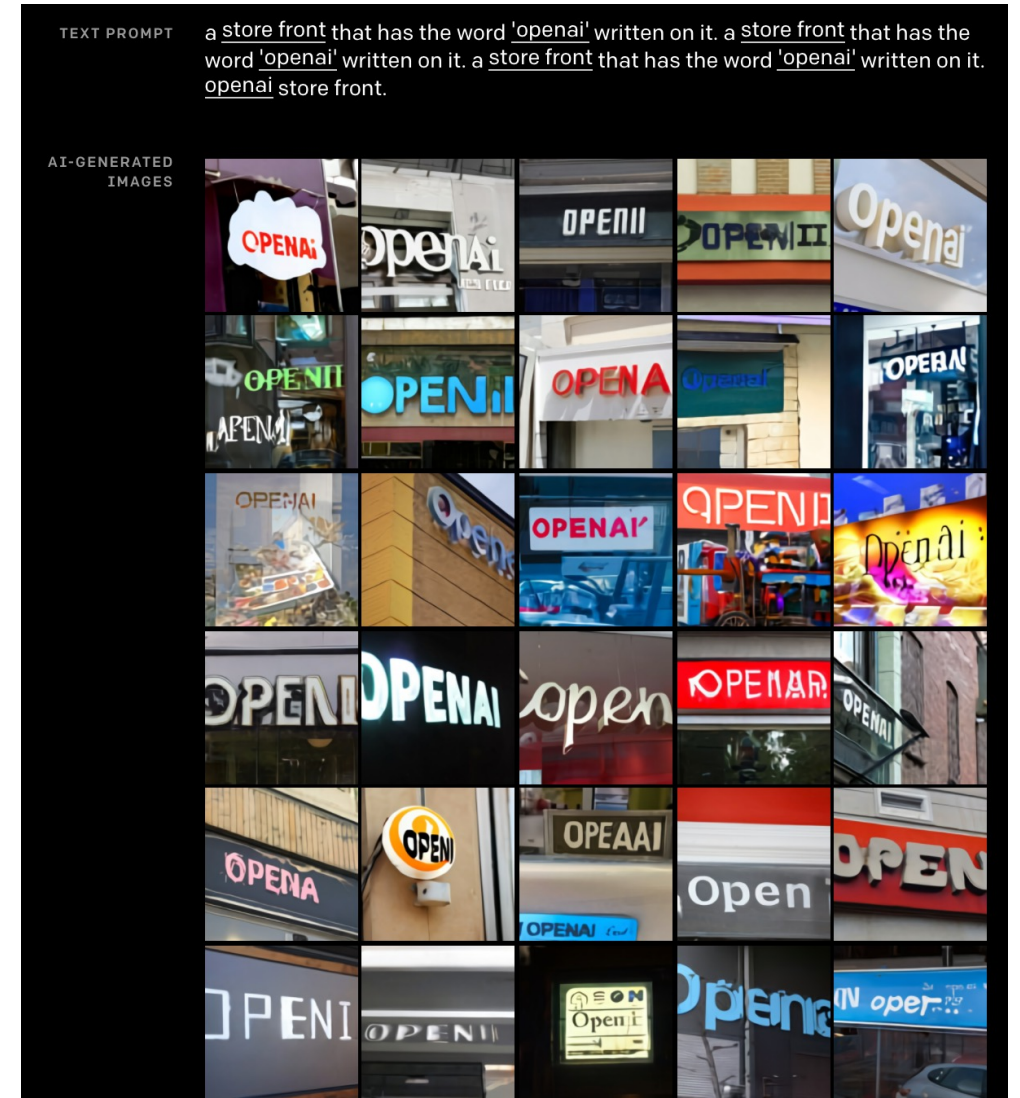
