– IN5550 –
Neural Methods in Natural Language Processing

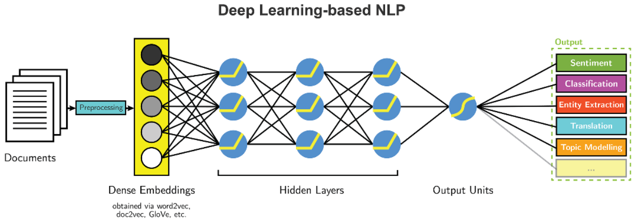
Sustainability 1: The possibilities and problems of large LMs

Lilja Øvrelid

(With thanks to Jeremy Barnes)

Neural NLP

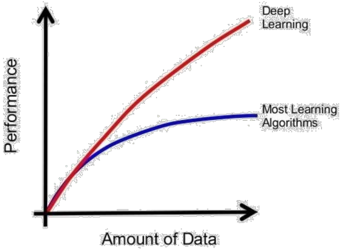
- Most NLP tasks approached as classification problems, using supervised machine learning
- Since at least 2014, neural networks completely dominating.
- Very successful on NLP benchmarks



https://ebookreading.net/view/book/EB9781787121423397.html

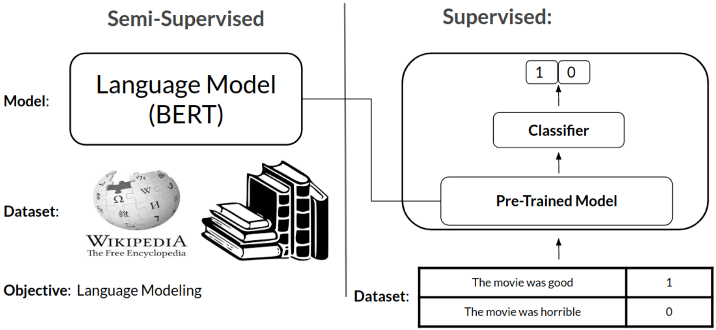
Why now?

- large amounts of data
- substantial computing power (GPUs, parallelization)
- architectures that combine the two
- transfer learning of pre-trained language models



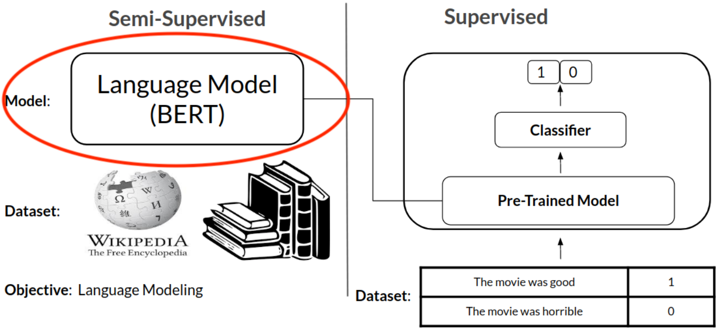
www.icaroai.com

NLP system



(https://blog.insightdatascience.com)

NLP system



(https://blog.insightdatascience.com)

NLP benchmarking and LMs

(March 2021)

SuperGLUE GLUE Paper </> Code Tasks

Leaderboard Version: **2.0**

	Rank	Name	Model	URL	Score	BoolQ	CB	COPA
	1	SuperGLUE Human Baselines	SuperGLUE Human Baselines		89.8	89.0	95.8/98.9	100.0
+	2	T5 Team - Google	T5		89.3	91.2	93.9/96.8	94.8
	3	Zhuiyi Technology	RoBERTa-mtl-adv		85.7	87.1	92.4/95.6	91.2
	4	Facebook AI	RoBERTa		84.6	87.1	90.5/95.2	90.6
	5	IBM Research AI	BERT-mtl		73.5	84.8	89.6/94.0	73.8
	6	SuperGLUE Baselines	BERT++		71.5	79.0	84.8/90.4	73.8
			BERT		69.0	77.4	75.7/83.6	70.6
			Most Frequent Class		47.1	62.3	21.7/48.4	50.0
			CBoW		44.5	62.2	49.0/71.2	51.6
			Outside Best		-	80.4	-	84.4
	-	Stanford Hazy Research	Snorkel [SuperGLUE v1.9]		-	-	88.6/93.2	76.2

