

# Transfer Learning в NLP

## Лекция 1

Введение в Transfer Learning. Self-supervised learning

# План занятия

1. Что такое Self-supervised Learning и Transfer Learning
2. Бенчмарк для оценки качества GLUE
3. GPT и языковое моделирование как self-supervised задача

# Self-supervised learning или как использовать много данных



В мире существует огромное количество “сырого” текста. Из него состоят книги, статьи, посты в соцсетях, веб-страницы и многое другое.

# Self-supervised learning или как использовать много данных



В мире существует огромное количество “сырого” текста. Из него состоят книги, статьи, посты в соцсетях, веб-страницы и многое другое.



Можем ли мы использовать эти данные, чтобы улучшить качество модели?

# Self-supervised learning или как использовать много данных



В мире существует огромное количество “сырого” текста. Из него состоят книги, статьи, посты в соцсетях, веб-страницы и многое другое.



Можем ли мы использовать эти данные, чтобы улучшить качество модели?



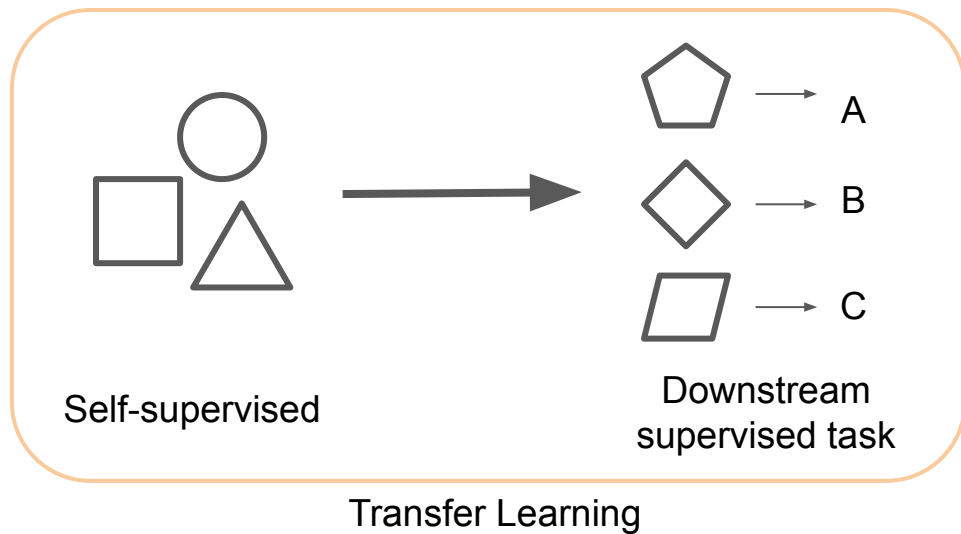
Конечно, да!

# Что же такое self-supervised learning

Self-supervised learning — процесс, при котором модель машинного обучения учит восстанавливать структуру данных на большом неразмеченном датасете для получения хороших промежуточных представлений.

# Что же такое self-supervised learning

Self-supervised learning — процесс, при котором модель машинного обучения учит восстанавливать структуру данных на большом неразмеченном датасете для получения хороших промежуточных представлений.



# General Language Understanding Evaluation (GLUE)

GLUE — англоязычный датасет, состоящий из 9 задач для оценки NLP-моделей на понимание общих концепций естественного языка.












## Какие задачи и датасеты есть в GLUE

Название	Задача	Train size	Описание
<b>CoLA</b>	Linguistic Acceptability	8.5k	Грамматическая корректность предложений
<b>SST-2</b>	Sentiment analysis	67k	Анализ эмоциональной окраски на отзывах
<b>MRPC</b>	Semantic similarity	3.7k	Являются ли пары предложений семантически похожими
<b>QQP</b>	Question semantic similarity	364k	Являются ли пары вопросов семантически похожими
<b>STS-B</b>	Semantic similarity	7k	Являются ли пары предложений семантически похожими
<b>MNLI</b>	NLI	393k	Классификация пар предложений на NLI
<b>QNLI</b>	QA/NLI	105k	Классификация пар вопрос+текст на содержание ответа
<b>WNLI</b>	Textual entailment	634	Является второе предложение следованием первого
<b>RTE</b>	Textual entailment	2.5k	Является второе предложение следованием первого

