

# NLP at Scale



**You**

Напиши план презентации для научно-исследовательского семинара по теме "Обзор T5 и GPT-3"



**ChatGPT**

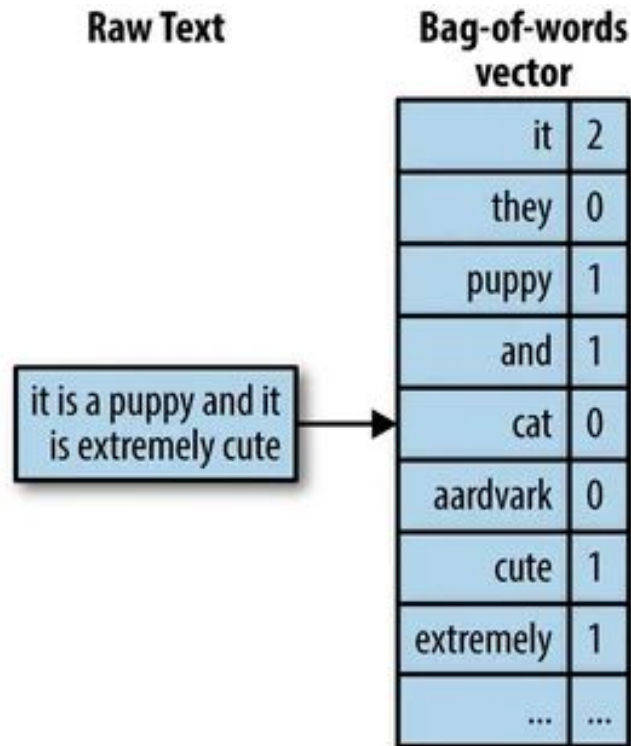
Конечно, вот примерный план презентации:

# Что такое NLP?

- Перевод
- Суммаризация (пересказ)
- Генерация
- Ответы на вопросы
- Что угодно ещё, связанное с текстами

