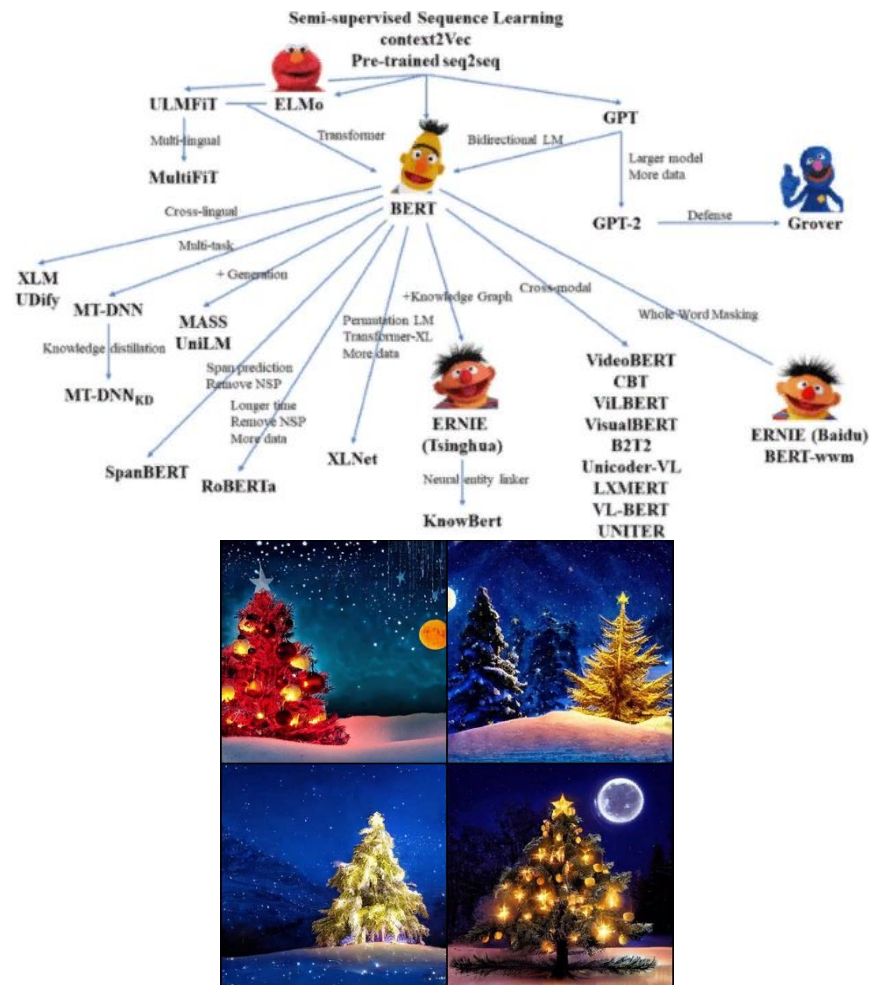


# Современное NLP

# Основные идеи

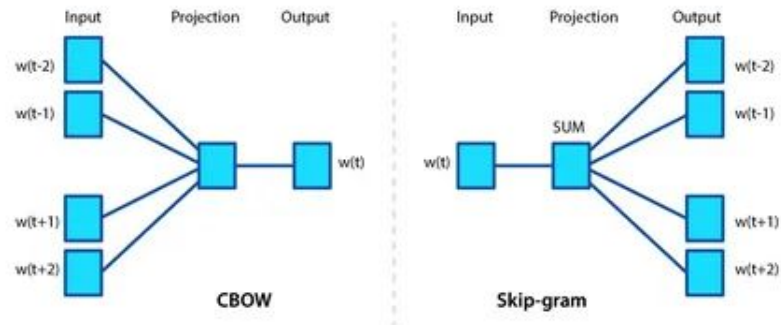
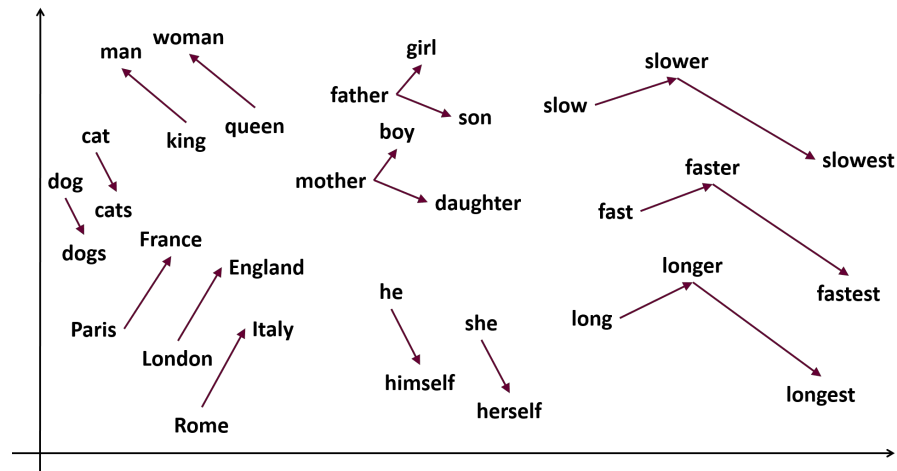
- почти целиком и полностью на основе глубоких нейронных сетей
- балом правят трансформерные модели (с механизмом внимания)
- обучаются на суперкомпьютерах и имеют до нескольких миллиардов параметров (175 в GPT-3)
- после (пред)обучения могут быть легко адаптированы под конкретные задачи (few- и zero-shot learning)
- объединяются с моделями из компьютерного зрения с впечатляющими результатами
