Towards a Green AI

Evolutionary solutions for an ecologically viable artificial intelligence

Nayat Sánchez-Pi and **Luis Martí**
Inria Research Center in Chile,
Santiago de Chile, Chile.

GECCO '21 Companion, July 10–14, 2021, Lille, France

ACM ISBN 978-1-4503-8351-6/21/07...$15.00
https://doi.org/10.1145/3449726.3461428

Instructors

Nayat Sánchez-Pi is currently the Director and CEO of the Inria Chile Research Center, created in 2012 by Inria, the French National Research Institute for Digital Sciences to facilitate scientific and industrial cooperation between France, Chile, and Latin America. Before that, she was a professor of Artificial Intelligence and Human-Computer Interaction at the Department of Informatics and Computer Science of the Institute of Mathematics and Statistics of the Rio de Janeiro State University. Prof. Sánchez-Pi's research interests have broadened over the years and span topics that range from artificial intelligence, machine learning, and data mining to ambient intelligence, ubiquitous computing, and multi-agent systems. She received a degree in Computer Science in 2000 from the University of Havana and a Ph.D. degree in Computer Science in 2011 from the Universidad Carlos III de Madrid. She has led numerous research projects applying evolutionary computation, machine learning, and other artificial intelligence methods.

Luis Martí is currently the scientific director of Inria Chile, the Chilean Center of Inria, the French National Institute for Computational Sciences. Before that, he was a senior researcher of the TAU team at Inria Saclay since 2015. He was also an Adjunct Professor (tenured) at the Institute of Computing of the Universidade Federal Fluminense. Previous to that, Luis was a CNPq Young Talent of Science Fellow at the Applied Robotics and Intelligence Lab of the Department of Electrical Engineering of the Pontifícia Universidade Católica do Rio de Janeiro, Brazil. Luis did his Ph.D. at the Group of Applied Artificial Intelligence of the Department of Informatics of the Universidad Carlos III de Madrid, Madrid, Spain, and got his Computer Science degree from the University of Havana. He is mainly interested in artificial intelligence, and, in particular, machine learning, neural networks, evolutionary computation, optimization, machine learning, hybrid systems, and all that.

Computing for efficient green 'stuff'

OPTIMIZATION
Plant and export layouts.
Robustness.

MODELING
Wind and wave modeling for solar, wind and tidal energies

PREDICTION
Being able to predict the demand and production leads to efficient production.

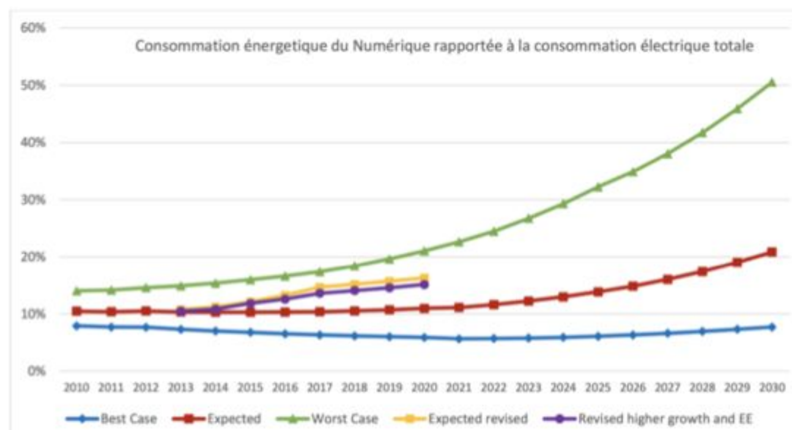


Figure 2 : Évolution 2010-2020 de la consommation énergétique du Numérique rapportée à la consommation électrique mondiale[7]
[Source: calculé par The Shift Project à partir des données publiées par Andrae et Edler (2015)]

AI FOR HUMANITY

L'INTELLIGENCE ARTIFICIELLE
AU SERVICE DE L'HUMAIN.

Energy consumption by the digital sector could increase tenfold by 2030, accounting for between

20 and 50%

of global electricity use.

5
7

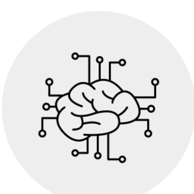
Cédric Villani, Marc Schoenauer, Yann Bonnet, Charly Berthet, Anne-Charlotte Cornut, François Levin, Bertrand Rondepierre. **AI for Humanity.** https://www.aiforhumanity.fr

Inria Chile

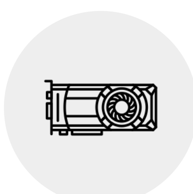
Ok, computing can help us reduce the environmental footprint...