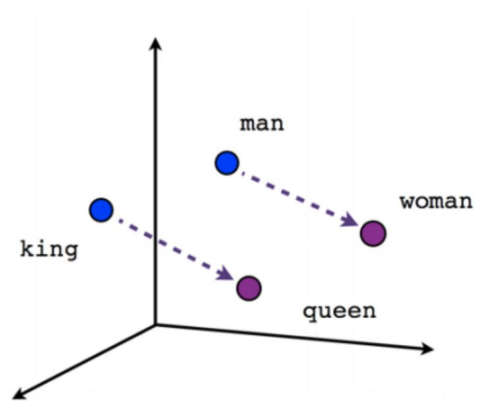
# Немного об истории

- Bag-of-words

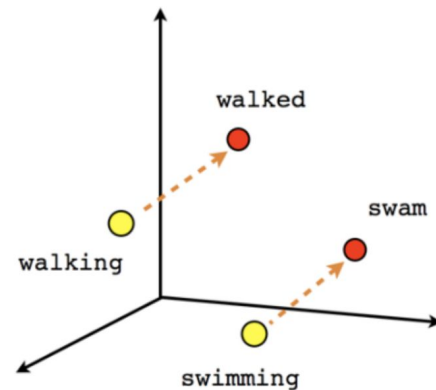


# Немного об истории

- Bag-of-words
- Word2Vec



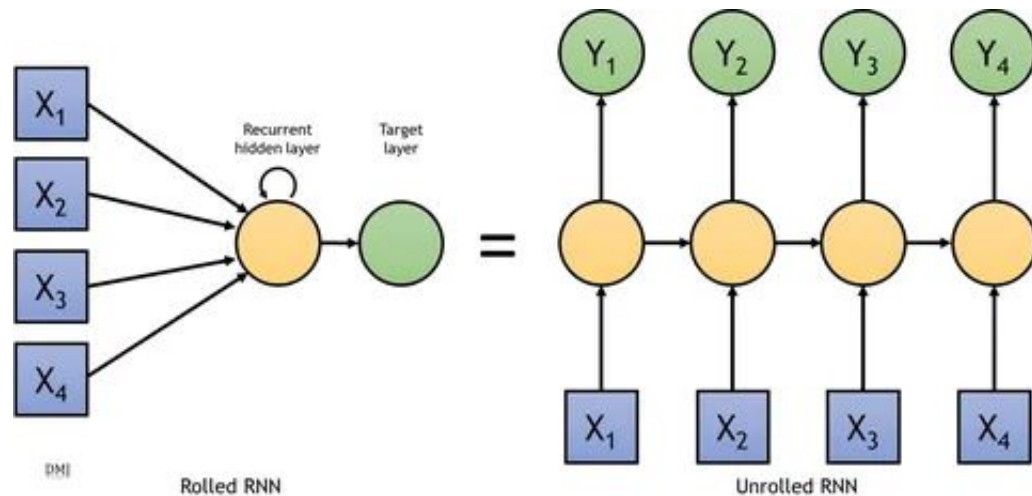
Male-Female



Verb tense

# Немного об истории

- Bag-of-words
- Word2Vec
- RNN



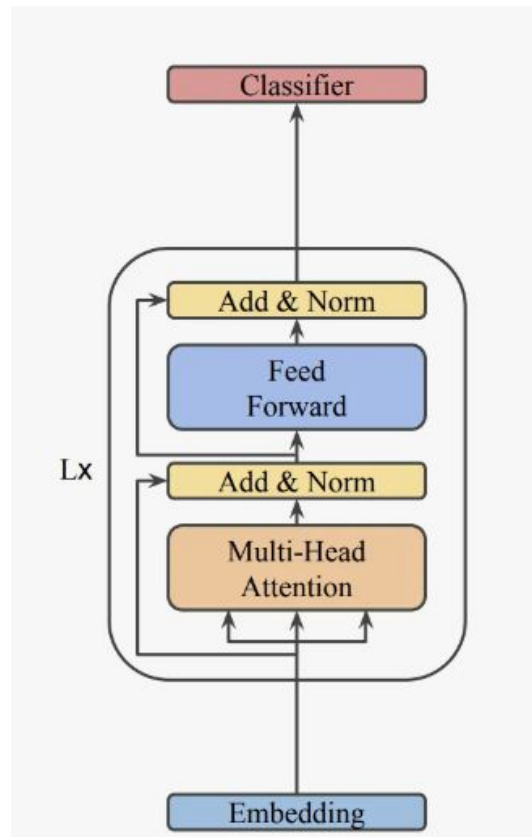
# Немного об истории

- Bag-of-words
- Word2Vec
- RNN
- Transformer!!!