- отрабатывают на сравнимом с человеческим уровне



Рождественская ночь с красивым месяцем и нарядной ёлкой

# Вспоминаем word2vec

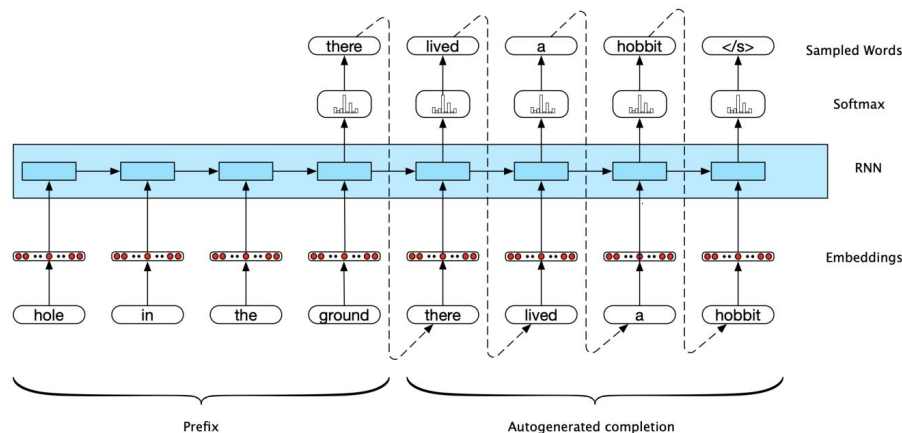
- BoW и TF-IDF накладывали на нас многочисленные ограничения
- Хотелось бы научиться представлять слова векторами так, чтобы над этими векторами можно было выполнять осмысленные операции (складывать, например)
- Делать это можно 2-мя похожими способами: CBOW и skip-gram
- Обучение – пример Self-Supervision подхода



