THE SOCIAL COSTS OF AI

ETHICS, AI & ECONOMICS

23. MARCH 2022

AIMEE VAN WYNSBERGHE
IWE

TIM MENSINGER
BONN GRADUATE SCHOOL OF ECONOMICS

STEFAN HÖSE
DRZE

MARKUS FRANKE
IWE & DRZE

1. WHO AM I?

2. THE RESEARCH QUESTION

3. WHAT IS BERT AND WHY DO WE NEED IT?

4. PROJECT STATUS

Who am I?

WHO AM I?

- From 2020: PhD Student in Economics, BGSE

- Until 2020: Economics, Computer Science & Math

- Projects: Machine Learning in Economics
  - Modeling human decision making
  - Predicting the effects of programs

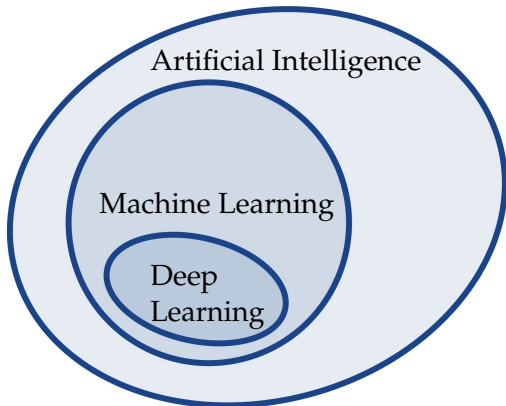- Research: Computational Econometrics

THE RESEARCH QUESTION

AI: An Example

Let's fix an example for this presentation

- Language modeling (NLP)

- Autocorrection on phones and computers

- Amazon Alexa etc.

- Google Translate etc.

EXAMPLE CONTD.

Artificial Intelligence

Machine Learning

Deep Learning

- AI: Teaching computers languages

- ML: Using specific models to do the above task

- DL: Using ML models but with **many** *parameters*

WHAT ARE PARAMETERS?

- Task: Write about your day

- Catch: You can only use max. 10, 100, or 1000 words

- Here: Number of words ≡ Number of parameters

THE PROBLEM

- **Many parameters** $\implies$ large computational costs

- Large computational costs $\implies$ large energy consumption

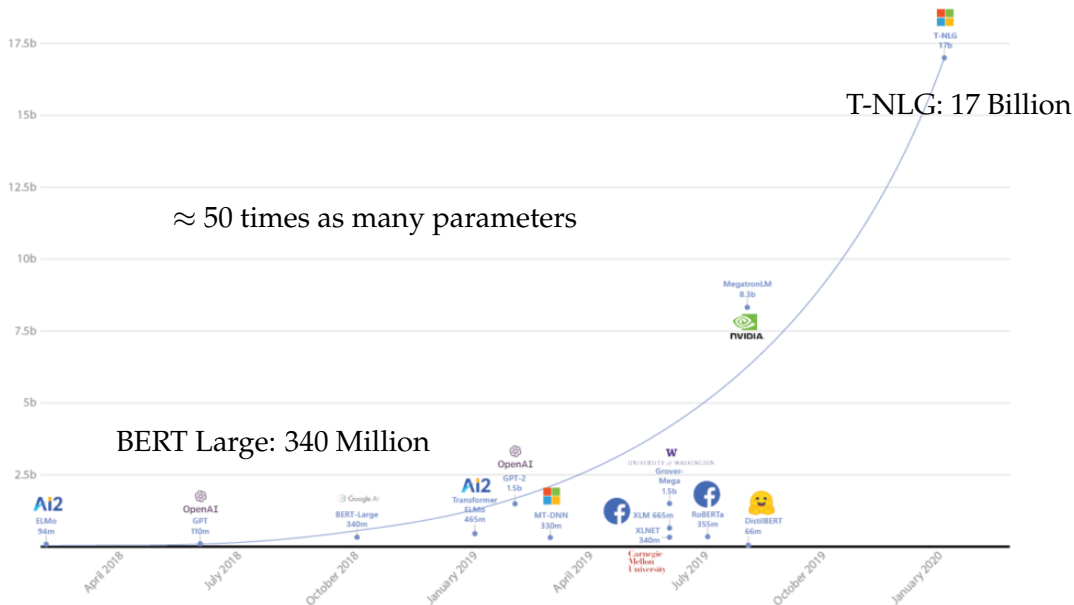- Large energy consumption $\implies$ **large CO_2 emissions**

ARE THE EMISSION REALLY THAT LARGE?

ARE THE EMISSION REALLY THAT LARGE?

Yes!

Training a large NLP model $\approx$ 125 round-trip flights New York - Beijing
—Strubell et al. (2020)

T-NLG: 17 Billion

$\approx$ 50 times as many parameters

BERT Large: 340 Million

11

BUT WHAT CAUSES THE EMISSION?