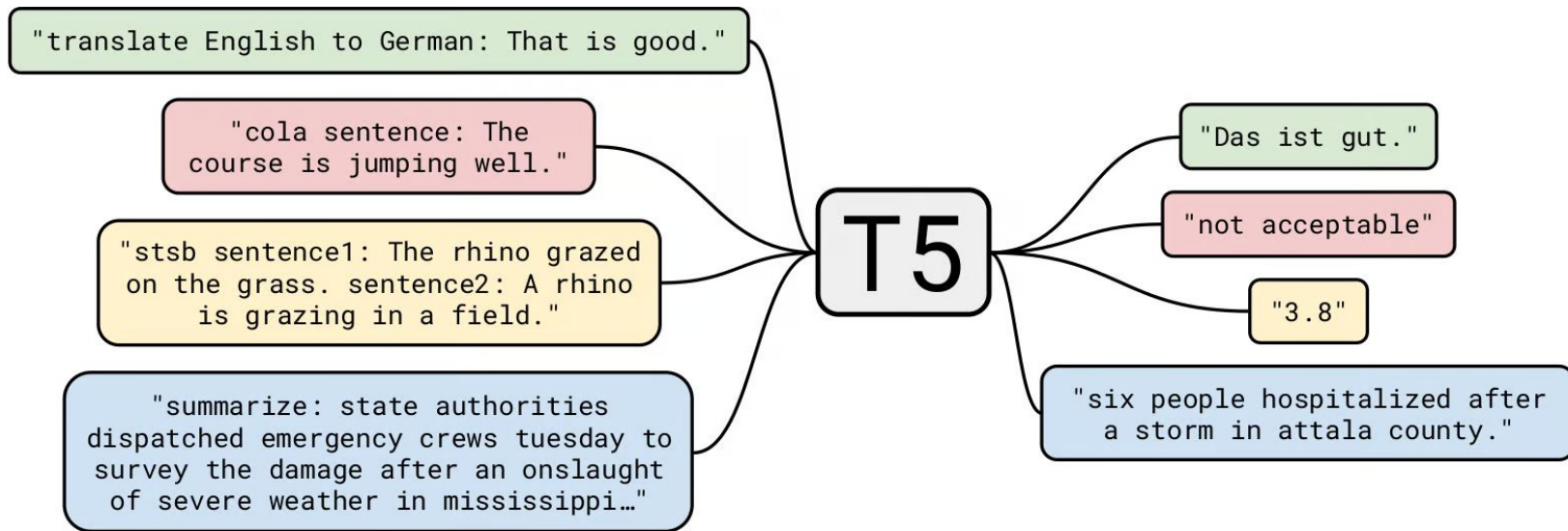


# Немного об истории

- Bag-of-words
- Word2Vec
- RNN
- Transformer!!!
- BERT, RoBERTa, ALBERT



# Универсальный Text-To-Text Transfer Transformer (T5)





Вот и все, мы решили NLP

Спасибо за внимание!



# К сожалению, нет

## Проблемы имеющихся подходов:

- Проблемы метода pre-train + fine-tune:

1. Дорого и долго

2. Низкая устойчивость

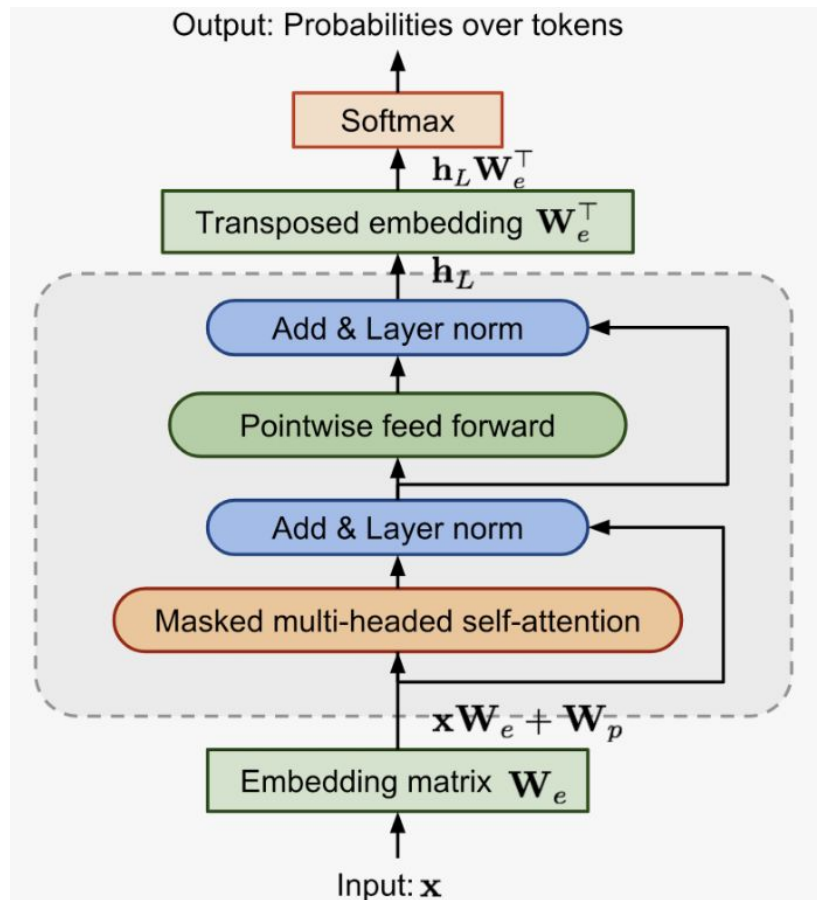
- Проблемы обучения на бенчмарки:

1. Низкая обобщающая способность

2. Отсутствие “понимания” языка

# Решение – GPT-3

- На основе нашего любимого трансформера
- На выходе предсказывает только следующий токен по входному контексту



# Размер имеет значение



T5 (11B параметров)



GPT-3 (175B параметров)

# Размер имеет значение

1. Взять колоссальную модель
2. Скормить ей несколько сотен гигабайт текстов
3. Подождать пару месяцев
4. ???????
5. PROFIT

# Размер имеет значение

- Нет необходимости в fine-tune
- Полностью unsupervised режим обучения
- Хорошее обобщение, за счет большой и качественной выборки



Вот и все, мы решили NLP

Спасибо за внимание!