 — single-sentence задачи     — pair-sentence задачи

# GLUE — ещё и публичный лидерборд

GLUE SuperGLUE															
Paper </> Code Tasks Leaderboard i FAQ Diagnostics Submit Profile Logout															
Rank	Name	Model	URL	Score	CoLA	SST-2	MRPC	STS-B	QQP	MNLI-m	MNLI-mm	QNLI	RTE	WNLI	AX
1	Microsoft Alexander v-team	Turing ULR v6		91.3	73.3	97.5	94.2/92.3	93.5/93.1	76.4/90.9	92.5	92.1	96.7	93.6	97.9	55.4
2	JDExplore d-team	Vega v1		91.3	73.8	97.9	94.5/92.6	93.5/93.1	76.7/91.1	92.1	91.9	96.7	92.4	97.9	51.4
3	Microsoft Alexander v-team	Turing NLR v5		91.2	72.6	97.6	93.8/91.7	93.7/93.3	76.4/91.1	92.6	92.4	97.9	94.1	95.9	57.0
4	DIRL Team	DeBERTa + CLEVER		91.1	74.7	97.6	93.3/91.1	93.4/93.1	76.5/91.0	92.1	91.8	96.7	93.2	96.6	53.3
5	ERNIE Team - Baidu	ERNIE		91.1	75.5	97.8	93.9/91.8	93.0/92.6	75.2/90.9	92.3	91.7	97.3	92.6	95.9	51.7
6	AliceMind & DIRL	StructBERT + CLEVER		91.0	75.3	97.7	93.9/91.9	93.5/93.1	75.6/90.8	91.7	91.5	97.4	92.5	95.2	49.1
7	DeBERTa Team - Microsoft	DeBERTa / TuringNLRv4		90.8	71.5	97.5	94.0/92.0	92.9/92.6	76.2/90.8	91.9	91.6	99.2	93.2	94.5	53.2
8	HFL iFLYTEK	MacALBERT + DKM		90.7	74.8	97.0	94.5/92.6	92.8/92.6	74.7/90.6	91.3	91.1	97.8	92.0	94.5	52.6
9	PING-AN Omni-Sinitic	ALBERT + DAAF + NAS		90.6	73.5	97.2	94.0/92.0	93.0/92.4	76.1/91.0	91.6	91.3	97.5	91.7	94.5	51.2
10	T5 Team - Google	T5		90.3	71.6	97.5	92.8/90.4	93.1/92.8	75.1/90.6	92.2	91.9	96.9	92.8	94.5	53.1
11	Microsoft D365 AI & MSR AI & GATECH	MT-DNN-SMART		89.9	69.5	97.5	93.7/91.6	92.9/92.5	73.9/90.2	91.0	90.8	99.2	89.7	94.5	50.2

