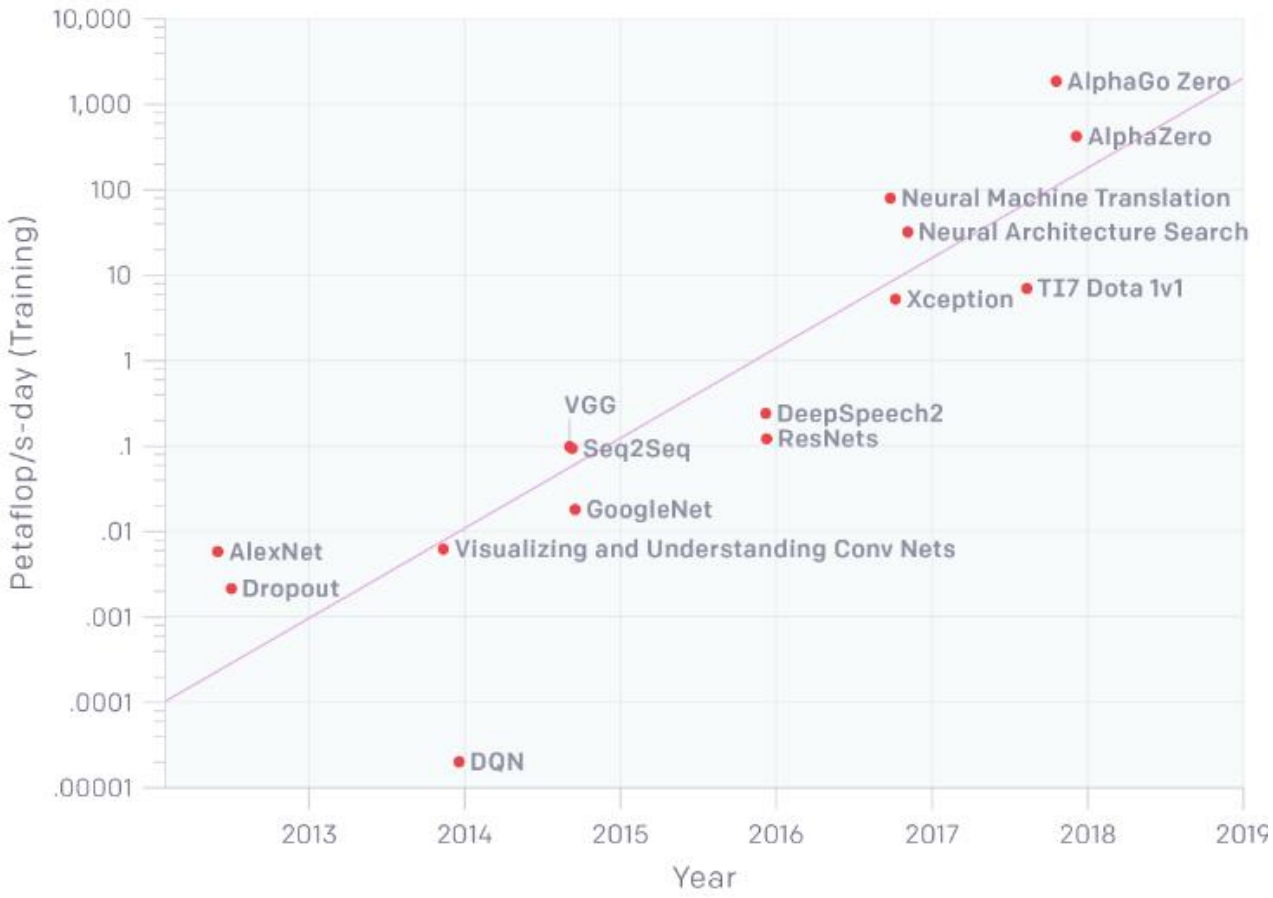
# Green AI

Roy Schwartz, Jesse Dodge, Noah A. Smith, Oren Etzioni

**2019**

# Introduction

- The amount of compute used to train deep learning models has increased 300,000x in 6 years. (Far exceed Moore's Law)
- The rising of leaderboard:
  - Hype around state of the art accuracy results
  - No mention of cost or efficiency



# Red AI

- Refers to seeking state-of-the-art result focusing only on accuracy
- Use massive computational ressources
  - Essentially buying better results
  - Concerning social impact : who is making the decision ? Concentration of power, rich get richer

- Though Red AI brought valuable insight and knowledge, it is way too dominant (and not sustainable AT ALL)
  - Personal addenda : Google recently announced contributing to private small nuclear reactor to feed AI