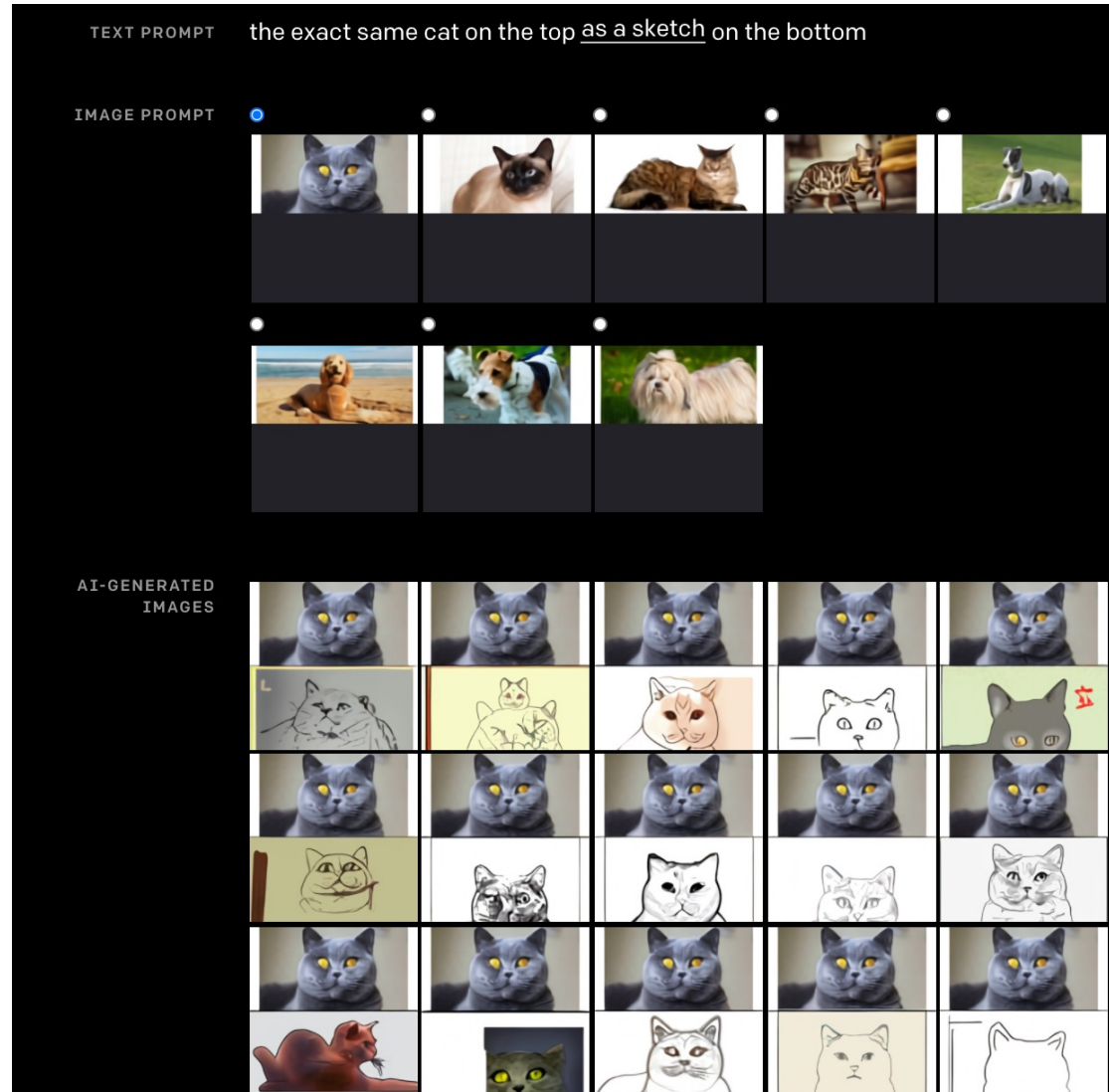