...but, what about the impact of computing itself?

Inria Chile



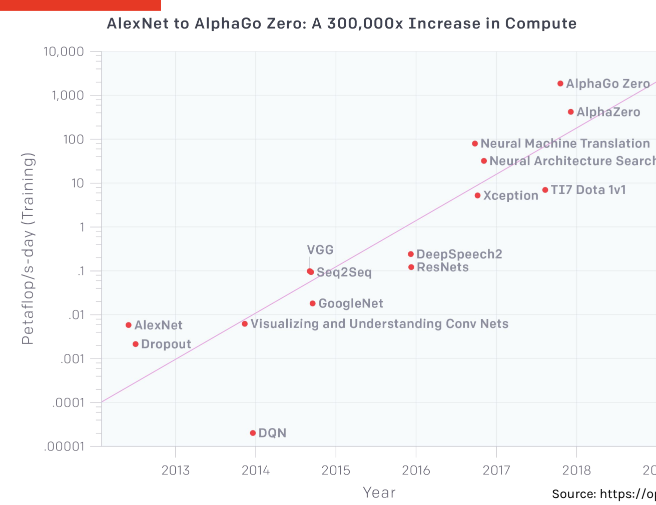
BETTER MODELS
Deep, reinforcement, transfer, active learning, etc.



BETTER HARDWARE
GPUs were developed for rendering graphics but are the best hardware for machine learning.

BETTER DATASETS
Important effort to create challenges datasets: vision, NLP, etc.

Inria Chile



Source: https://openai.com/blog/ai-and-compute/

Inria Chile

Direct impact of state of the art of NLP deep learning methods

CO_2 Emissions of deep learning

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

Emma Strubell, Ananya Ganesh and Andrew McCallum. **Energy and Policy Considerations for Deep Learning in NLP.** Annual Meeting of the Association for Computational Linguistics (ACL short). Florence, Italy. July 2019.

Inria Chile

Common carbon footprint benchmarks

in lbs of CO2 equivalent

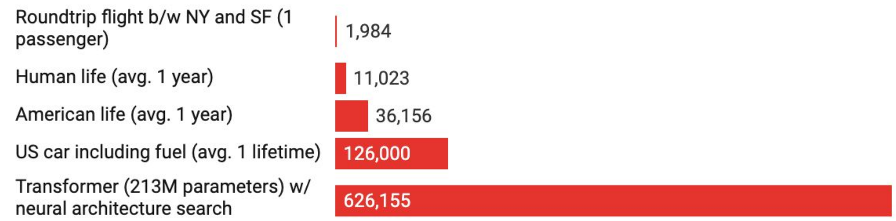


Chart: MIT Technology Review • Source: Strubell et al. • Created with Datawrapper

https://www.technologyreview.com/s/613630/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/

Inria Chile

	Date of original paper	Energy consumption (kWh)	Carbon footprint (lbs of CO2e)	Cloud compute cost (USD)
Transformer (65M parameters)	Jun, 2017	27	26	\$41-\$140
Transformer (213M parameters)	Jun, 2017	201	192	\$289-\$981
ELMo	Feb, 2018	275	262	\$433-\$1,472
BERT (110M parameters)	Oct, 2018	1,507	1,438	\$3,751-\$12,571
Transformer (213M parameters) w/ neural architecture search	Jan, 2019	656,347	626,155	\$942,973-\$3,201,722
GPT-2	Feb, 2019	-	-	\$12,902-\$43,008

Note: Because of a lack of power draw data on GPT-2's training hardware, the researchers weren't able to calculate its carbon footprint.

Table: MIT Technology Review • Source: Strubell et al. • Created with Datawrapper

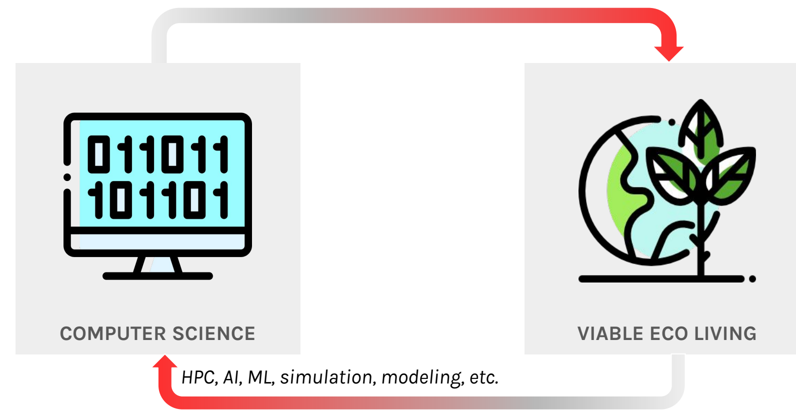
https://www.technologyreview.com/s/613630/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/

Inria Chile

The solution

Inria Chile

Green AI



Measuring AI's footprint

- Carbon emissions generated.
- Electric energy consumed
- Elapsed time used for execution
- Number of parameters (weights) and hyperparameters.
- Floating point operations.
- Focused on the cost of training.

