Towards Climate Awareness in NLP Research

Daniel Hershcovich Department of Computer Science University of Copenhagen dh@di.ku.dk

> Julia Anna Bingler ETH Zurich binglerj@ethz.ch

Nicolas Webersinke Mathias Kraus FAU Erlangen-Nuremberg {nicolas.webersinke, mathias.kraus}@fau.de

Markus Leippold University of Zurich markus.leippold@bf.uzh.ch

EMNLP 2022

UNIVERSITY OF COPENHAGEN

https://arxiv.org/abs/2205.05071

Climate impact of NLP

Models grow exponentially and take longer to train



Efficient Methods for Natural Language Processing: A Survey (Treviso et al., 2022)

1. Model publicly available?



REUSE

La CompVis/stable-diffusion-v1-4	
☞ • Updated about 10 hours ago • ↓ 809k • ♡ 3.15k	
<pre> bigscience/bloom </pre>	
♭ • Updated 24 days ago • ↓ 12.2k • ♡ 1.71k	
🖵 CompVis/stable-diffusion-v-1-4-origin	nal
☞ Updated 5 days ago 。 ♡ 1.4k	



https://huggingface.co/models

- 1. Model publicly available?
- 2. Time to train final model
- 3. Time for all experiments

		Q stacey > Projects > estuary ∟				
Z	(i)	Runs (107)				
	~	Q	=	ţţ		
		Name (56 visualized)	State	Tags	acc	Runtime -
	Ê	💿 🌒 50K examples (b 64)	finished		0.4042	1d 7h 56m 5s
	٤	💿 🛑 rmsprop 2GPU	finished	1GPU	0.4364	12h 39m 0s
	A	💿 🛑 4 GPU, b 32, e 50	finished	gpu and batch	0.6129	12h 34m 43s
EDUCE	0	💿 🛑 new_cluster	finished		0.4525	12h 31m 23s

https://wandb.ai/

Energy consumption scales with size

Model	Consumption [MWh]
1.3B	2.1
2.7B	4.8
6.7B	11.8
13B	22.9

A Holistic Assessment of the Carbon Footprint of Noor, a Very Large Arabic Language Model (Lakim et al., BigScience 2022)



- 1. Model publicly available?
- 2. Time to train final model
- 3. Time for all experiments
- 4. Power of GPU and CPU

	A100 80GB PCIe	A100 80GB SXM
GPU Memory	80GB HBM2e	80GB HBM2e
GPU Memory Bandwidth	1,935 GB/s	2,039 GB/s
Max Thermal Design Power (TDP)	300W	400W ***

https://www.nvidia.com/en-us/data-center/a100/#specifications

CO₂e footprint depends on time and location

CO2 Grams Emitted, BERT Language Modeling



Measuring the Carbon Intensity of AI in Cloud Instances (Dodge et al., FAccT 2022)

- 1. Model publicly availa
- 2. Time to train final mo
- 3. Time for all experime
- 4. Power of GPU and C
- 5. Location for computations
 6. Energy mix at location



https://lowcarbonpower.org/ map-gCO2eq-kWh



RELOCATE

- 1. Model publicly availa
- 2. Time to train final mo
- 3. Time for all experime
- 4. Power of GPU and C
- 5. Location for compute
- 6. Energy mix at locatic



Select an Experiment

Analyze Project

Showing results for experiment: Infrastructure Hosted at United States

Power Consumption : 0.0003 kWh Carbon Equivalent : 0.0001 kg

7. CO₂eq for final model
8. CO₂eq for all experiments

https://codecarbon.io/

or

Automated reporting

Convenient and reproducible

nist.py		
🛃 mnis	st.py ×	
1 (2	import tensorflow as tf	
3 (from codecarbon import Emission	
	C EmissionsTracker	codecarbon
	<pre>mnist = tf.keras.da</pre>	codecarbon
6	Press <₽ to insert, → to replace Next Tip	
	<pre>(x_train, y_train), (x_test, y_test) = mnist.load_data()</pre>	
8	x_train, x_test = x_train / 255.0, x_test / 255.0	
9		
	<pre>model = tf.keras.models.Sequential(</pre>	
	φ Γ	
	<pre>tf.keras.layers.Flatten(input_shape=(28, 28)),</pre>	
	<pre>tf.keras.layers.Dense(128, activation="relu"),</pre>	
15	<pre>tf.keras.layers.Dropout(0.2),</pre>	

But measures vary significantly between tools

<u>Evaluating the carbon footprint of NLP methods:</u> <u>a survey and analysis of existing tools</u> (Bannour et al., SustaiNLP 2021)

Greenhouse Gas Protocol



https://www.greenelement.co.uk/blog/carbon-footprint-scope-1-2-3/



- 1. Model publicly available?
- 2. Time to train final model
- 3. Time for all experiments
- 4. Power of GPU and CPU
- 5. Location for computations
- 6. Energy mix at location
- 7. CO₂eq for final model
- 8. CO₂eq for all experiments

9. Average CO₂eq for inference per sample

Climate awareness in NLP



Greenwashing?

"Better than some alternative" *≠* sustainable



Principles of climate performance assessment







RELEVANCE COMPLETENESS CONSISTENCY



TRANSPARENCY

ACCURACY

Pragmatic reporting



Carbon offsetting?





Mitigation by reduction is essential and feasible

Model card example

ClimateBert			
1. Model publicly available?	Yes		
2. Time to train final model	8 hours		
3. Time for all experiments	288 hours		
4. Power of GPU and CPU	0.7 kW		
5. Location for computations	Germany		
6. Energy mix at location	470 gCO ₂ eq/kWh		
7. CO_2eq for final model	2.63 kg		
8. CO_2eq for all experiments	94.75 kg		
9. Average CO_2 eq for inference per sample	0.62 mg		

ClimateBert: A Pretrained Language Model for Climate-Related Text (Webersinke et al., 2021)

Model card example

$MCWQ \ \texttt{mT5-base+RIR}$

Information	Unit
1. Model publicly available?	Yes
2. Time to train final model	592 hours
3. Time for all experiments	1315 hours
4. Energy consumption	2209.2 kWh
5. Location for computations	Denmark
6. Energy mix at location	191 gCO2eq/ kWh
7. CO2eq for final model	189.96 kg
8. CO2eq for all experiments	421.96 kg

<u>Compositional Generalization in Multilingual Semantic Parsing over Wikidata</u> (Cui et al., TACL 2022)



Some avenues for positive climate impact

Detecting greenwashing (e.g., <u>Bingler et al., 2022</u>)

Analyzing public discourse (e.g., <u>Hansen & Hershcovich</u>, <u>NLP4PI 2022</u>)

Understanding consumer behavior

Facilitating behavior change

How Good Is NLP? A Sober Look at NLP Tasks through the Lens of Social Impact (Jin et al., Findings 2021)

Summary



Climate awareness should be mainstream





Consistent and pragmatic reporting via model cards Code and templates: <u>github.com/danielhers/</u> <u>climate-awareness-nlp</u>



Efficient is not enough – net positive impact needed Contact: <u>danielhers.github.io</u> <u>dh@di.ku.dk</u>