TEXT PROMPT a snail made of harp. a snail with the texture of a harp.

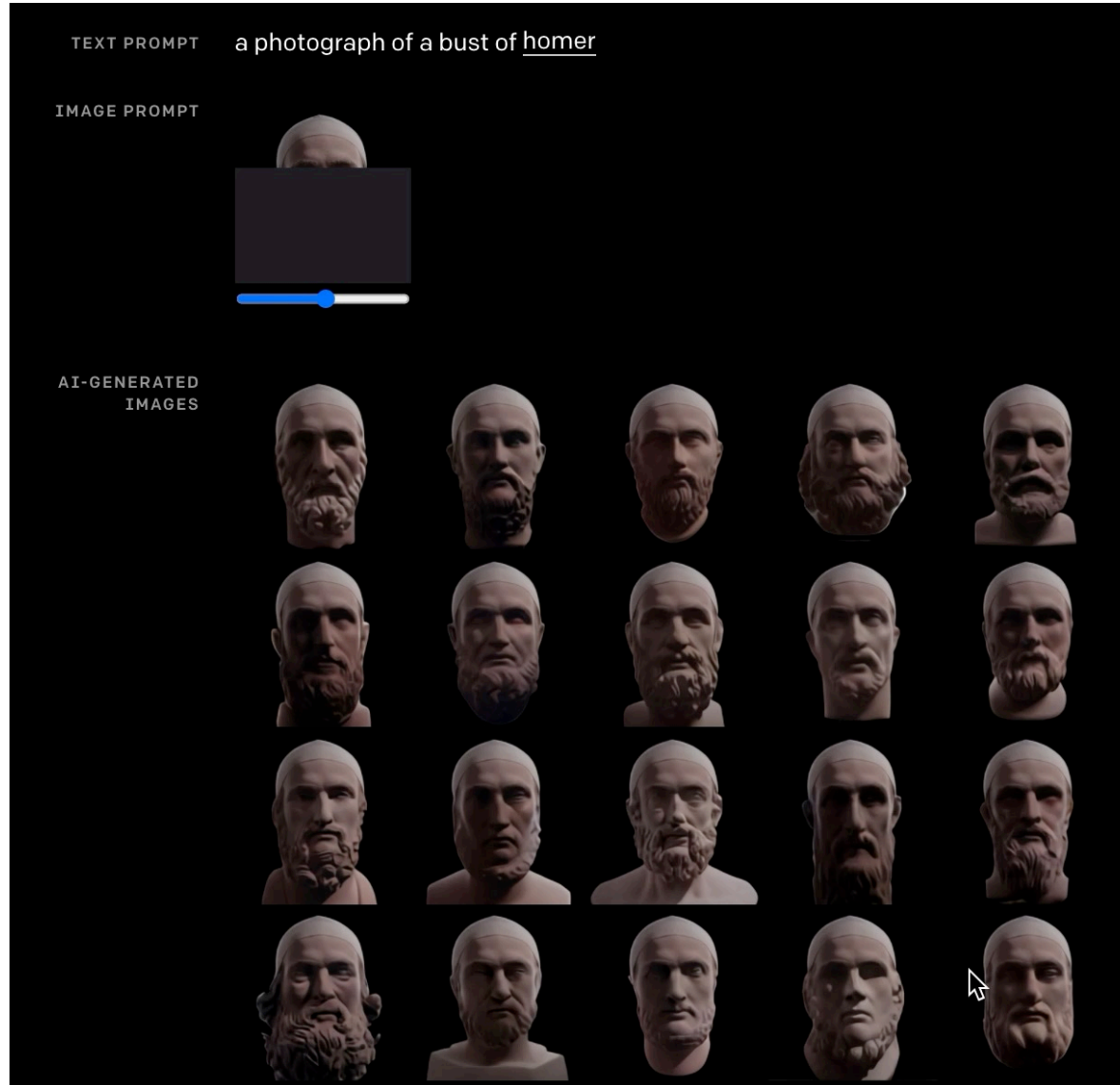
AI-GENERATED
IMAGES



DALL-E (2021)



DALL-E (2021)



ИСТОЧНИКИ

Блог OpenAI

1. Sparse Transformers: <https://openai.com/blog/sparse-transformer/>
2. Image GPT: <https://openai.com/blog/image-gpt/>
3. DALL-E: <https://openai.com/blog/dall-e/>

Статьи

1. [https://cdn.openai.com/papers/Generative Pretraining from Pixels V2.pdf](https://cdn.openai.com/papers/Generative_Pretraining_from_Pixels_V2.pdf)
2. <https://arxiv.org/pdf/2102.12092.pdf>