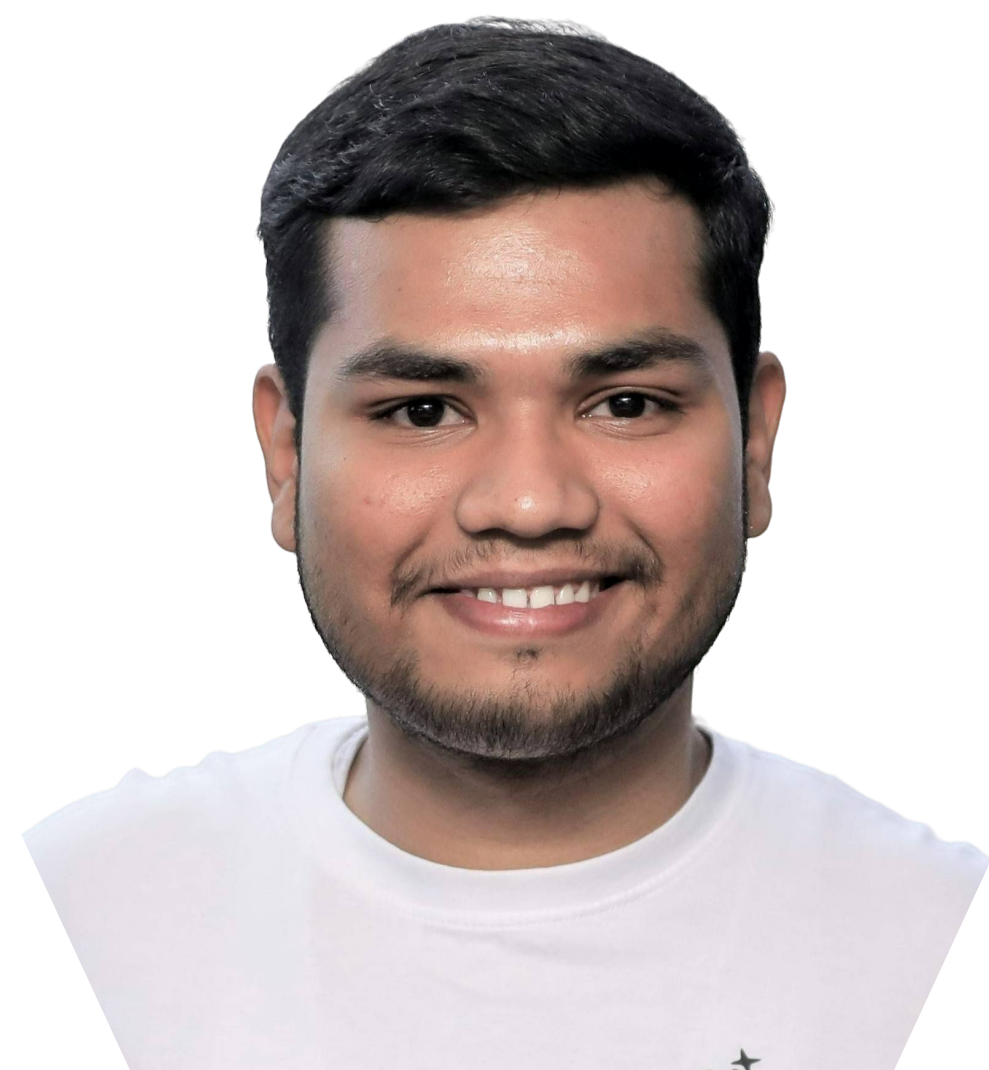


Diving into Sust**AI**nability

Date : 26th August 2023 | Speaker : Ayon Roy

Event : AI Catalyst Party by MLSA MIET | Venue : Microsoft Gurgaon, India



Hello World!

I am Ayon Roy

Executive Data Scientist @ NielsenIQ

Z by HP Global Data Science Ambassador

Mentored/Judged **100+** Hackathons

Delivered **70+** Technical Talks

Brought **Kaggle Days Meetup** Community in India for the 1st time

If you haven't heard about me yet, you might have been living under the rocks. Wake up !!