[1]:https://www.reuters.com/technology/artificial-intelligence/google-buy-power-small-modular-nuclear-reactor-company-kairos-ai-needs-2024-10-14/

# **Red AI** Dominance

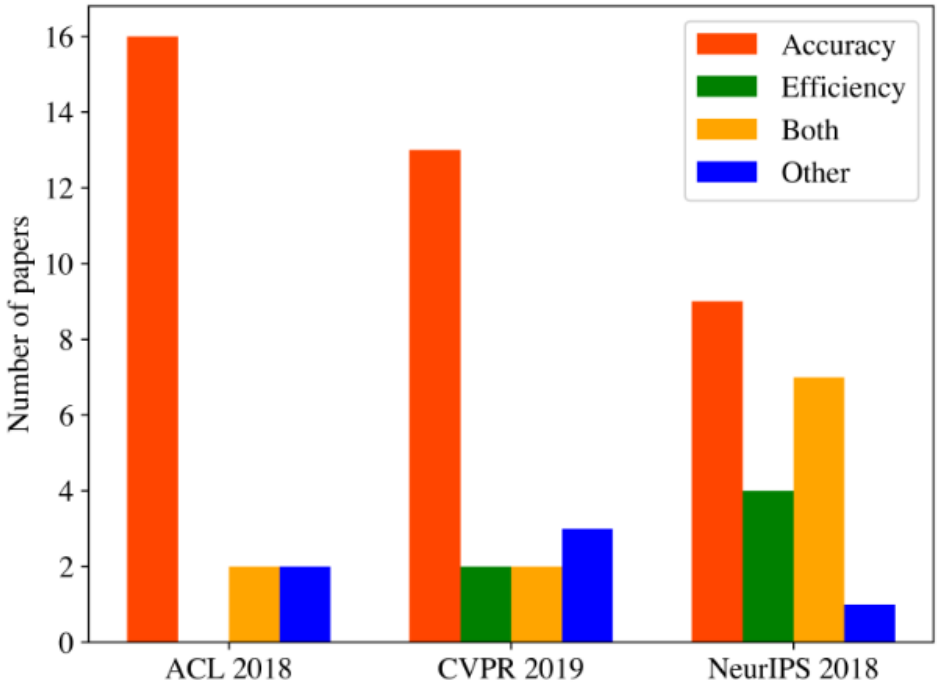



Figure 2: AI papers tend to target *accuracy* rather than *efficiency*. The figure shows the proportion of papers that target accuracy, efficiency, both or other from a sample of 60 papers from top AI conferences.

# **Red AI** Cost

$$Cost(R) \propto E \cdot D \cdot H$$

Equation 1: The equation of Red AI: The cost of an AI ($R$)esult grows linearly with the cost of processing a single ($E$)xample, the size of the training ($D$)ataset and the number of ($H$)yperparameter experiments.

- Processing a single (E)xample
  - XLNet (Bert-large + custom losses + more data) : 512 TPU chips for 2.5 days
  - AlphaGo : 1,920 CPUs and 280 GPUs to play a single game of Go at a cost of over $1,000 per hour. (Best version)
- Size of the training (D)ataset
  - Common Crawl contains 242 TB, expensive to store !
  - Benefit of increasing dataset size on accuracy seems to be logarithmic at best
- Number of (H)yperparameter experiments
  - Hardest part to quantify as a full R&D cycle use a lot of iteration before doing a full HP search
  - Google trained 12,800 neural net for object detection and language modelling task

# Green AI

- Goal : yield novel results **without** increased costs (and ideally reducing)
- A first step for **Green AI** is to define and use efficiency metrics