# Говорим про рекуррентные нейросети (RNNs)

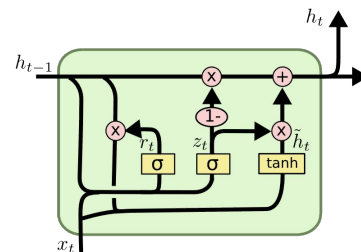
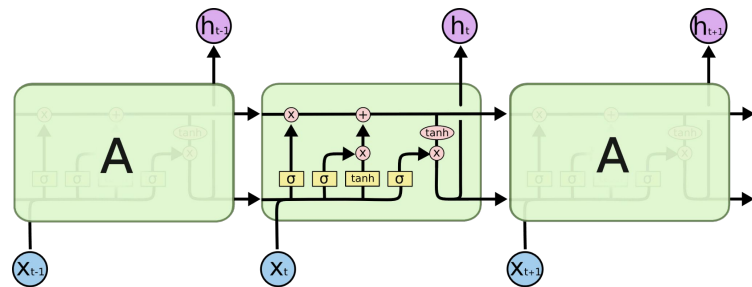
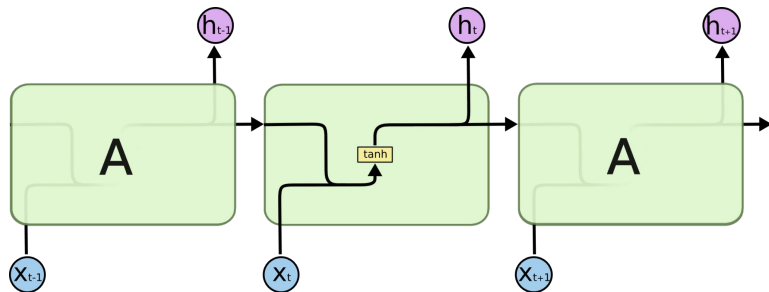
**Мотивация:** тексты (как и другие последовательности) устроены так, что нам хотелось бы время от времени обращаться к прошлой информации

- есть один нюанс – нейросети обучаются back-prop'ом: если последовательность длинная (а так почти всегда и бывает), то градиенты будут затухать и обучаться особо ничему не получится
- есть выход – будем накапливать информацию “умно”: например, LSTM



# Память в RNN на примере LSTM

- самая полезная штука на диаграмме – верхняя стрелочка; благодаря ней мы можем пробрасывать предыдущие состояния без изменений
- есть вариация – GRU



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

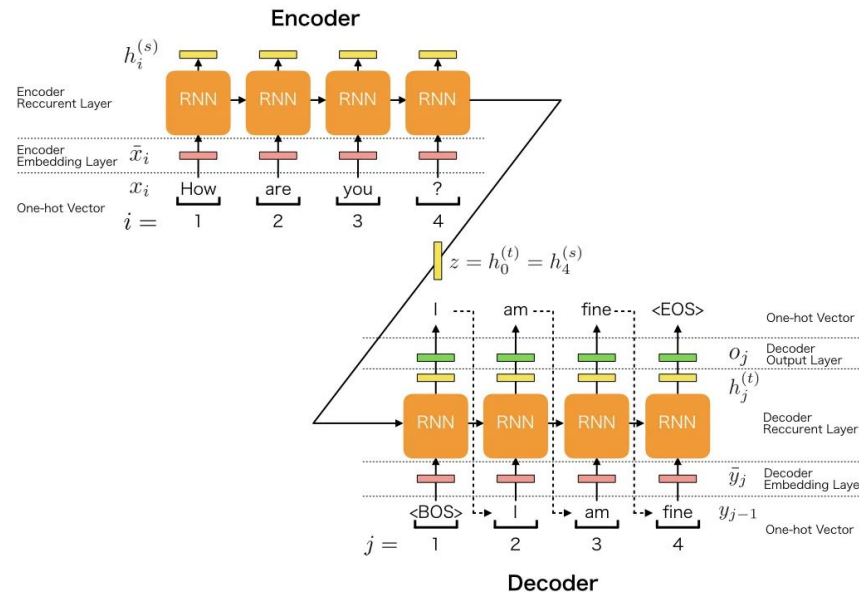
$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

# seq2seq задачи

- есть seq2seq задачи, в которых длины входной и выходной последовательностей не совпадают (например, машинный перевод)
- придумали Encoder-Decoder архитектуру
  - Скрытое состояние после всего текста — «контекст»;
  - Контекст передаётся в RNN, которая генерирует выходной текст + BeamSearch