Agenda

- Evolution of AI & its approaches
- The necessity for doing AI sustainably
- Brief about United Nations SDGs [[Sustainable Development Goals](#)]
- Exploring ways to use AI for achieving SDGs
- How can you approach **SustAInability**

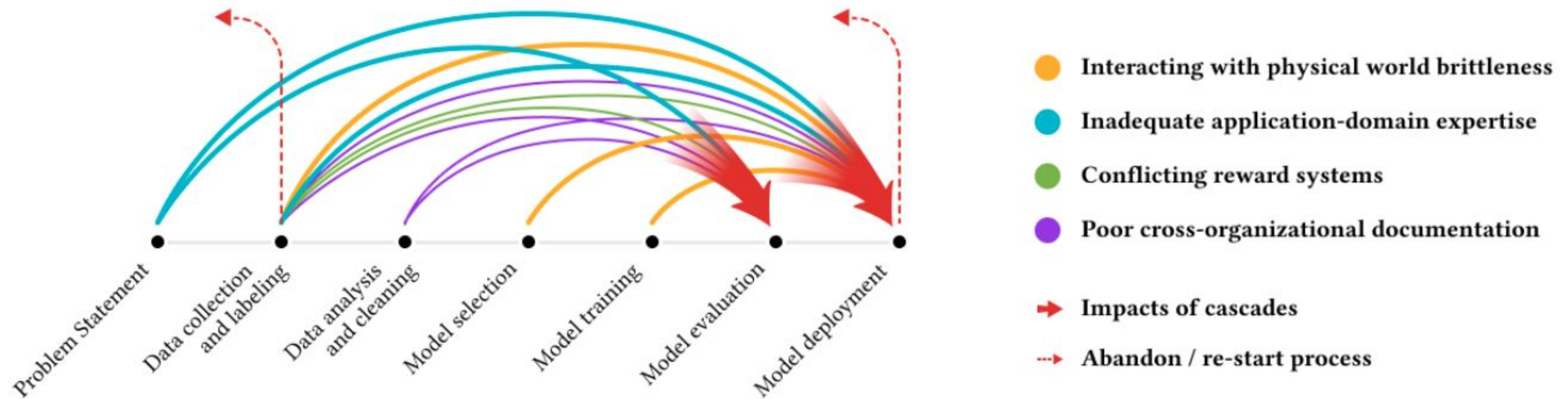
AI Systems

Artificial Intelligence Systems are projects which are undertaken with the long-term goal of simulating the human brain in real time, complete with artificial consciousness and artificial general intelligence.

How do we simulate the human brain in real time & bring artificial consciousness ?