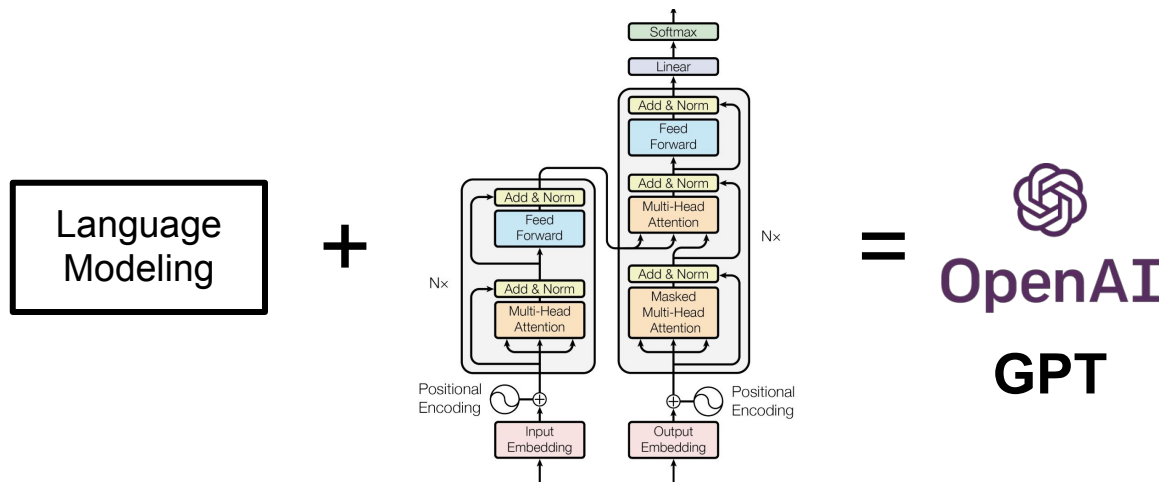
<https://gluebenchmark.com/>

# Каким образом сформулировать self-supervised задачу для понимания естественного языка?

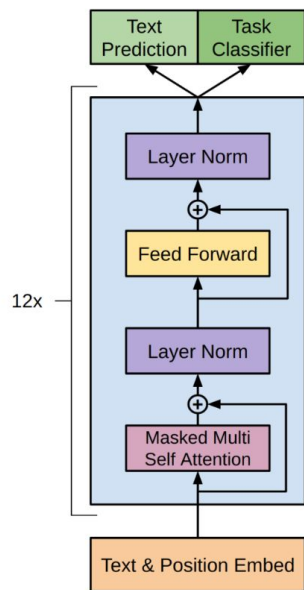
Языковое моделирование — self-supervised задача, которая позволяет изучить структуру языка.

# Каким образом сформулировать self-supervised задачу для понимания естественного языка?

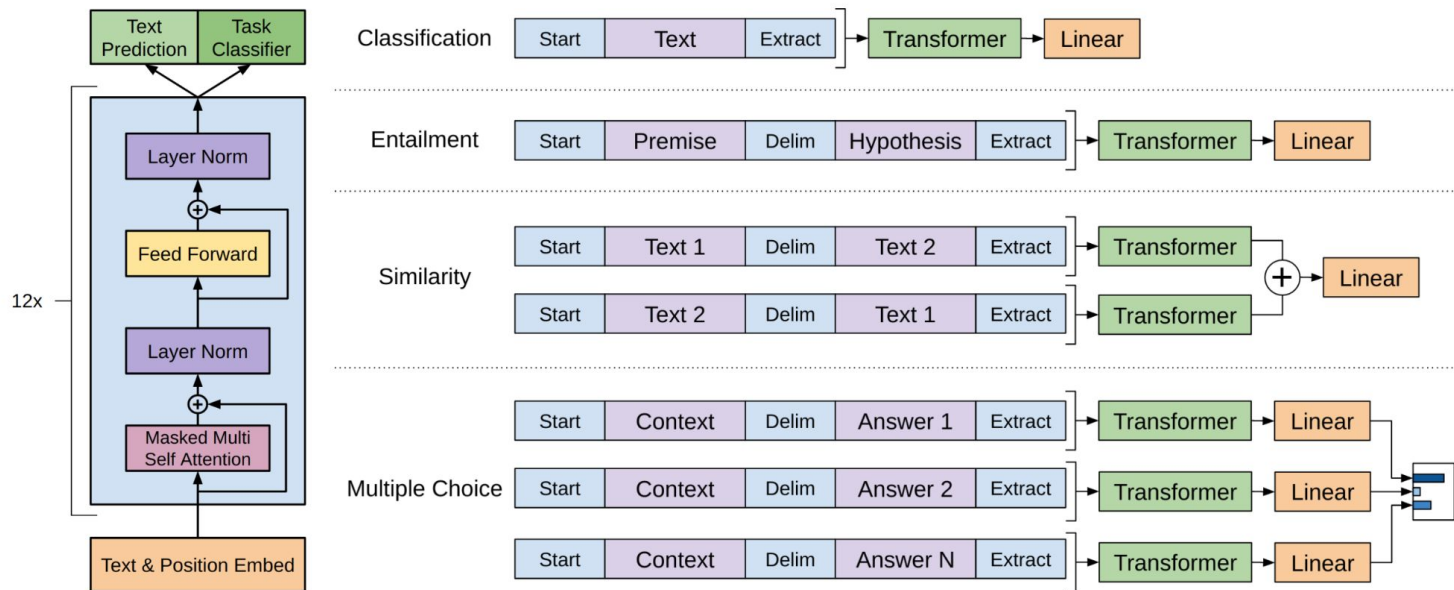
Языковое моделирование — self-supervised задача, которая позволяет изучить структуру языка.



