NLP MEETUP VIENNA

NATURAL LANGUAGE PROCESSING

NLP Meetup #4

18.2.2020 @ Sclable

NLP Meetup #4

sclable

19:00 - NLP News & Trends - **Liad Magen**

19:30 - Emotion Recognition in Textual Conversations - **Philipp Möhl**

20:00 - Networking & Refreshments

Stefan Gindl

Dr. Andreas Rath

Jason Hoelscher-Obermaier

Liad Magen

NLP - News & Trends

Liad Magen

News Topics

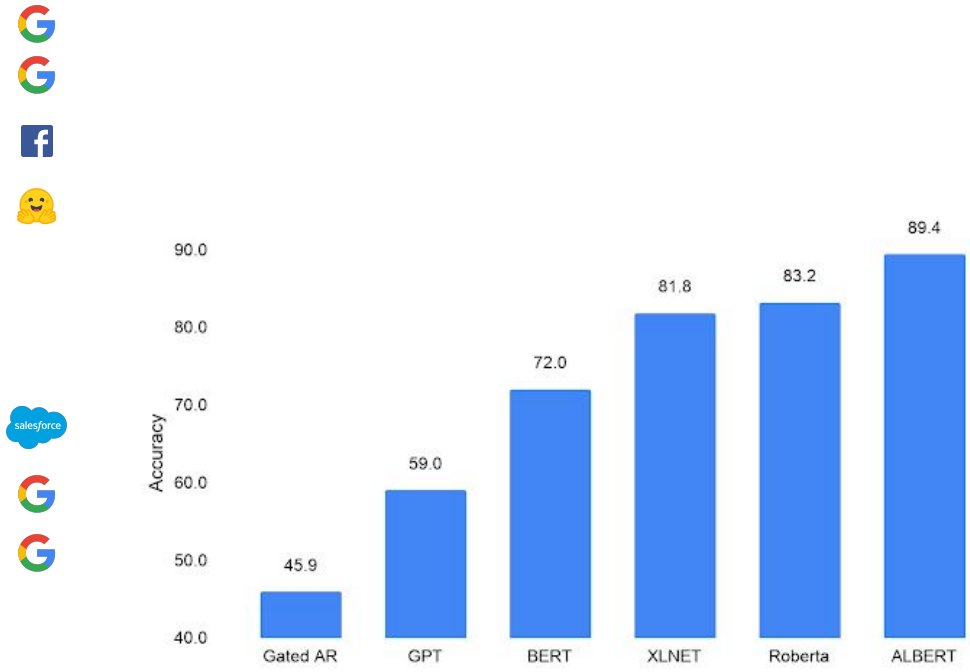
- Language Models
- Chatbots
- Information Retrieval/Extraction
- Applied AI Frameworks

Language Models

BERTology

BERT’s children

- AL**BERT**
- XLNet
- Ro**BERT**a
- DISTIL**BERT**
- CamemBERT
  FlauBERT
- CTRL
- T5
- Reformer



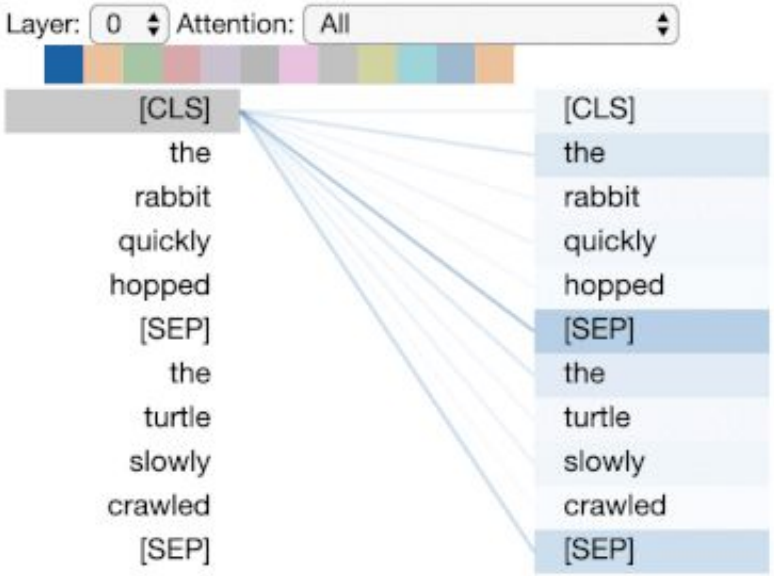
* Check out TRAX

BERT Analyzing tools

jessevig/bertviz: Tool for visualizing attention in the Transformer model (BERT, GPT-2, Albert, XLNet, RoBERTa, CTRL, etc.)