Data + Model (Algorithms) + Compute

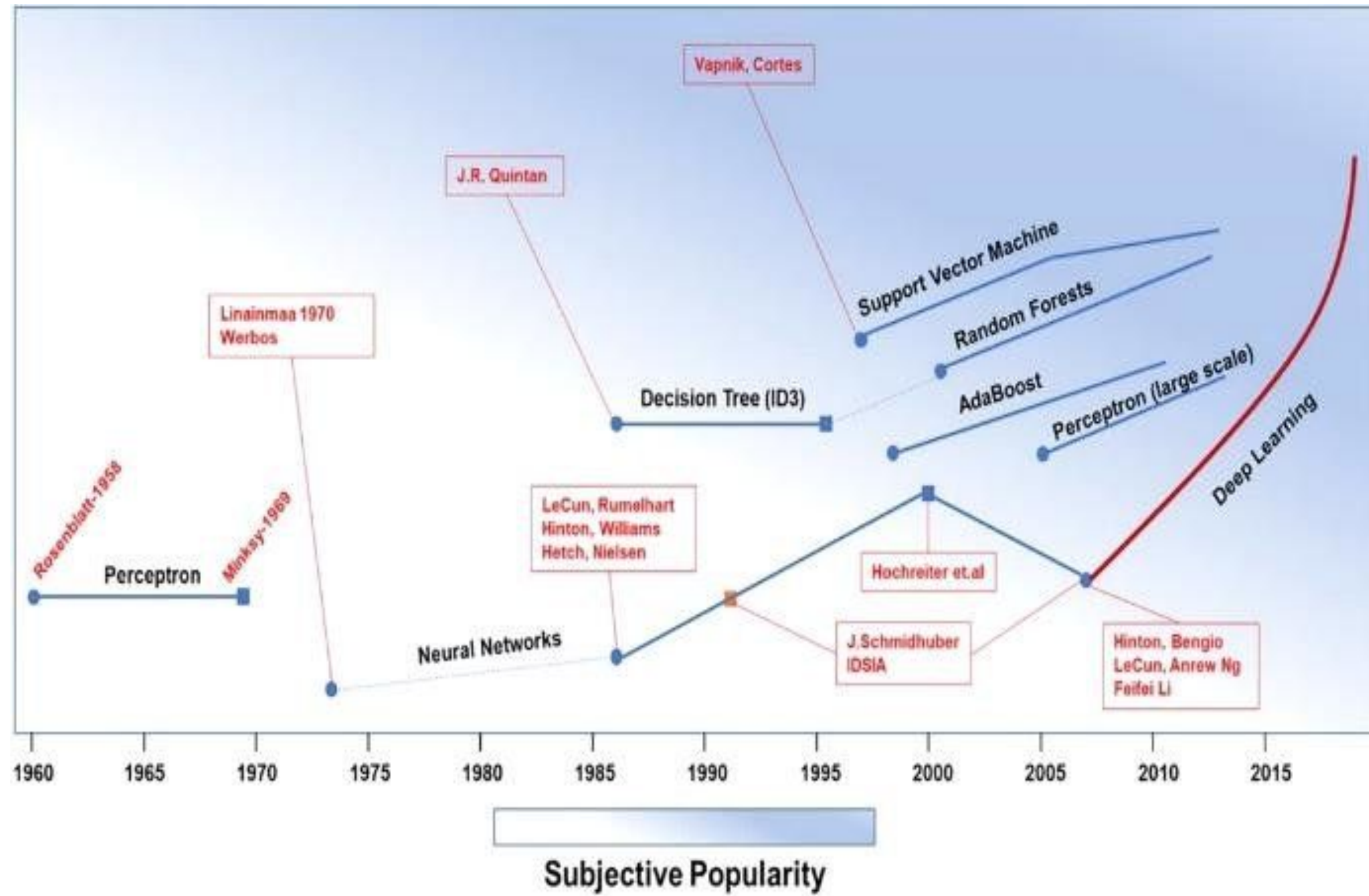
Stages to build AI Systems



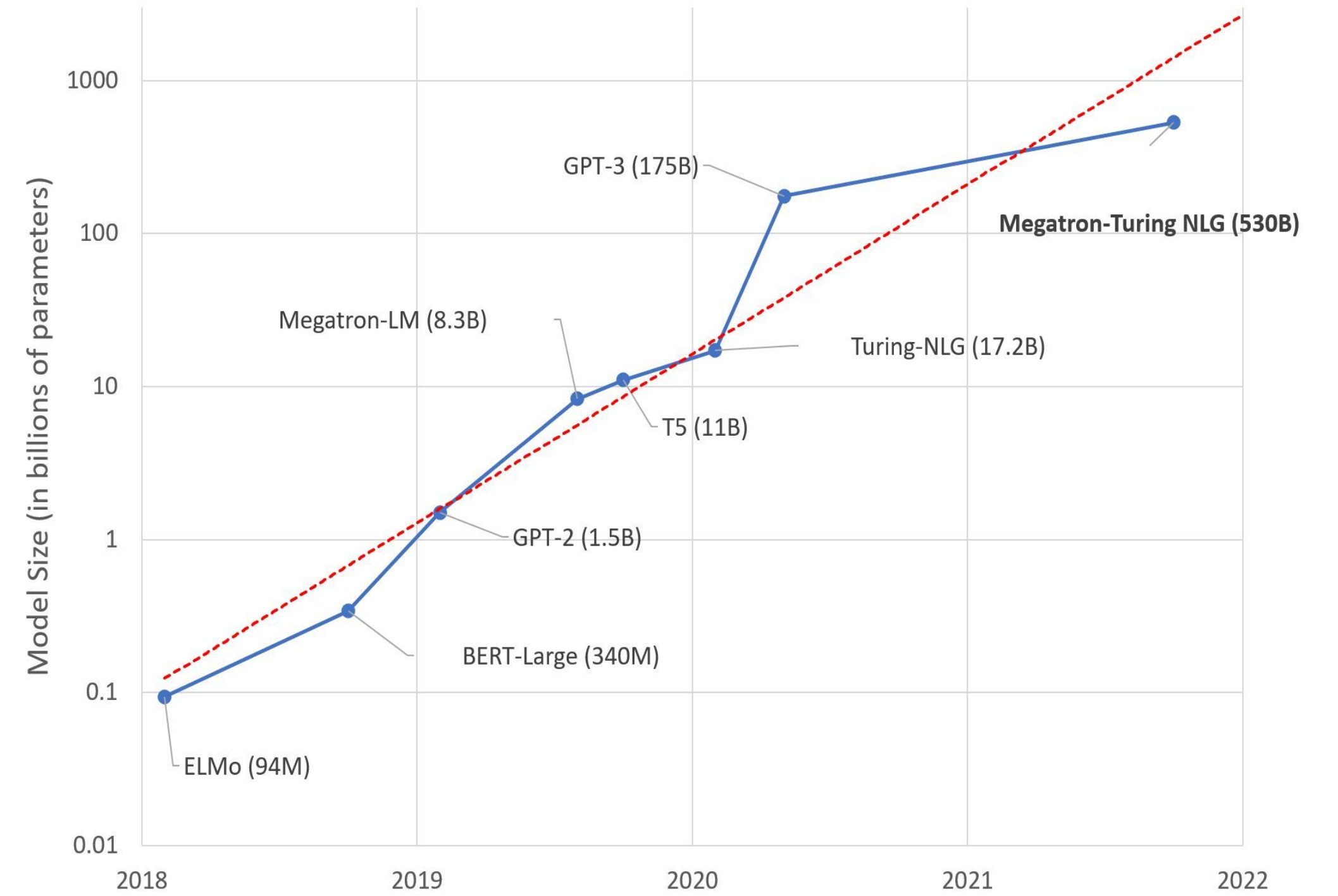
“Everyone wants to do the model work, not the data work” : Data Cascades in High-Stakes AI

<https://storage.googleapis.com/pub-tools-public-publication-data/pdf/0d556e45afc54afeb2eb6b51a9bc1827b9961ff4.pdf>

Evolution of AI



https://www.researchgate.net/publication/349864030_Review_of_machine_learning_and_deep_learning_application_in_mine_microseismic_event_classification



<https://huggingface.co/blog/large-language-models>

Visit - <https://ayon-roy.netlify.app>

Understanding growth in AI's support system

AI is driven by 3 primary factors - Data, Model (Algorithms) & Compute

While model development have taken the spotlight for a few years now & have transformed the way AI advancements are happening with significant improvements in efficiency.