# Как это работает?



# Как это работает?



# Детали

- Transformer “decoder” с 12 слоями и 117M параметров
- Размер скрытого состояния 768
- BPE словарь размера 40 тыс.
- Обучена на датасете BookCorpus

# Как выросли метрики?

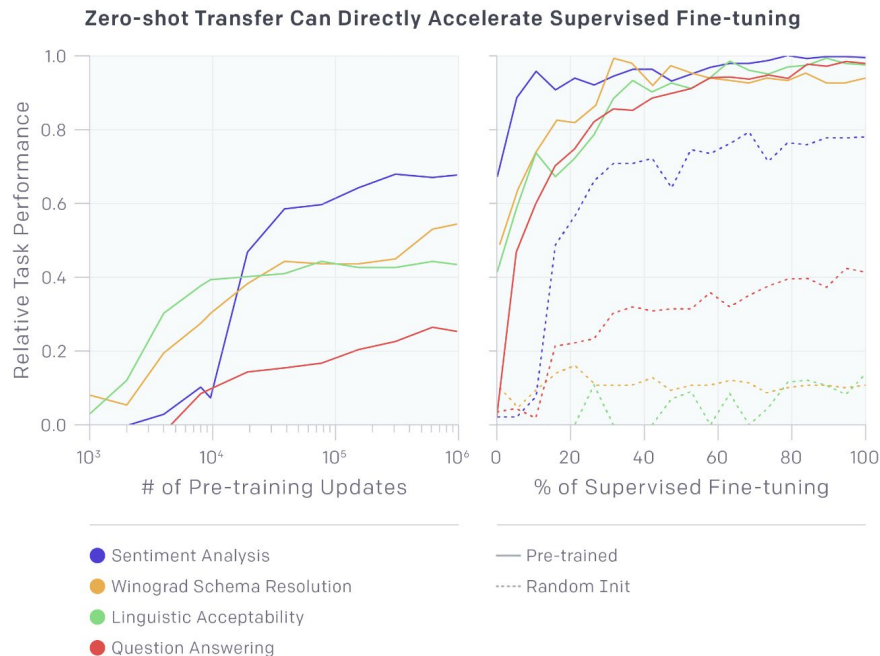
На момент выхода в 2018 GPT стала новой state-of-the-art моделью.

DATASET	TASK	SOTA	OURS
SNLI	Textual Entailment	89.3	<b>89.9</b>
MNLI Matched	Textual Entailment	80.6	<b>82.1</b>
MNLI Mismatched	Textual Entailment	80.1	<b>81.4</b>
SciTail	Textual Entailment	83.3	<b>88.3</b>
QNLI	Textual Entailment	82.3	<b>88.1</b>
RTE	Textual Entailment	<b>61.7</b>	56.0
STS-B	Semantic Similarity	81.0	<b>82.0</b>
QQP	Semantic Similarity	66.1	<b>70.3</b>
MRPC	Semantic Similarity	<b>86.0</b>	82.3
RACE	Reading Comprehension	53.3	<b>59.0</b>
ROCStories	Commonsense Reasoning	77.6	<b>86.5</b>
COPA	Commonsense Reasoning	71.2	<b>78.6</b>
SST-2	Sentiment Analysis	<b>93.2</b>	91.3
CoLA	Linguistic Acceptability	35.0	<b>45.4</b>
GLUE	Multi Task Benchmark	68.9	<b>72.8</b>

<https://openai.com/blog/language-unsupervised/>

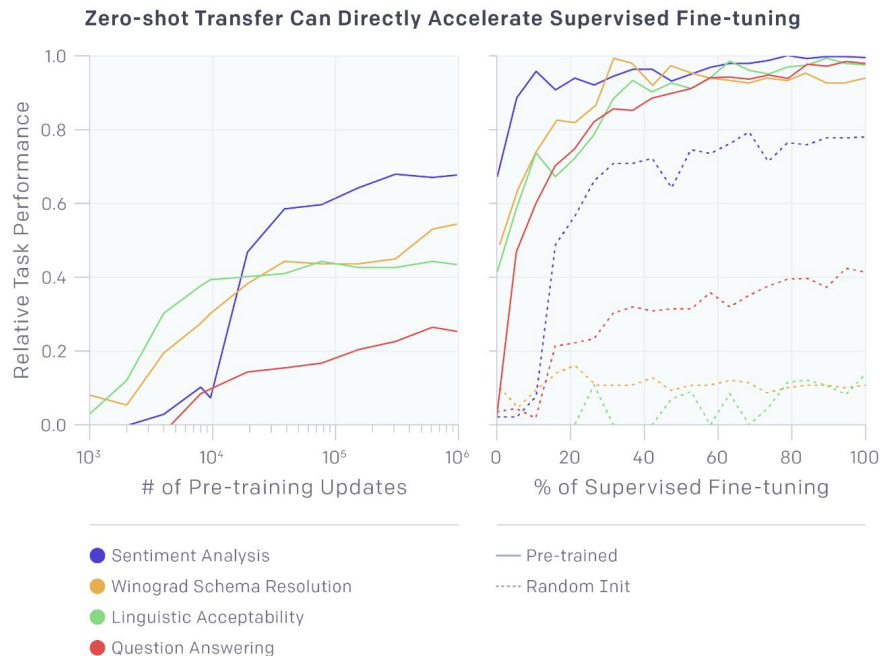


# И этап претренировки действительно важен!



Левый график: с увеличением количества pretraining шагов качество модели в zero-shot решении задач улучшается