[1906.04341] What Does BERT Look At? An Analysis of BERT's Attention

Partial explainability tool.



BERT Research: Can BERT capture Linguistic structure?

[1905.05950] BERT Rediscovers the Classical NLP Pipeline

POS ➔parsing ➔NER ➔semantic roles ➔coref.

[1906.02715] Visualizing and Measuring the Geometry of BERT

"We find evidence of a fine-grained geometric representation of word senses"

Revealing the Dark Secrets of BERT

BERT is severely overparameterized;

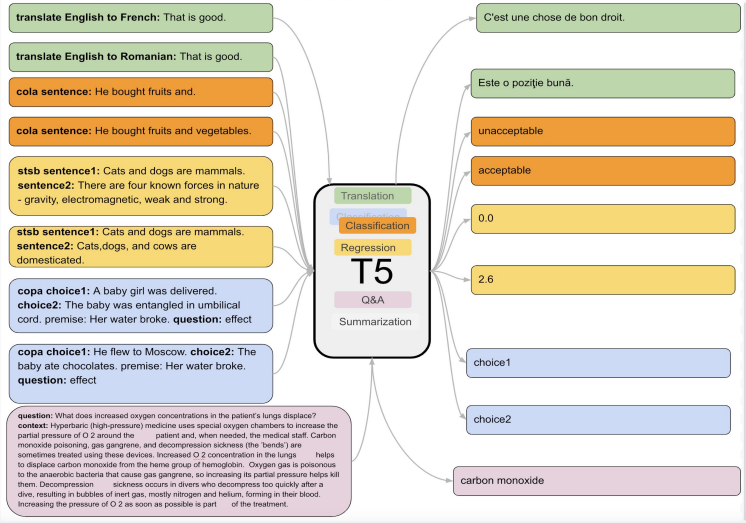
Not all attention heads are required

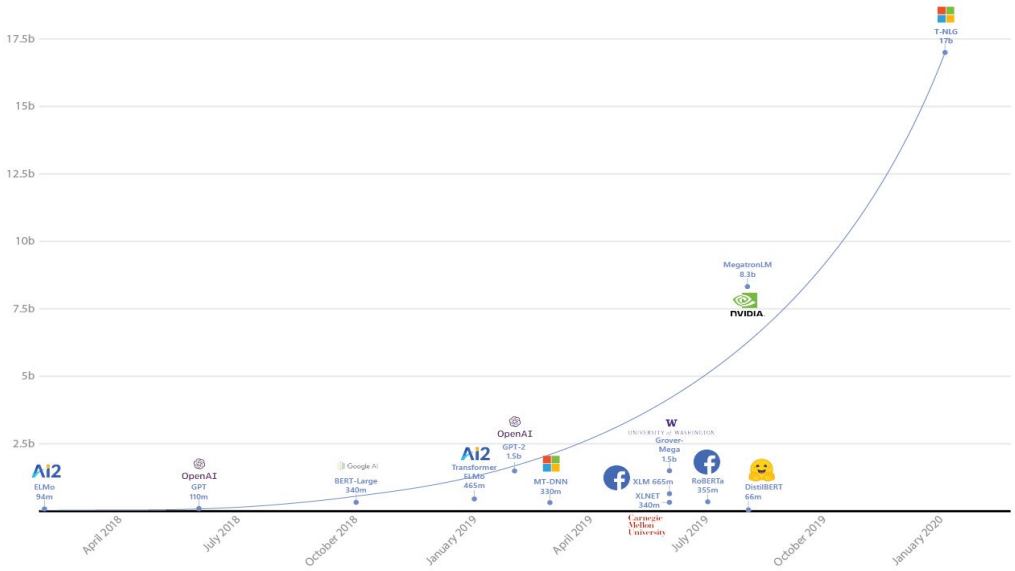
BERTrend

1. BERT leads a major field breakthrough
    a. **Transfer learning**
    b. Task independency
    c. (officially powering google search)
2. Active development
    a. Academic research
    b. Industry collaboration
3. Language dependent
    a. Performs better on single language models
    b. Independant releasing of LMs: French, Chinese, Finish...
4. Size doesn't always matters
    a. DistillBERT / ALBERT