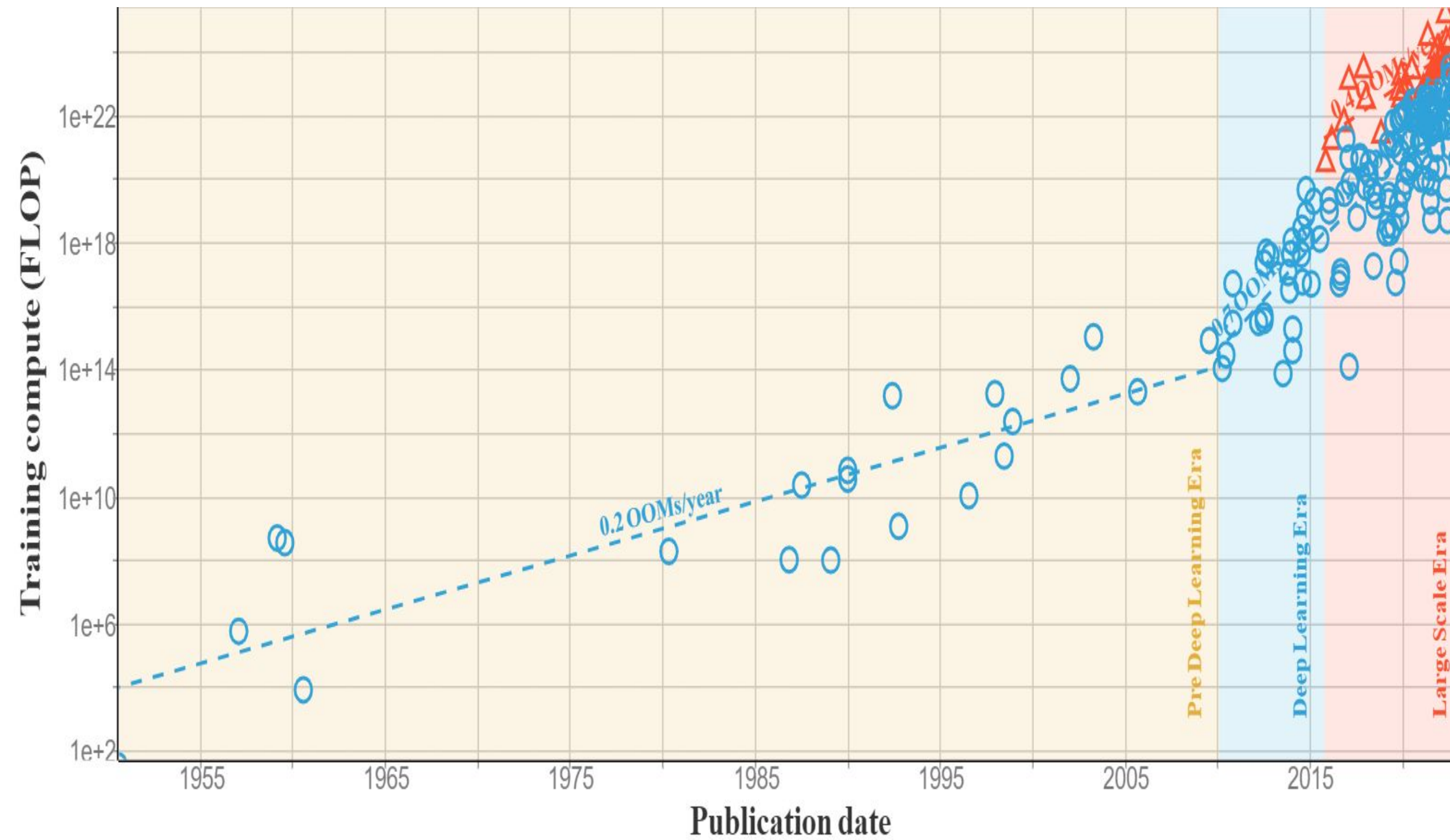
Data has been becoming increasingly available, particularly with the advent of “**big data**” in recent years.

At the same time, progress in computing hardware has been rapid, with increasingly powerful and specialised AI hardware.