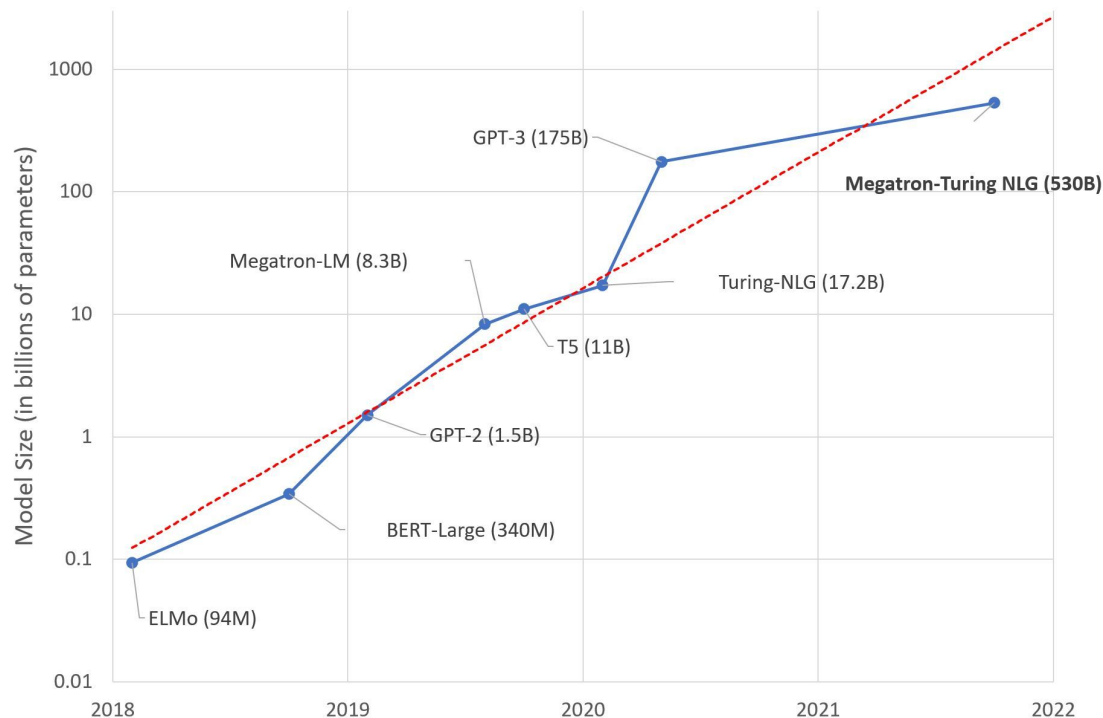
# И этап претренировки действительно важен!



Левый график: с увеличением количества pretraining шагов качество модели в zero-shot решении задач улучшается

Правый график: претренированная модель гораздо лучше справляется с задачами по сравнению со случайной инициализацией

# GPT-{1/2/3} просто добавь словёв!



# Итоги занятия

1. Познакомились с новым подходом в машинном обучении — self-supervised learning, который не требует дополнительной разметки данных, а использует структуру данных для обучения моделей.

# Итоги занятия

1. Познакомились с новым подходом в машинном обучении — self-supervised learning, который не требует дополнительной разметки данных, а использует структуру данных для обучения моделей.
2. Узнали о GLUE — коллекции датасетов, помогающих оценивать качество моделей.

# Итоги занятия

1. Познакомились с новым подходом в машинном обучении — self-supervised learning, который не требует дополнительной разметки данных, а использует структуру данных для обучения моделей.
2. Узнали о GLUE — коллекции датасетов, помогающих оценивать качество моделей.
3. Вспомнили про задачу языкового моделирования, которая на самом деле является self-supervised, и узнали о модели GPT, которая использует архитектуру «Трансформер» и позволяет улучшить качество на GLUE.