"My LM is bigger than yours"

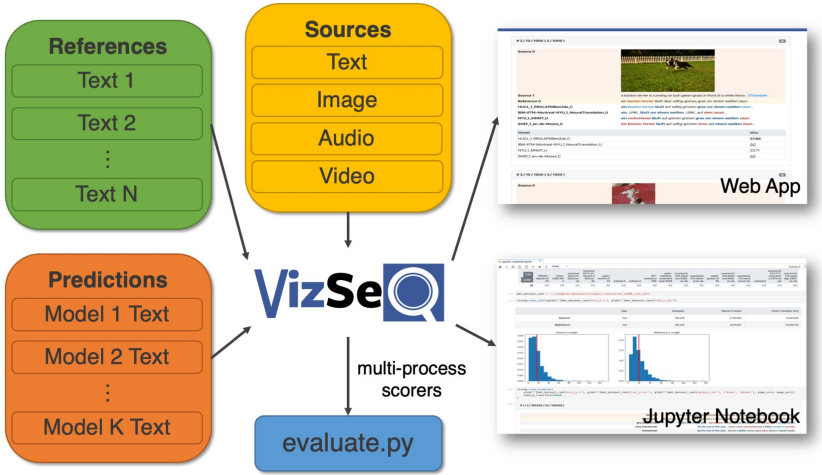
- “Text-to-Text Transfer Transformer” - T5 (11B)
- T-NLG (17B)

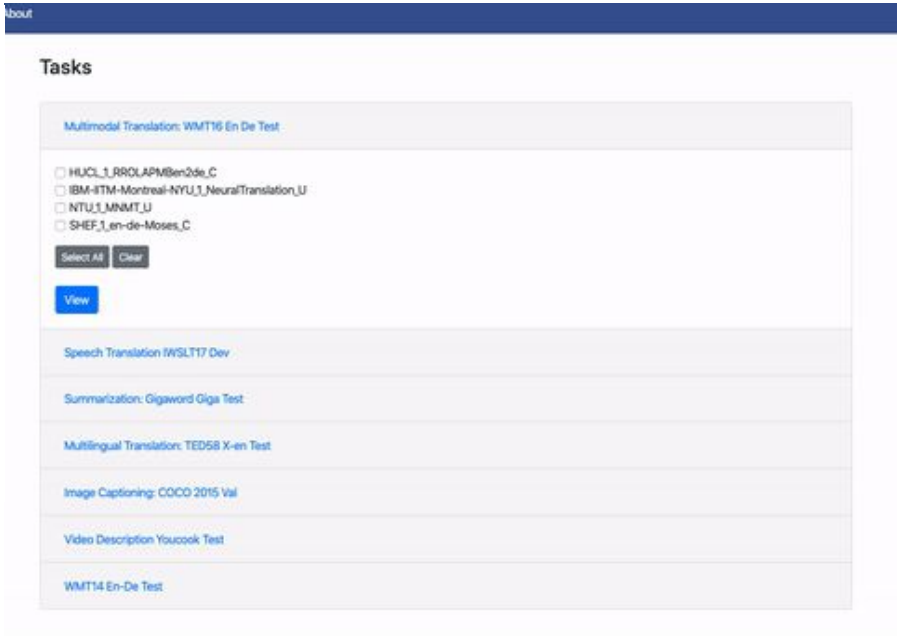




Measuring LM for text generation

facebookresearch/vizseq: A Research Toolkit for Natural Language Generation (Translation, Captioning, Summarization, etc.)