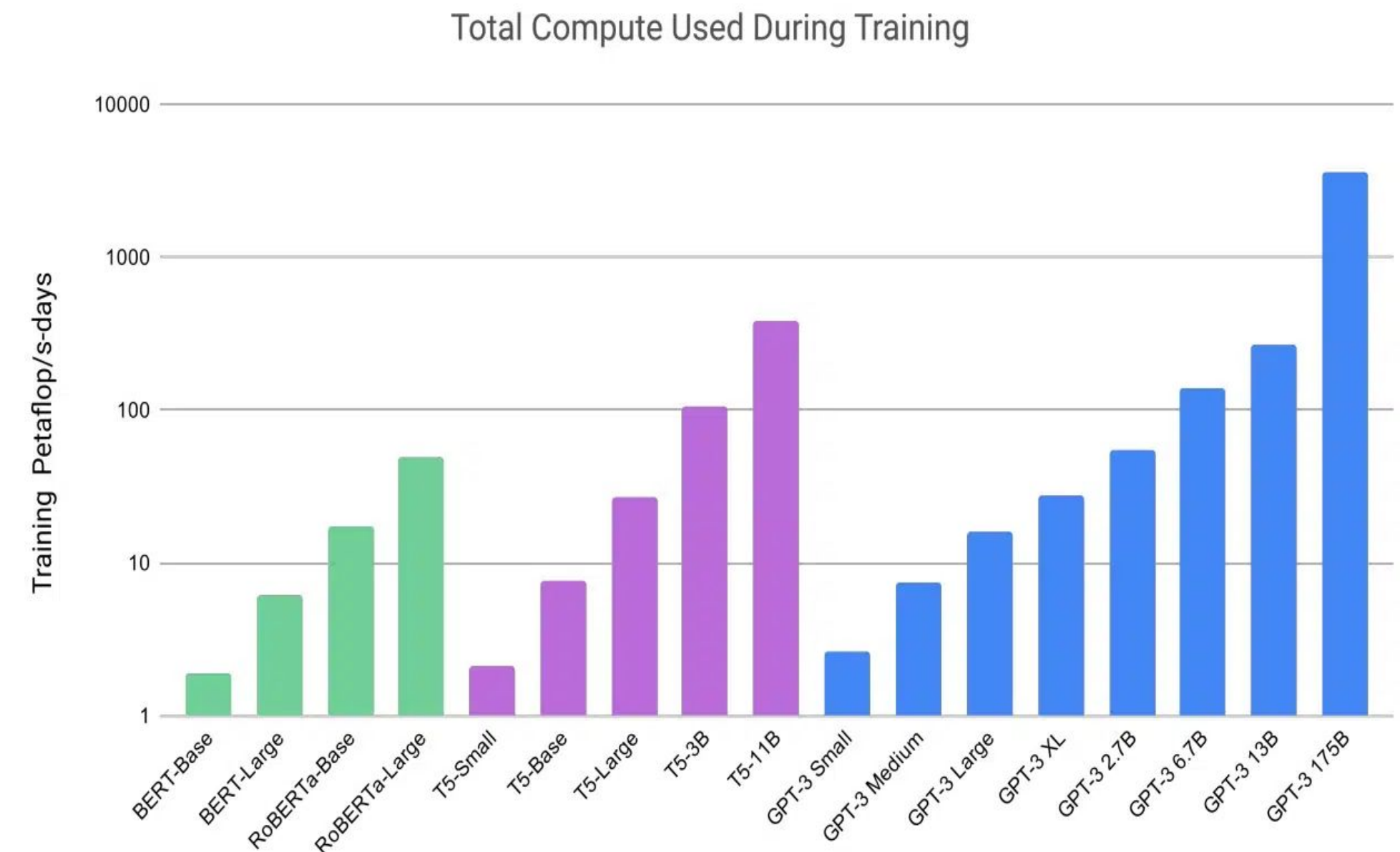
What is less obvious is the *relative* importance of these factors, and what this implies for the future of AI. A paper on “Scaling Laws for Neural Language Models” [Kaplan et al. \(2020\)](#) studied these developments through the lens of **scaling laws**, identifying three key variables:

- Number of parameters of a machine learning model
- Training dataset size
- Compute required for the final training run of a machine learning model (henceforth referred to as **training compute**)

