State of the Art

word2vec

https://arxiv.org/abs/1301.3781

Google, 2013

GloVe

https://nlp.stanford.edu/projects/glove/

Stanford, 2014

CoVe

https://arxiv.org/abs/1708.00107v2

Salesforce.com, 2016

CoVe approach

- BiLSTM encoder, LSTM decoder

- trained for machine translation

- use context from encoder as context vectors

ELMo

"Embeddings from Language Models"

https://allennlp.org/elmo

https://arxiv.org/abs/1802.05365

Allen Institute, Feb 2018

94 million parameters

ElMo approach

- BiLSTM

- trained for language modelling
    - both forward and reverse, concatenated

- *contextual* embeddings – functions of the entire input sequence

GPT

"Generative Pre-Training"

https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

OpenAI, June 2018

110 million parameters

GPT approach

- Transformer
- trained for language modelling
  - unidirectional

BERT

"Bidirectional Encoder Representations from Transformers"

https://arxiv.org/abs/1810.04805

Google, Oct 2018

340 million parameters

BERT approach

- Transformer

- trained for language modelling
    - bidirectional

- fine-tuning

GPT-2

https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

OpenAI, Feb 2019

1.5 billion parameters

40GB of training data

GTP-2 and fake news

https://openai.com/blog/better-language-models/

https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction

RoBERTa

https://ai.facebook.com/blog/roberta-an-optimized-method-for-pretraining-self-supervised-nlp-systems/

https://arxiv.org/abs/1907.11692

Facebook, July 2019

355 million parameters

XLNet

https://proceedings.neurips.cc/paper/2019/hash/dc6a7e655d7e5840e66733e9ee67cc69-Abstract.html

CMU/Google Dec 2019

Turing-NLG

https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/

Microsoft, Feb 2020

17 billion parameters

GPT-3

https://arxiv.org/abs/2005.14165

OpenAI, May 2020

175 billion parameters

trained on ~500 billion tokens (570 GB)

PaLM

https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html

Google, April 2022

540 billion parameters

AI Dungeon

https://play.aidungeon.io/main/landing