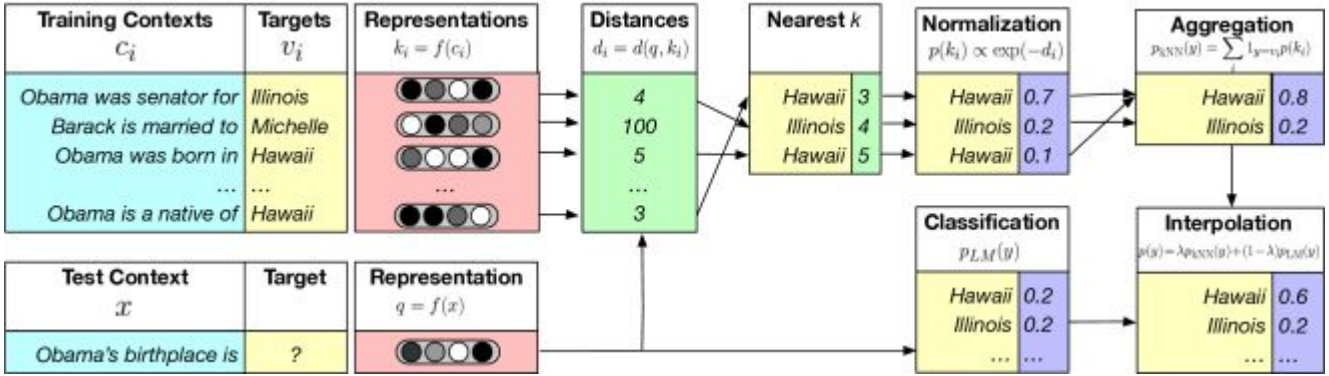




New LM Approaches

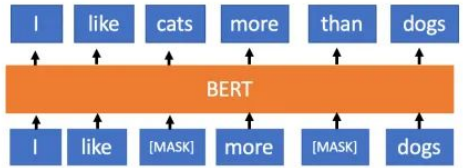
[1906.08237] **XLNet**: Generalized Autoregressive Pretraining for Language Understanding

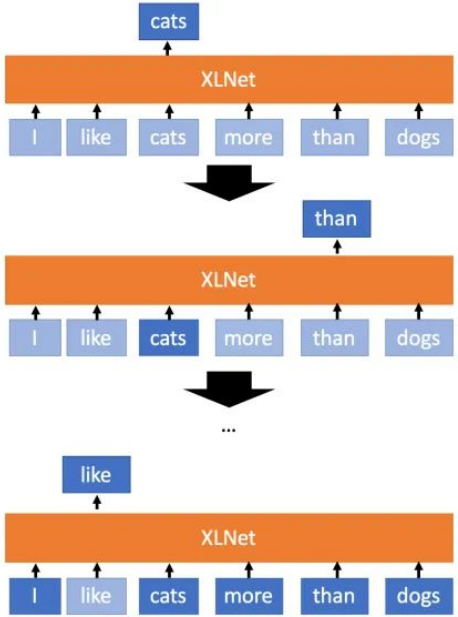
[1911.00172] Generalization through Memorization: **Nearest Neighbor Language Models**



XLNet vs BERT

XLNet randomly selects the next word to predicts:



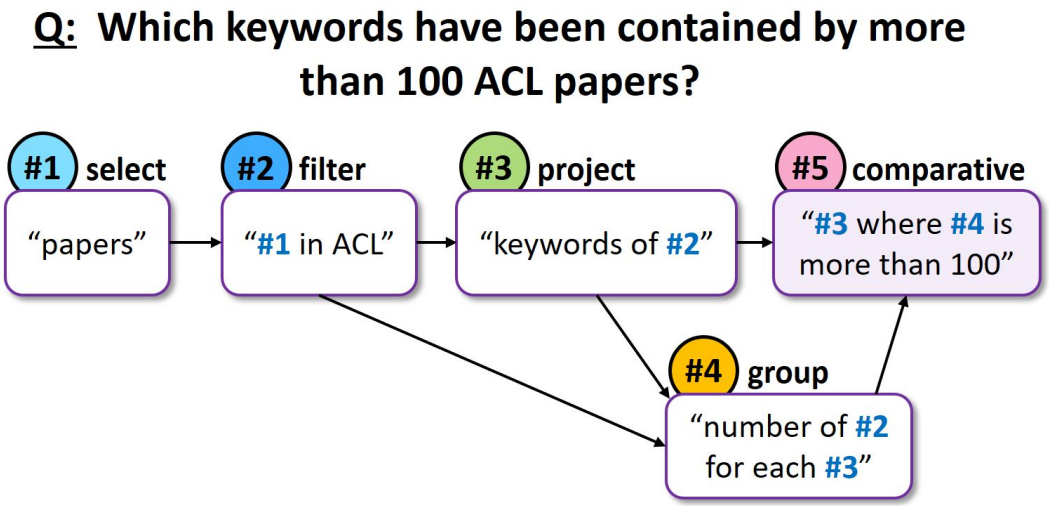


Information Extraction

BREAK Dataset

A Question Understanding Benchmark | BREAK

83K questions for training models to reason over complex questions

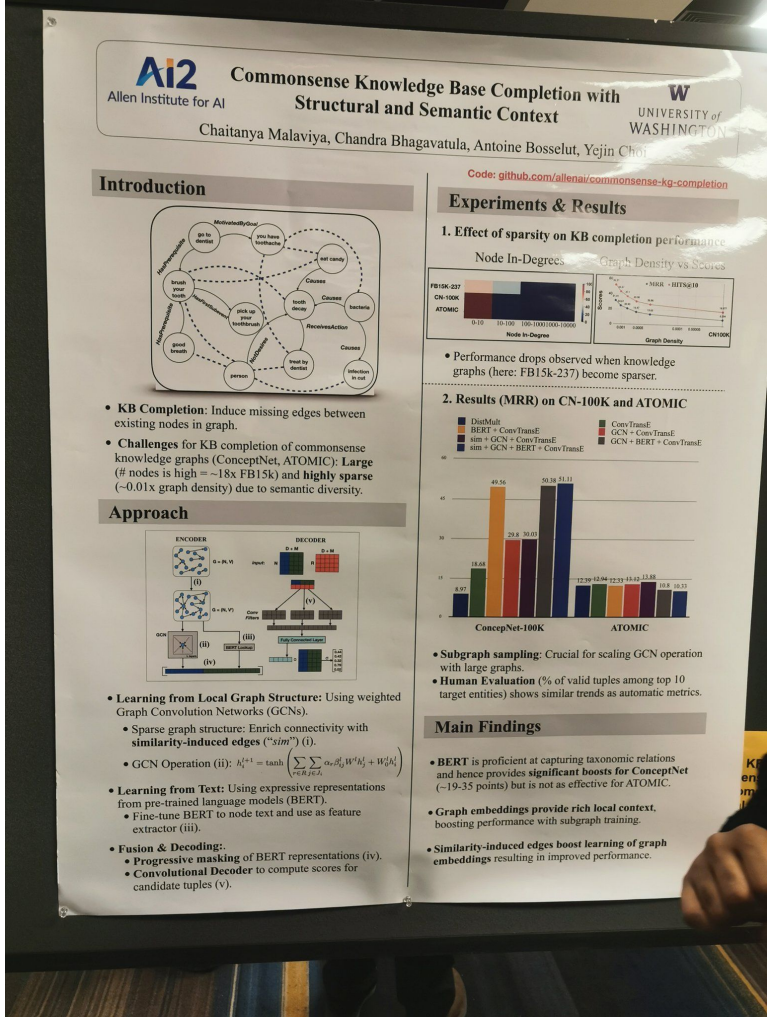


AllenNLP

Graph knowledge base

[1910.02915] Commonsense Knowledge Base Completion with Structural and Semantic Context

