

Обработка естественного языка. Лекция 8

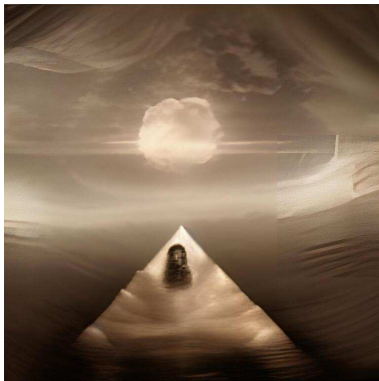
Изображение и звук

Михаил Пожидаев

24 октября 2023 г.

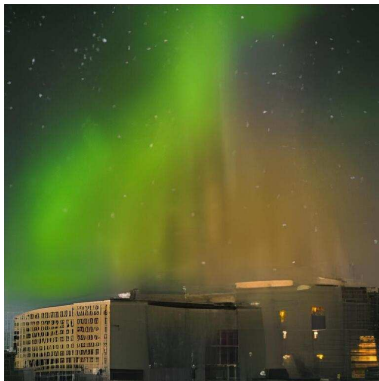
Картинка

«Луч света в тёмном царстве»



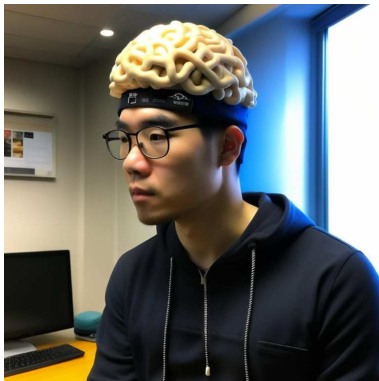
Картинка

«Северное сияние над университетом»



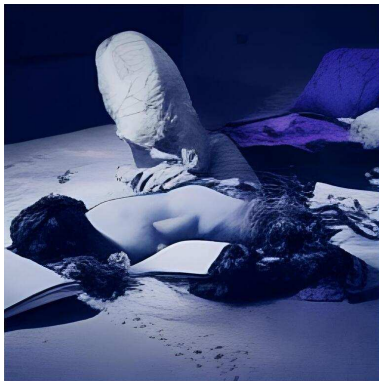
Картинка

«Мозг студента перед сессией»



Картинка

«Кошмарный сон студента»



Диффузионный процесс

Обучение на основе анализа добавленного шума

Бесконтрольное зашумление

Для набора изображений итерационно добавляется шум с некоторым распределением. Модель запоминает различия между исходным и зашумлённым изображениями, после чего становится способна производить обратную операцию.

Контролируемое зашумление

Блоки анализатора добавленного шума могут быть снабжены дополнительными блоками внимания, которые связывают процесс анализа шума с некоторым закодированным текстом.

Stable diffusion

Генерация изображений по тексту

Модель Stable diffusion содержит три блока:

1. Кодер текста на основе Трансформера (допустимо использовать BERT или GPT). Отвечает за «понимание» текста.
2. Диффузионная модель, обрабатывающая шум в промежуточном пространстве описаний изображений с учётом результатов кодера текста. Отвечает за связь текста с изображением.
3. Генератор изображений, «рисующий» картинку по описанию.

MFCC

Мел-кепстральные коэффициенты

мел

Единица высоты звука, основанная на восприятии этого звука органами слуха.

В окне сигнала с преобразованием Фурье

Небольшой набор значений (вектор), который качественно характеризует речевой сигнал для фрагмента (окна) в записи.

Tacotron 2

Seq2seq для синтеза речи

Feature prediction net

Составной кодер, который обрабатывает последовательность символов для построения вектора скрытого состояния.

NN-vocoder WaveNet

Декодер, который превращает выходной вектор скрытого состояния в мел-спектрограмму.

Feature prediction net и NN-vocoder WaveNet содержат дополнительную связь на основе механизма внимания.

Feature prediction net

Слой кодера в Tacotron2

- ▶ Первый слой: Embedding-слой для построения векторного представления текста.
- ▶ Второй слой: три одномерных свёрточных ячейки с последующей нормализацией и обработкой функцией активации.
- ▶ Третий слой: обработка LSTM в двух направлениях с конкатенацией получившегося результата.

Спасибо за внимание!

Всё о курсе: <https://marigostra.ru/materials/nlp.html>

E-mail: msp@luwrain.org

Канал в Телеграм: <https://t.me/MarigostraRu>