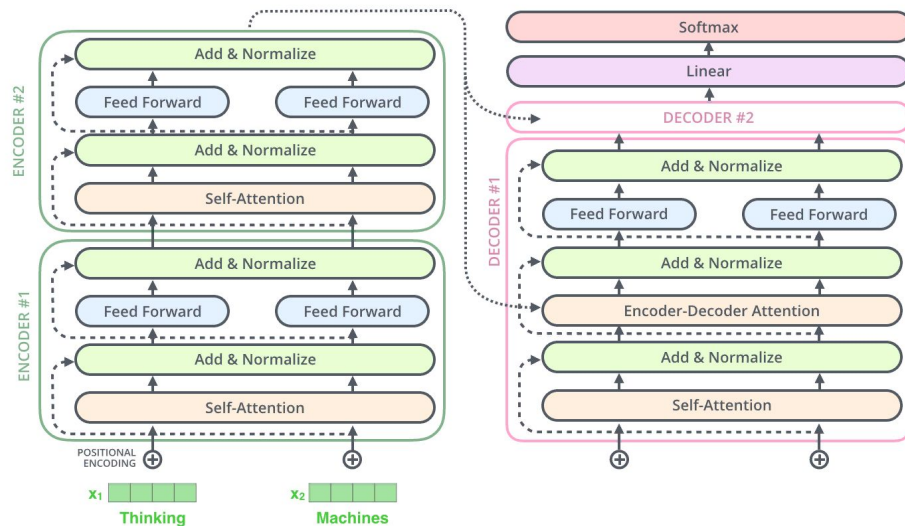
**Проблема:** все равно будем терять информацию о первых словах (или о середине, если используем BiLSTM)

# Механизм внимания

- а давайте смотреть на всю входную последовательность целиком
- но особое внимание будем уделять наиболее значимым словам (Attention)
- идёмте смотреть :) <https://jalammar.github.io/illustrated-transformer/>

## Важные концепты:

- Self-Attention
- Positional Encoding



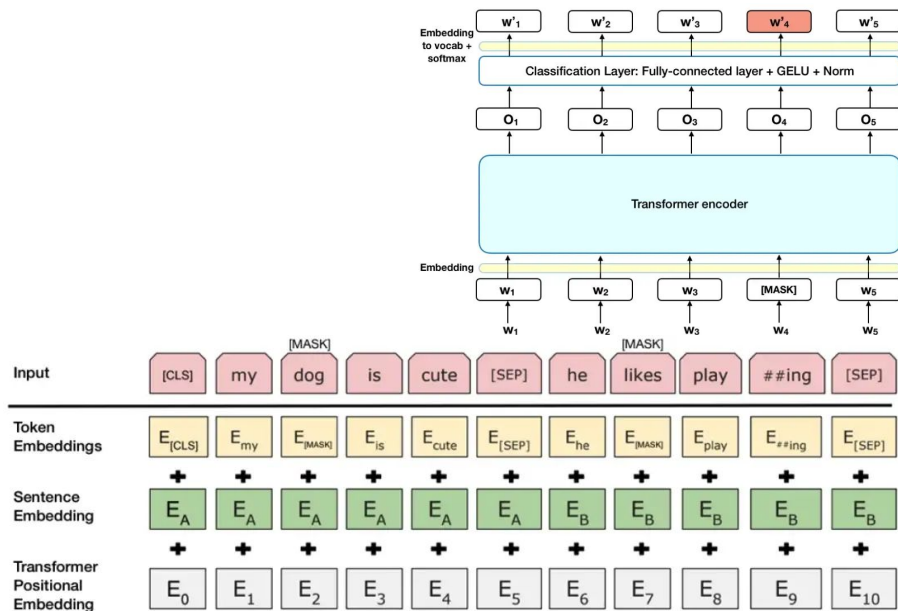
# BERT

Основное улучшение в сравнении с классическим трансформером –  
двунаправленное обучение (за счет Masked Language Model)

**MLM:** • перед обучением 15% токенов  
заменяются на [Mask]

• затем хотим их восстановить

**NSP:** • во время обучения модель  
получает пары предложений, её задача –  
“понять”, относиться ли второе  
предложение к первому