NLP benchmarking and LMs

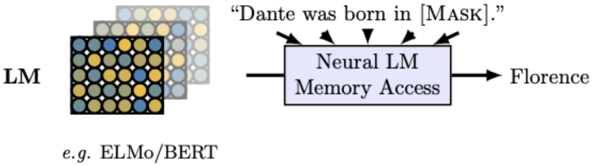
(March 2022)

SuperGLUE GLUE Paper </> Code Tasks

Leaderboard Version: **2.0**

	Rank	Name	Model	URL	Score	BoolQ	CB	COPA	M
+	1	Liam Fedus	ST-MoE-32B		91.2	92.4	96.9/98.0	99.2	89
	2	Microsoft Alexander v-team	Turing NLR v5		90.9	92.0	95.9/97.6	98.2	88
	3	ERNIE Team - Baidu	ERNIE 3.0		90.6	91.0	98.6/99.2	97.4	88
+	4	Zirui Wang	T5 + UDG, Single Model (Google Brain)		90.4	91.4	95.8/97.6	98.0	88
+	5	DeBERTa Team - Microsoft	DeBERTa / TuringNLRv4		90.3	90.4	95.7/97.6	98.4	88
	6	SuperGLUE Human Baselines	SuperGLUE Human Baselines		89.8	89.0	95.8/98.9	100.0	81
+	7	T5 Team - Google	T5		89.3	91.2	93.9/96.8	94.8	88
	8	SPoT Team - Google	Frozen T5 1.1 + SPoT		89.2	91.1	95.8/97.6	95.6	87

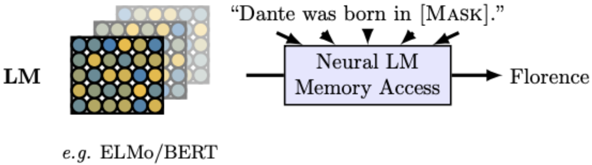
Why?



(Petroni et al, 2019)

- Learn a lot about language and the world

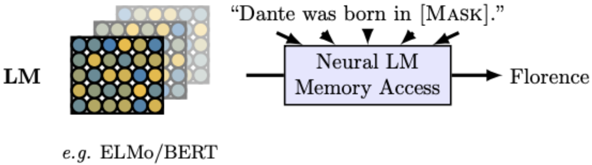
Why?



(Petroni et al, 2019)

- Learn a lot about language and the world
- but not everything

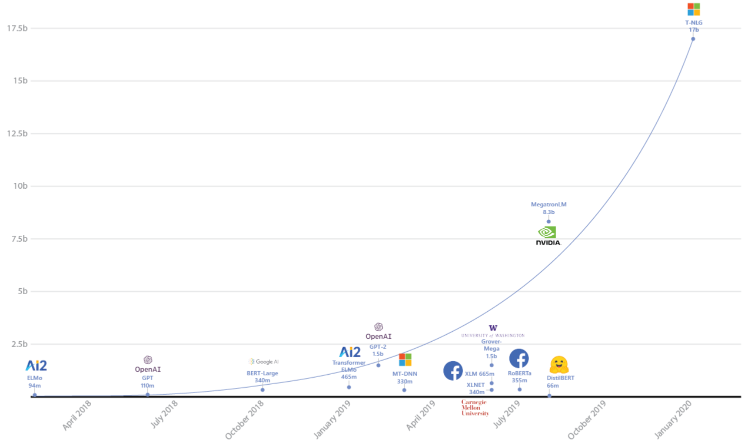
Why?



(Petroni et al, 2019)

- Learn a lot about language and the world
- but not everything
- Limited by the training data
- If trained on non-curated text from the internet, they tend to reflect the biases found there.