The solution is a multi-step solution

Better Hardware

It is unlikely that we get right of GPUs (or TPUs) at training time.

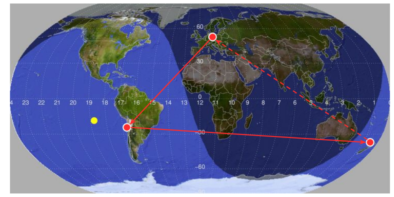
There are **hardware alternatives** at **use time**:

- ➔ Field Programmable Gate Arrays (FPGAs), Application-Specific Instruction-set Processors (ASIPs), etc.

We should also keep exploring the use of ***low-precision computing:***

Reducing the quality (and therefore length) of the **floating-point representation of numbers.**

Self-scaling and cloud computing



Self-scaling computing facilities make available a pool of shared resources.

Optimally schedule computing time.

Cloud computing allows **to pick the location** where programs will be run.

Code is mobile!

We can, for example, "track the sun" and ensure that the AI/ML processes use renewable sources.

We need cloud (computing) transparency!

Smarter experiments

Finding the **right configuration** of the **hyperparameters** probably where more energy is consumed.

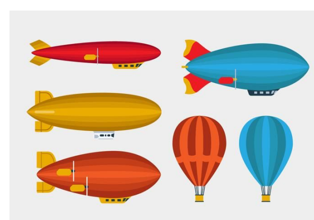
This is an optimization problem => NP-Hard problem

However, better approaches like **evolutionary computing** are here to help!

...but they need populations of individuals, hence more energy.

This is a multi-objective optimization problem!

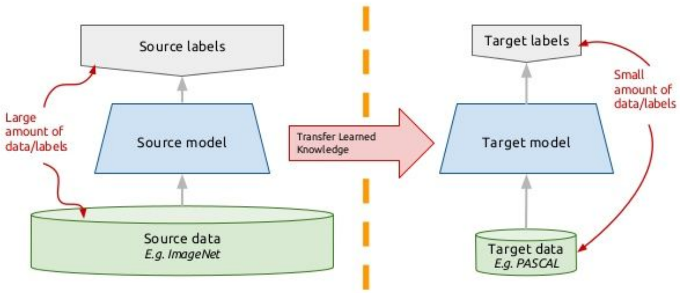
Self adaptation



To look for **methods** that **adapt** their complexity **automatically** to the **complexity** of the **problem** being solved.

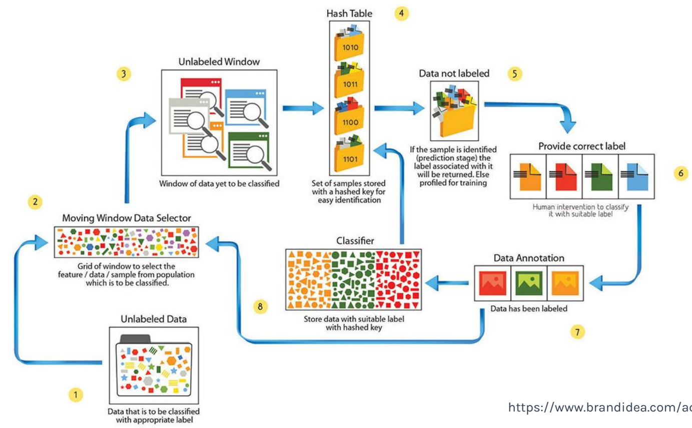
Neural networks based on **adaptive resonance theory (ART)** and **growing neural gas (GNG)** have rules to adapt themselves to the complexity of the problem.

Model reuse and transfer learning



Inria Chile

Active learning and sample efficiency



https://www.brandidea.com/activeLearning.html

Inria Chile

Final remarks

- We should not only work in solving key problems in the area of green energies.
- We should also readdress computing under an environment-aware point of view.
  - to report the **budget/accuracy curve** observed that shows how much energy needs to be used to achieve the results.
  - releasing pretrained models to save others the cost of pre-training.

Inria Chile

Inria

Merci ! - Obrigado! - Thank you! - ¡Gracias!

Questions?