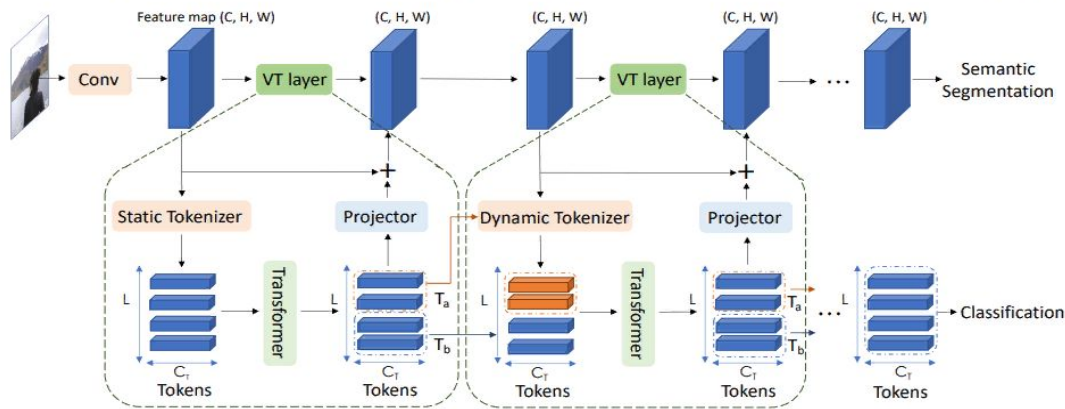
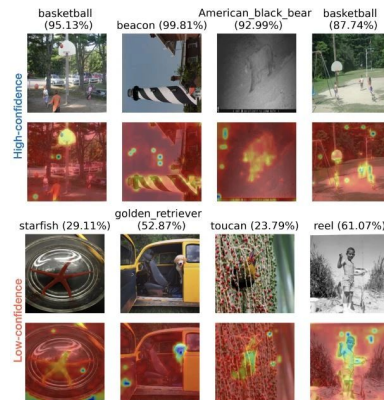




# Заканчиваем с трансформерами. NLP only?

Нет.

- ❑ [Visual Transformer](#)
- ❑ Interpreting RecSys Recommendations
- ❑ [Decision Transformer \(for RL\)](#)
- ❑ [Transformers for Large Scale Audio](#)

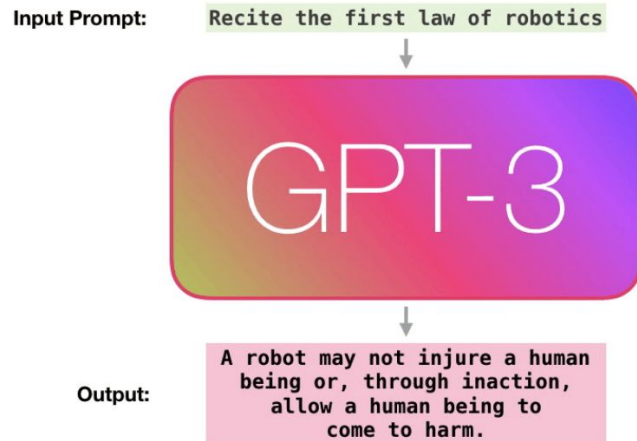


# GPT-3

- Контекст в 2048 токенов
- 175 миллиардов параметров (800 Gb для хранения)
- “Generative pre-training” значит, что модель обучается предсказывать следующий токен

## Применения:

- Overall: Text Generation
- GitHub Copilot
- <https://russiannlp.github.io/rugpt-demo/>



# Совсем молодая ChatGPT

расскажи анекдот про openai

Раз в OpenAI пришел один сотрудник и сказал: "Я хочу сделать ИИ, который угадывает желания человека". На что его коллега ответила: "А что, ГПТ-3 недостаточно?".

Напиши пожалуйста сочинение на тему нравственного выбора от имени талантливого школьника 10-го класса

Нравственный выбор – это важное решение, которое мы должны принимать каждый день. Иногда оно может быть сложным, потому что мы должны выбрать между тем, что нам нравится и тем, что правильно. Но я уверен, что если мы постоянно стремимся к хорошему и правильному, мы сможем делать правильные выборы.

В моей школе мы изучаем много разных предметов, но самым важным, по моему мнению, является этика. Этот предмет помогает нам поня

<https://openai.com/blog/chatgpt/>

[https://t.me/mashkka\\_ds](https://t.me/mashkka_ds)