Size of LMs



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

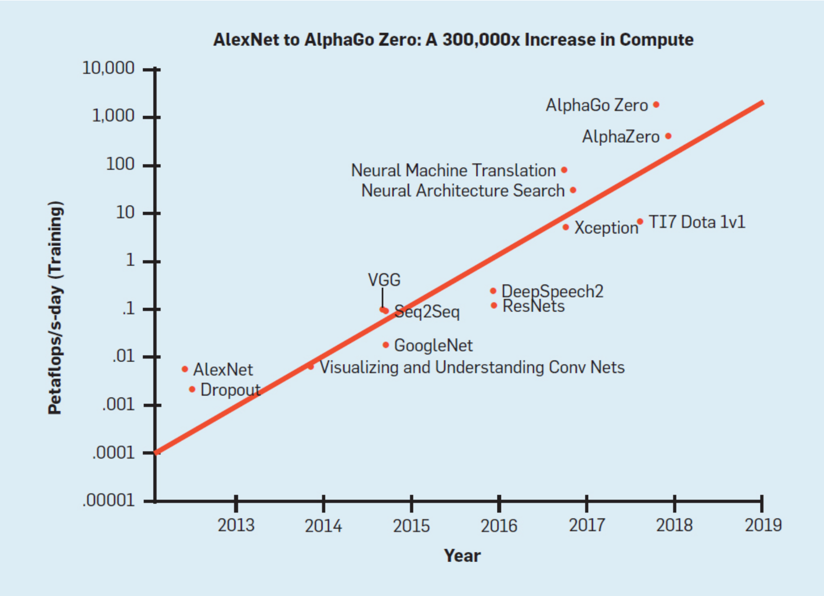
Size and data consumption

Year	Model	# of Parameters	Dataset Size
2019	BERT [39]	3.4E+08	16GB
2019	DistilBERT [113]	6.60E+07	16GB
2019	ALBERT [70]	2.23E+08	16GB
2019	XLNet (Large) [150]	3.40E+08	126GB
2020	ERNIE-Gen (Large) [145]	3.40E+08	16GB
2019	RoBERTa (Large) [74]	3.55E+08	161GB
2019	MegatronLM [122]	8.30E+09	174GB
2020	T5-11B [107]	1.10E+10	745GB
2020	T-NLG [112]	1.70E+10	174GB
2020	GPT-3 [25]	1.75E+11	570GB
2020	GShard [73]	6.00E+11	–
2021	Switch-C [43]	1.57E+12	745GB

Table 1: Overview of recent large language models

* from Bender & Gebru et al (2021): On the dangers of stochastic parrots: Can language models be too big?

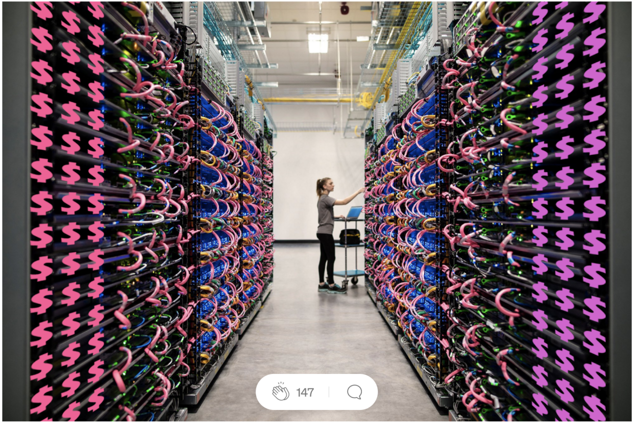
Compute



* from (Schwartz et al (2020): Green AI

Cost

The Staggering Cost of Training SOTA AI Models



https://medium.com/syncedreview/the-staggering-cost-of-training-sota-ai-models-e329e80fa82

Carbon footprint

Consumption	CO_2e (lbs)
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000

Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192
w/ neural architecture search	626,155

*from Strubell et al. (2019) Energy and policy considerations for deep learning in NLP.

Towards Green NLP?

How can we mitigate the negative effects of large LMs?

- SustaiNLP workshop
- Enhance reporting of computational budgets
- Efficiency as a core evaluation metric

* from (Schwartz et al (2020): Green AI