REALM: Retrieval-Augmented Language Model Pre-Training



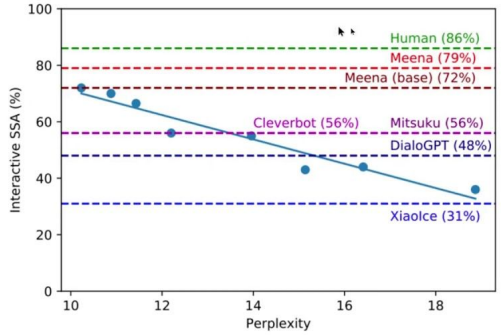
Chatbots

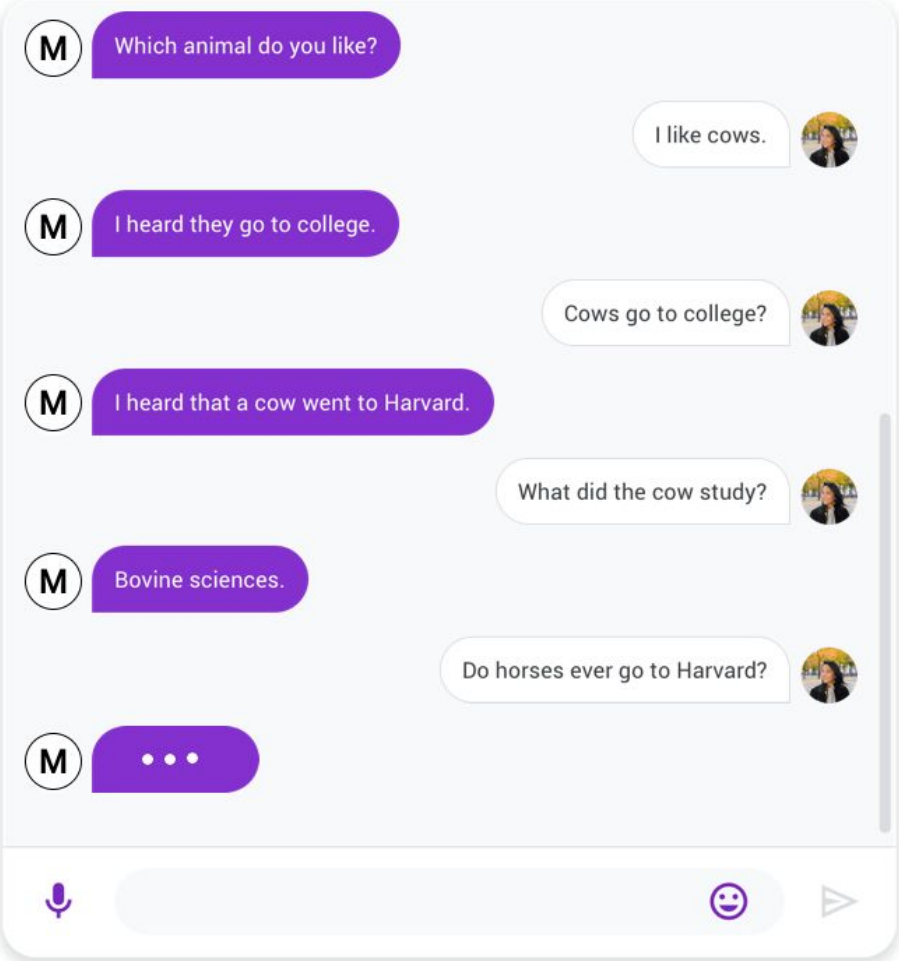
Meet Meena

[2001.09977] Towards a Human-like Open-Domain Chatbot

- 2.6B params - GPT-2 x 2
- 40B words from social media conversations
- SSA ⇒ Sensibleness and Specificity Average
  Does the response make sense?
  Is it specific to the content?

SSA vs Perplexity





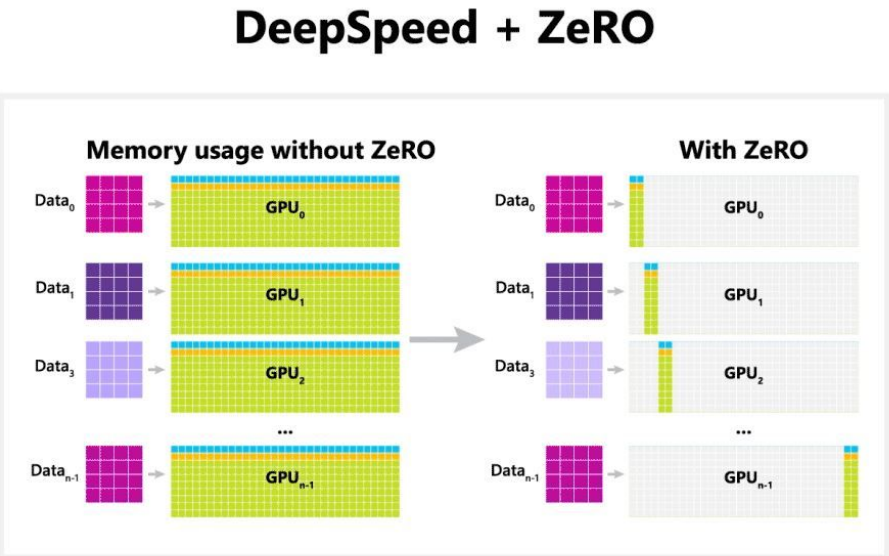
Applied AI

Deep Speed

DeepSpeed

x4 - x8 memory reduction

32GB GPUs



Scale
- 100B parameter
- 10X bigger

Speed
- Up to 5X faster

Cost
- Up to 5X cheaper

Usability
- Minimal code change

THiNC

- Framework agnostic (PyTorch, TF...)
- Configuration system
- High order functions
- Typed variables

Thank you!

Bibliography

https://www.scihive.org/paper/1911.00172

https://thegradient.pub/gpt2-and-the-nature-of-intelligence/

https://allenai.github.io/Break/

https://medium.com/dair-ai/nlp-newsletter-reformer-deepmath-electra-tinybert-for-search-vizseq-open-sourcing-ml-68d5b6eed057https://ai.googleblog.com/2020/01/towards-conversational-agent-that-can.html