CC BY NC SA

Outils pour l'analyse linguistique : présentation et organisation du cours

Karën Fort

karen.fort@univ-lorraine.fr / https://members.loria.fr/KFort

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

UNIVERSITÉ DE LORRAINE

Introduction
D'où je parle

Organisation

Karën Fort (elle)

et vous ?

Introduction

Organisation
- Contrat d'enseignement
- Cours
- Bibliographie

Je dis ce que je fais

et je fais ce que je dis

- tout le monde arrive à l'heure
- tout le monde assiste à tous les cours (me signaler rapidement les cas particuliers)
- si vous parlez, parlez-**nous**
- pas de réseau social, pas de mail, pas de téléphone, pas de nourriture

- n'hésitez pas à poser des questions : il y a des questions idiotes, mais j'y répondrai aussi (une fois) !
- n'hésitez pas à me dire si :
  - je vais trop|pas assez vite
  - vous connaissez déjà ce dont je parle
- n'hésitez pas à me remettre en cause

Me contacter

- courriel (pour les devoirs ou en cas de question) : karen.fort@univ-lorraine.fr
- bureau : 215
- LORIA : B116

Comportement

Rappel de quelques règles de politesse :

- quand vous écrivez un courriel, saluez au début et à la fin
- quand vous envoyez un devoir, nommez le fichier intelligemment (*Nom-NumTD.doc*, par exemple)
- écoutez les consignes attentivement

Nétiquette

http://fr.wikipedia.org/wiki/N%C3%A9tiquette

Organisation du cours

- 19 h, mélange de CM et TD :
  - principalement des séances de 3h
  - deux séances de 2h

- évaluation :
  - examen sur table

Plan provisoire du cours

- intro + corpus
- encodages
- rappels de TAL
- TXM 1 / TXM Zipf
- TXM 2 / TXM MPT
- GREW
- GREW multilingue
- TAL récent

À propos de l'utilisation de ChatGPT pour ce cours

- ChatGPT n'est pas un moteur de recherche
- ChatGPT présente d'autres problèmes...

Empreinte carbone de l'entraînement de GPT 2

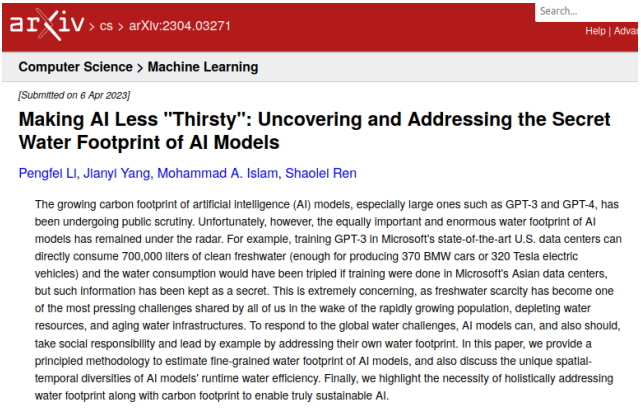
Consumption	CO_2e (lbs)
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000
Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192
w/ neural architecture search	626,155

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

[Strubell et al., 2019]

Note : ne concerne qu'une source sur 4 [Bannour et al., 2021]

Consommation d'eau de ChatGPT



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. d'eau pour entraîner un modèle de ce type

$\rightarrow$ 1/2 l. d'eau toutes les 30 pages générées, sans compter la production de l'équipement

Qui de l'inférence (utilisation) ?

- Selon OpenAI, l'impact d'une requête chatGPT est de 4,32 g. CO2
  - Selon un rapport de Google publié en 2009, l'impact d'une requête Google est estimé à 0,2 g. CO2
  - l'impact carbone d'une requête chatGPT est **22 fois plus élevé** que celui d'une requête classique sur un moteur de recherche

avec l'autorisation d'Aurélie Névéol.
https://piktochart.com/blog/carbon-footprint-of-chatgpt/

Suggestions

Utilisez :

- Pour les questions de code : `https://stackoverflow.com/` (utilisé pour entraîner ChatGPT, la source des solutions, donc)
- Votre moteur de recherche de base

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.