# К сожалению, нет

- Модель учится полностью unsupervised и намного лучше понимает, как работает язык
- Но качество ответов не очень и к SOTA по бенчмаркам мы даже не приблизились
- Как обойтись без fine-tune?





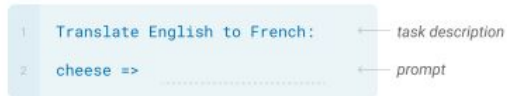
# In-context learning

Из-за огромного размера модели и контекстного окна модель получила способность обучаться прямо из входа во время инференса!

## The three settings we explore for in-context learning

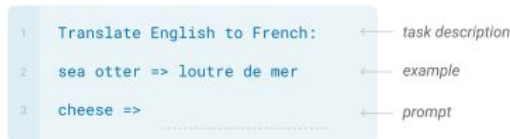
### Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.



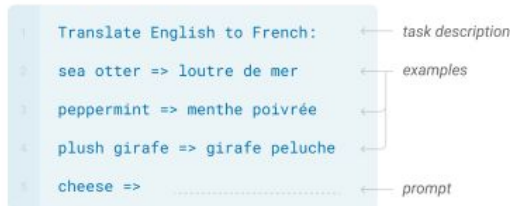
### One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.



### Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.



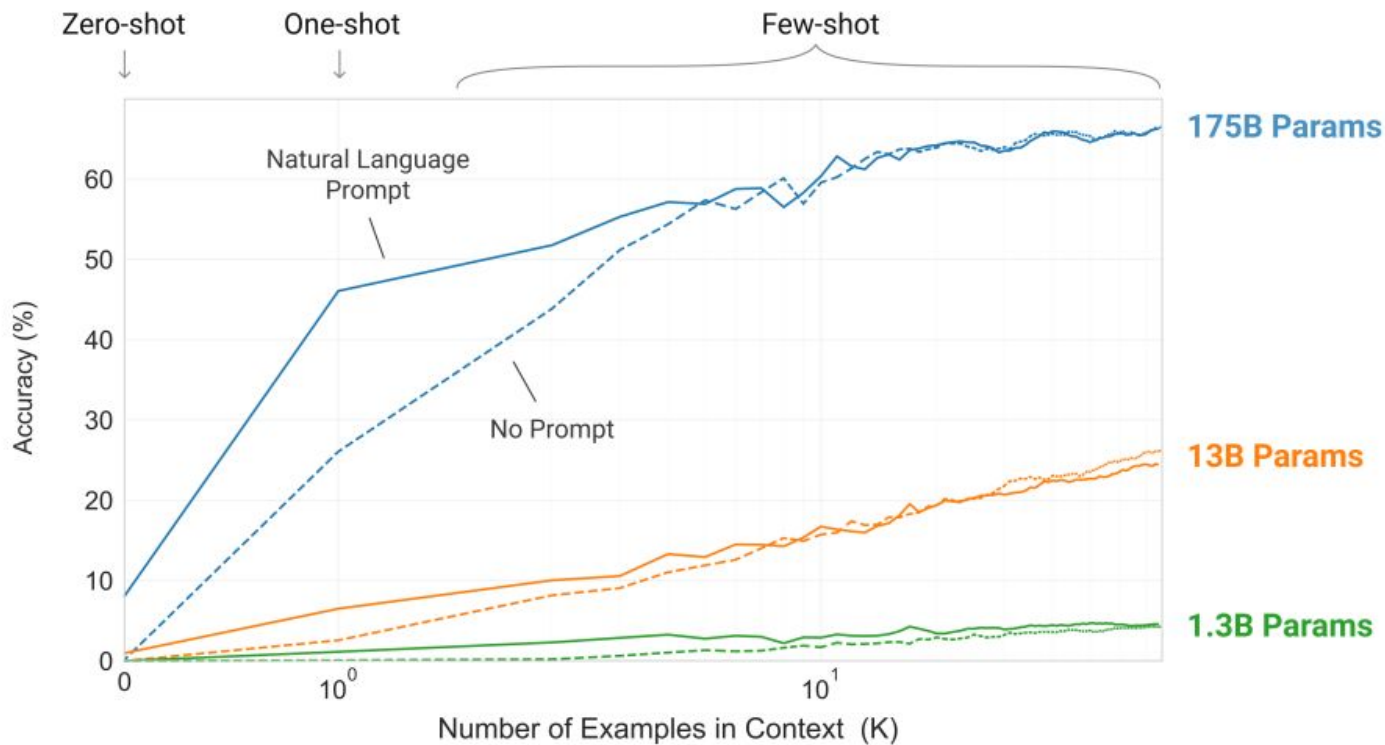
## Traditional fine-tuning (not used for GPT-3)