For state-of-the-art large models:

- **(Pre-) Training**
  $\approx$ 1 month on 1,000 computers; Narayanan et al. (2021)

- **Fine-Tuning**
  $\approx$ few days on several computers; Devlin et al. (2018)

- **Deployment**
  $\approx$ millisecond

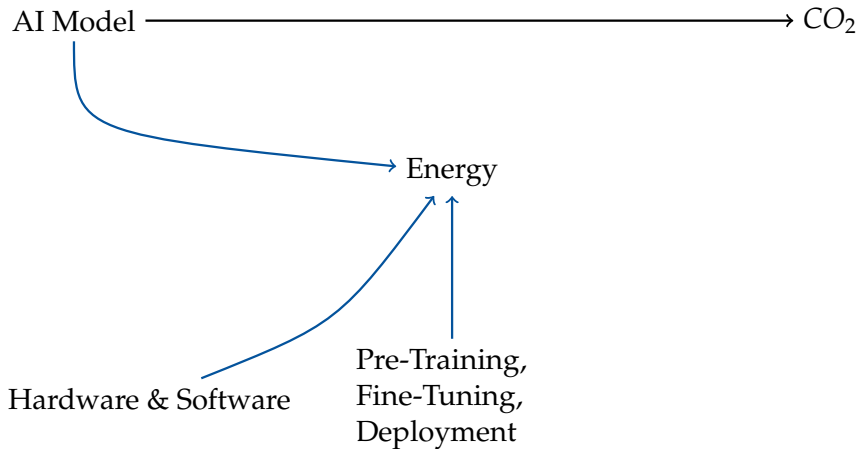# HOW OFTEN DO THESE STAGES HAPPEN?

Let's consider a language model used by Google:

- **(Pre-) Training and Model Development**
  ≈ 100 times (*further resources needed*)

- **Fine-Tuning and Model Development**
  $\approx 1{,}000$ times (*further resources needed*)

- **Deployment**
  ≈ 5 billion Google searches every day
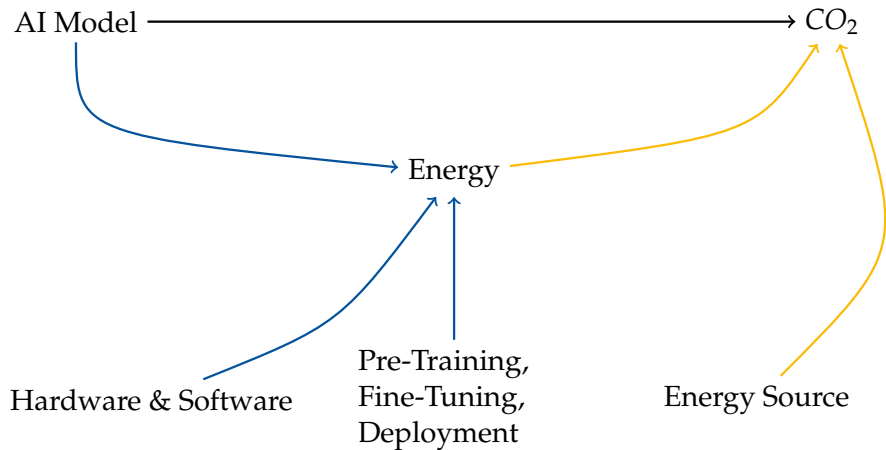
THE PROBLEM CONTD.

AI Model $\longrightarrow CO_2$

THE PROBLEM CONTD.

AI Model ───────────────────────→ CO_2

Energy

Hardware & Software

Pre-Training,
Fine-Tuning,
Deployment

AI Model $\longrightarrow$ CO_2

Energy

Hardware & Software

Pre-Training,
Fine-Tuning,
Deployment

Energy Source

RESEARCH QUESTIONS

- How do we quantify emissions of AI models?

- Can we nudge the AI community towards sustainability?

What is BERT and why do we need it?

BERT

- Language model developed at Google by Devlin et al. (2018)

- Is used e.g. for Google search queries

- Main example in Strubell et al. (2020) and Schwartz et al. (2020)

- We use it, because
  - Perfect model size for our resources
  - The literature is already familiar with the model
  - We can compare our results to the literature

PROJECT STATUS

PROJECT STATUS

- Build test computing infrastructure ✔
  - Setup energy measurement architecture ✔

  - Setup computing environment ✔

  - Run BERT model ⟳

- Build real computing infrastructure ✖
 - Get new computing hardware ✖

- Analyze energy data ✖

References I

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*.

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*.

Kaack, L. H., Donti, P., Strubell, E., & Rolnick, D. (2020). Artificial intelligence and climate change: Opportunities, considerations, and policy levers to align ai with climate change goals. *Invited policy brief, Heinrich Böll Foundation*.

Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2021). Aligning artificial intelligence with climate change mitigation. *Working paper*.

Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the carbon emissions of machine learning.

Narayanan, D., Shoeybi, M., Casper, J., LeGresley, P., Patwary, M., Korthikanti, V., Vainbrand, D., Kashinkunti, P., Bernauer, J., Catanzaro, B., Phanishayee, A., & Zaharia, M. (2021). Efficient large-scale language model training on GPU clusters. *CoRR*.

REFERENCES II

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*.

Strubell, E., Ganesh, A., & McCallum, A. (2020). Energy and policy considerations for modern deep learning research. *Proceedings of the AAAI Conference on Artificial Intelligence*.

Contact
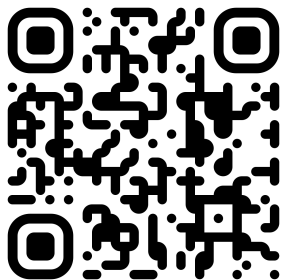
tmensinger.com

tmensinger@uni-bonn.de

github.com/timmens/social-cost-ai