# Dall-E и prompt-tuning

GPT-3 как языковая модель

Диффузия для генерации картинок

Между ними нейросеть CLIP,  
наученная линковать текстовое описание и картинки

- prompt-tuning – новый важный концепт в DL

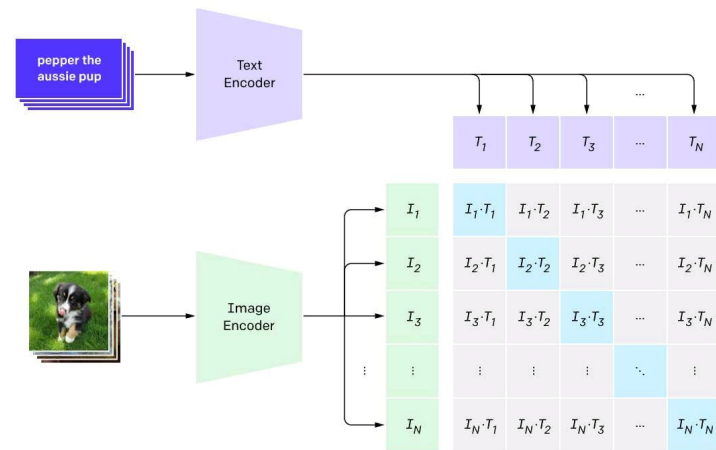


Фото грабителя

A photo of a burglar

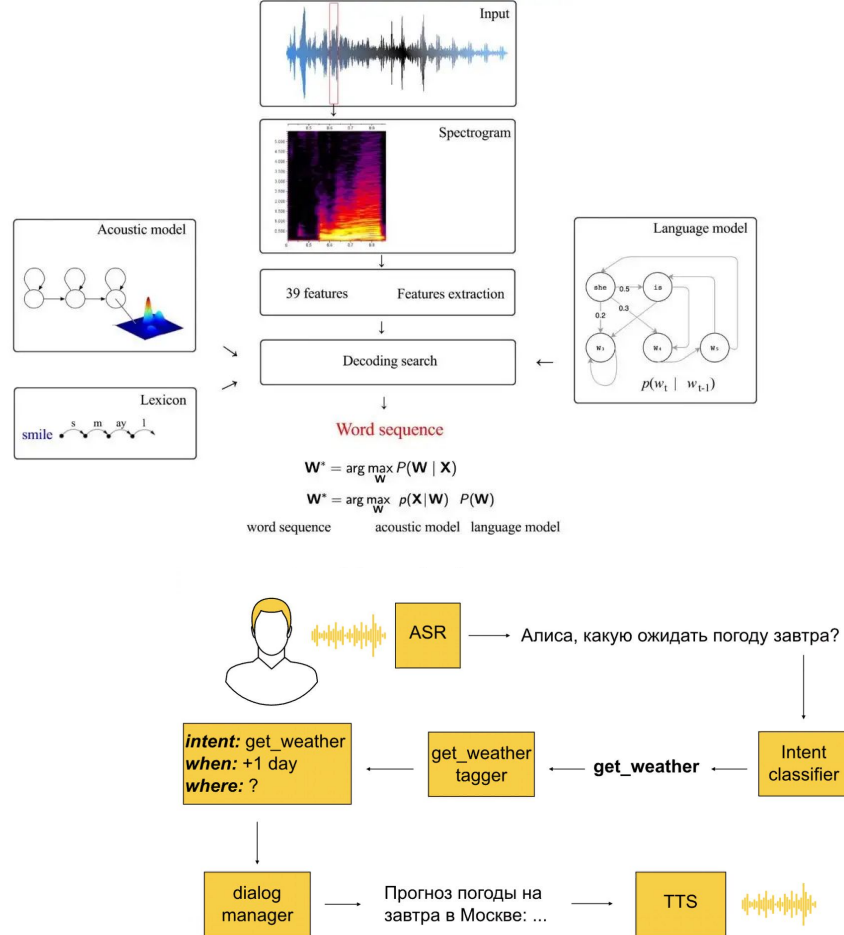
एक चोर की तस्वीर (हिन्दी)

<https://openai.com/blog/dall-e/>

<https://www.assemblyai.com/blog/how-dall-e-2-actually-works/>

# Задачи не про тексты

- Optical Character Recognition (OCR)
- Automatic Speech Recognition (ASR)
- Text-to-Speech (TTS)
- Голосовые ассистенты



<https://habr.com/ru/company/yandex/blog/349372/>

<https://jonathan-hui.medium.com/speech-recognition-asr-model-training-90ed50d93615>

# Aftermath