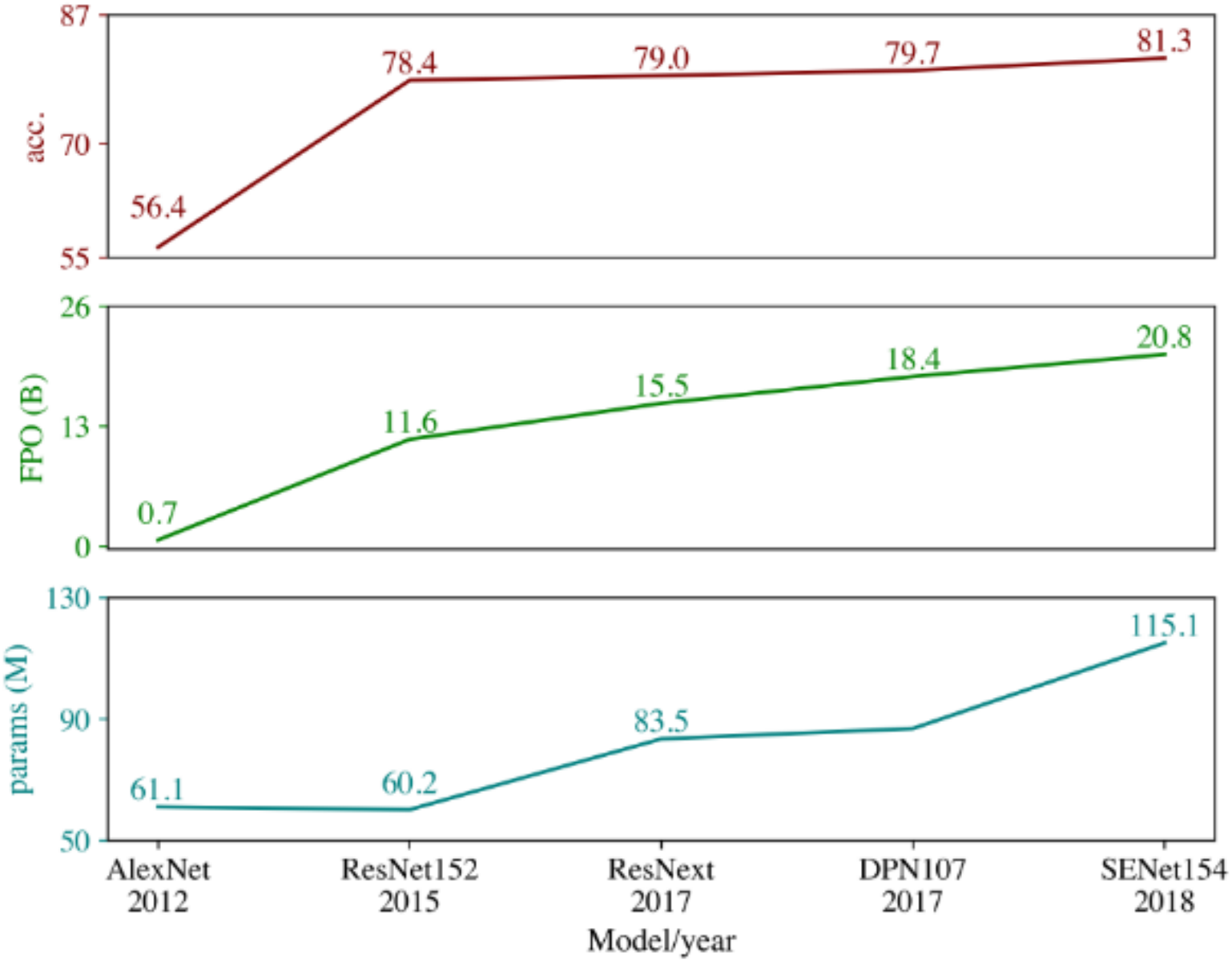
# Green AI : Unreliable Metrics

- Carbon footprints : ultimately part of what we need to minimize but it is highly dependent of the localization and infrastructure
  - Personal addenda : the paper mention that it is what we want to "directly minimize" though recent paper also point an important Water Footprint
- Electricity usage : correlated with carbon while being time and localization independent but it is still hardware dependent
- Elapsed real time : natural measure for efficiency but it depends on the hardware and the current running softwares (libraries, concurrent jobs etc.)

# **Green AI** : Independent Metrics

- Number of parameters : correlated with the amount of work and used memory and does not depend on the hardware ! Though it's not easily comparable across architecture
- Floating Point Operation (FPO) : define base cost for *ADD* and *MUL* operations and recursively compute the total cost.
  - Directly compute the amount of work done (tied to energy consumption)
  - Agnostic to hardware, localisation and time
  - Strongly correlated with run time
  - Ignore memory consumption and depends on exact implementation…

**Green AI** :
Example of
Analysis




(a) Different models.

# Related work

- Long term trends have led the field of machine learning to an unsustainable position (not only with LLMs)
- Some company tries to offset their footprint by purchasing carbon credits (controversed), it is not clear that buying carbon credit is effective and it is only voluntary based.
- Machine learning can also be used as a tool to tackle climate change

# Conclusion

- **Green AI** raise many exciting research directions
- It is a valuable option and we should increase it's prevalence but both **Green** and **Red** AI both have contributions to make

Two potential research directions are:

- Research on data efficiency
- Effort to report number of R&D experiments

# Personal Conclusion

- Good article with lots of interesting points
- Very good word color idea

- Stuck in a "Carbon-only" mindset
- Written in 2019 so it cannot question current hype around foundational models