### Fine-tuning

The model is trained via repeated gradient updates using a large corpus of example tasks.



# In-context learning

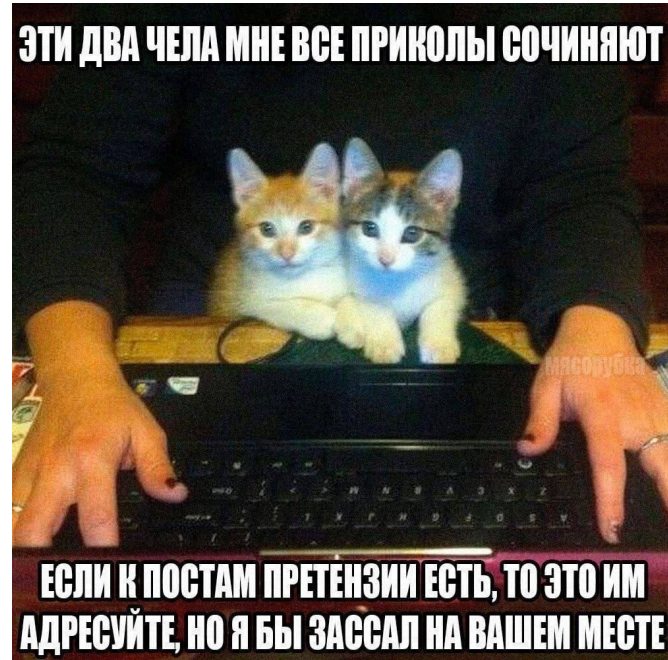
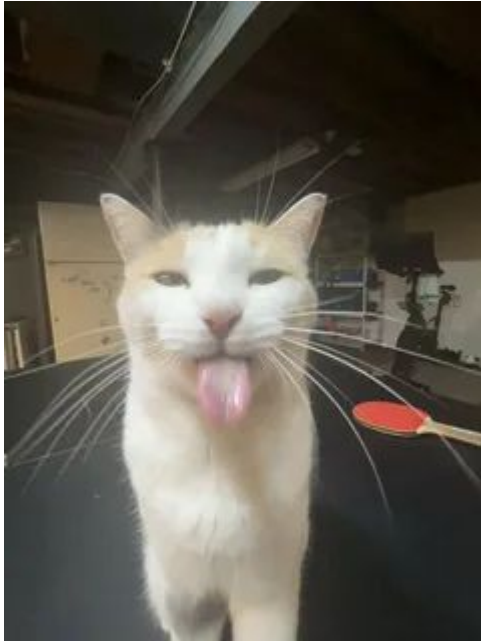


# Ограничения GPT-3

- Огромный размер
- Трудности с генерацией текстов (особенно больших)
- Только одно направление генерации и обучения
- Плохое “понимание” мира в целом
- Смещенность модели в сторону обучающих данных

Вот теперь все, мы так и не решили NLP

Спасибо за внимание!



# ИСТОЧНИКИ

- [Language Models are Few-Shot Learners](#)
- [Language models are unsupervised multitask learners](#)
- [Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer](#)