Python for beginners: presentation and organization of the course

Karën Fort

karen.fort@loria.fr / https://members.loria.fr/KFort

iDMC Institut des sciences du Digital Management & Cognition
UFR COMPOSANTE DE L'UNIVERSITÉ DE LORRAINE

UNIVERSITÉ DE LORRAINE

Karën Fort (she/her)

You?

I say what I do

and I do as I say

- everybody arrives on time
- everybody attends all the classes (please email me in case of emergency)
- do not change groups
- if you speak, speak to **us**
- no social network, no email, no (smart)phone, no food

- do not hesitate to ask questions: there are dummy questions, but I'll answer them too (once)!
- do not hesitate to tell me if:
  - I'm going too fast|not fast enough
  - you already know what I'm talking about
  - my English is not good (correct me, please)
- do not hesitate to question what I'm saying

How to contact me

- email: karen.fort@loria.fr
- office: at IDMC or LORIA (B116), email before

Material

All my material is available online:

- on my website: https://members.loria.fr/KFort/idmc-nancy-from-2024/
- not on Arche
- under a CC licence

Behavior

Some basic politeness rules:

- when you write an email, salute the person
- when you send a piece of work, name it so that we can find it (*Name-NumTD.doc*, for example)
- listen carefully to what we decide

Netiquette

https://en.wikipedia.org/wiki/Etiquette_in_technology

Organization of the course

- 8 weeks, 4h a week:
  - 2h lecture with practice (CM), with me, followed by
  - 2h practice (TD), with me and Nasser-Eddine Monir (nasser-eddine.monir@inria.fr)

Evaluation

- 50% TD:
  - all TD (except the 1st one) are to be done and put on Arche
  - 3 (randomly picked) TD will be graded
  - the worst grade *can* be removed
- 50% exam:
  - on paper (no machine)

About TD (labs)

- work by pairs (or alone)
- never the same pair
- the work should be finished in class (no homework)
- your names should appear in the file name:
  - TD0-FortMonir.ipynb (sometimes, it will be .py)
- send your work to your teacher by email at the end of the class:
  - karen.fort@loria.fr
  - nasser-eddine.monir@inria.fr

On using (or not) ChatGPT

- ChatGPT is not a search engine
- ChatGPT presents other issues...

ChatGPT Carbon footprint: training

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

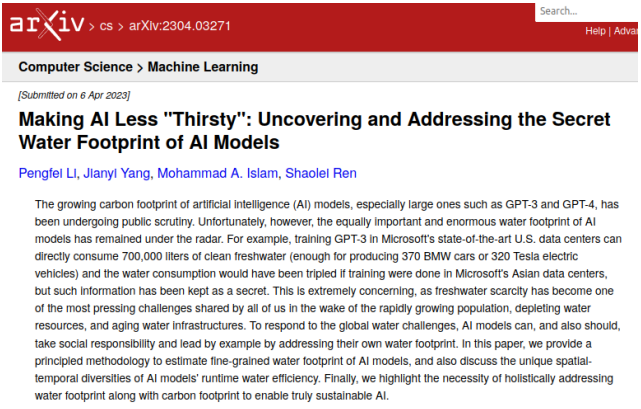
Table 1: Estimated CO_2 emissions from training common NLP models, compared to familiar consumption.[1]

[Strubell et al., 2019]

Note: this concerns only 1 source out of four [Bannour et al., 2021]

→ inference estimated at 24.86 tCO_2e per day

ChatGPT water consumption



arXiv > cs > arXiv:2304.03271

Search...

Help | Adva

Computer Science > Machine Learning

[Submitted on 6 Apr 2023]

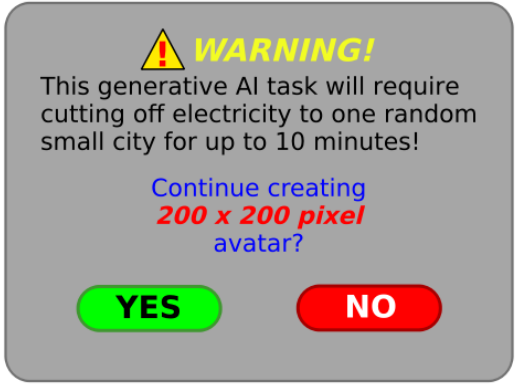
Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models

Pengfei Li, Jianyi Yang, Mohammad A. Islam, Shaolei Ren

The growing carbon footprint of artificial intelligence (AI) models, especially large ones such as GPT-3 and GPT-4, has been undergoing public scrutiny. Unfortunately, however, the equally important and enormous water footprint of AI models has remained under the radar. For example, training GPT-3 in Microsoft's state-of-the-art U.S. data centers can directly consume 700,000 liters of clean freshwater (enough for producing 370 BMW cars or 320 Tesla electric vehicles) and the water consumption would have been tripled if training were done in Microsoft's Asian data centers, but such information has been kept as a secret. This is extremely concerning, as freshwater scarcity has become one of the most pressing challenges shared by all of us in the wake of the rapidly growing population, depleting water resources, and aging water infrastructures. To respond to the global water challenges, AI models can, and also should, take social responsibility and lead by example by addressing their own water footprint. In this paper, we provide a principled methodology to estimate fine-grained water footprint of AI models, and also discuss the unique spatial-temporal diversities of AI models' runtime water efficiency. Finally, we highlight the necessity of holistically addressing water footprint along with carbon footprint to enable truly sustainable AI.

$\rightarrow$ 700,000 l. of water for the training of such a model

$\rightarrow$ 1/2 l. of water for 30 pages generated, not counting the equipment production


WARNING!
This generative AI task will require cutting off electricity to one random small city for up to 10 minutes!
Continue creating
200 x 200 pixel
avatar?
YES
NO

Suggestions

Use:

- https://stackoverflow.com/ (used to train ChatGPT, the source)
- https://www.w3schools.com/python/default.asp
- Your favorite search engine

About this course

- it will not cover everything you'll need
- it should help you help yourself (find your own answers)
- however, I tried to take suggestions from previous students into account

Let's code!

- this and the notebooks are/will be available **on my website**, also as HTML
- you can/should use your own notebook
- for this course (only), you can use an online notebook, like: `https://colab.research.google.com`
- from next week, you **should** have Jupyter Notebook installed (install party on 4th/5th/6th)

Bannour, N., Ghannay, S., Névéol, A., and Ligozat, A.-L. (2021).
Evaluating the carbon footprint of NLP methods: a survey and analysis of existing tools.
In EMNLP, Workshop SustaiNLP, Punta Cana, Dominican Republic.

Strubell, E., Ganesh, A., and McCallum, A. (2019).
Energy and policy considerations for deep learning in NLP.
In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 3645–3650, Florence, Italy. Association for Computational Linguistics.