Growth in Computing power requirement



<https://epochai.org/blog/compute-trends>



<https://blogs.nvidia.com/blog/2022/10/10/llms-ai-horizon/>

Growth in Data usage

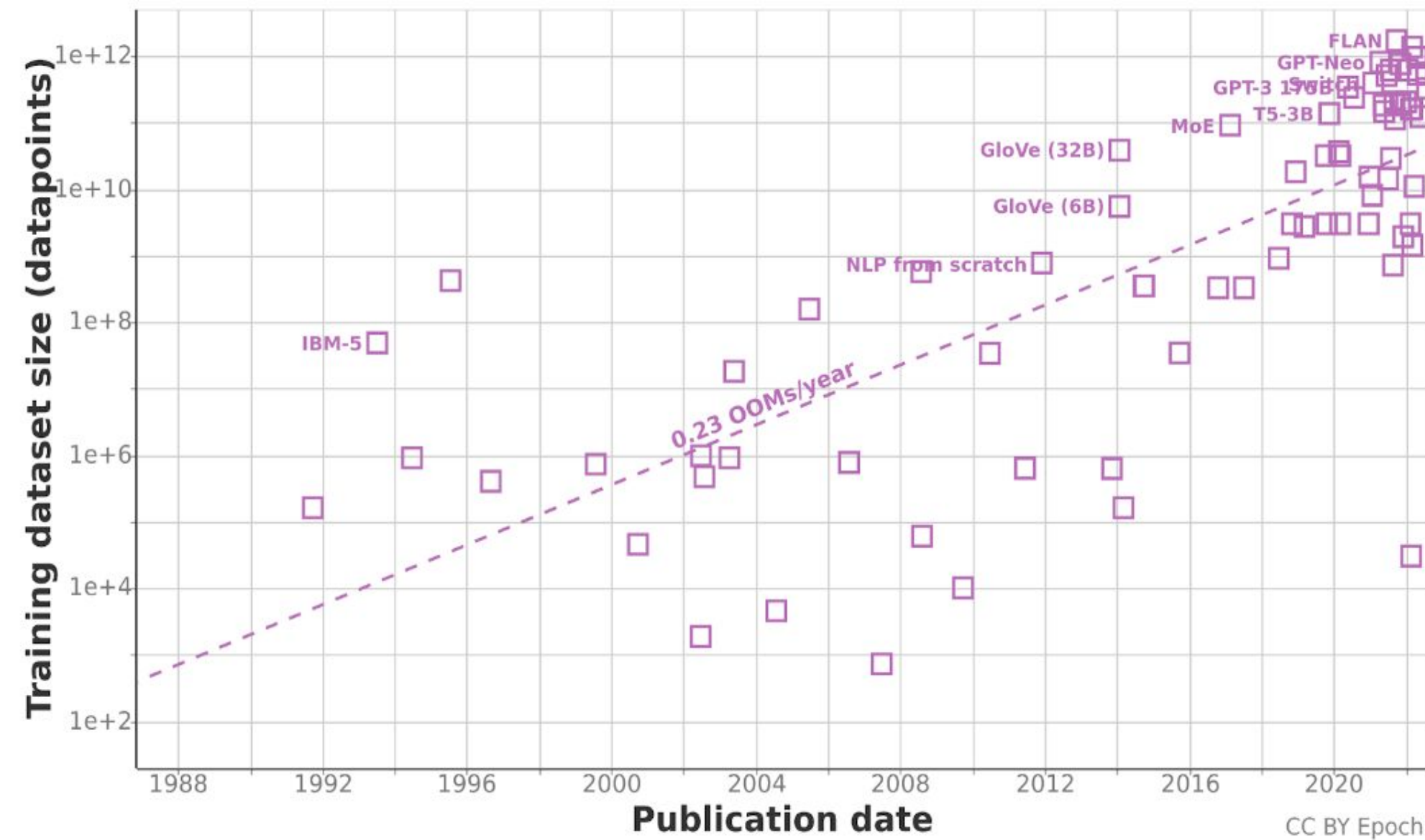


Figure 3: Evolution of language datasets

CC BY Epoch

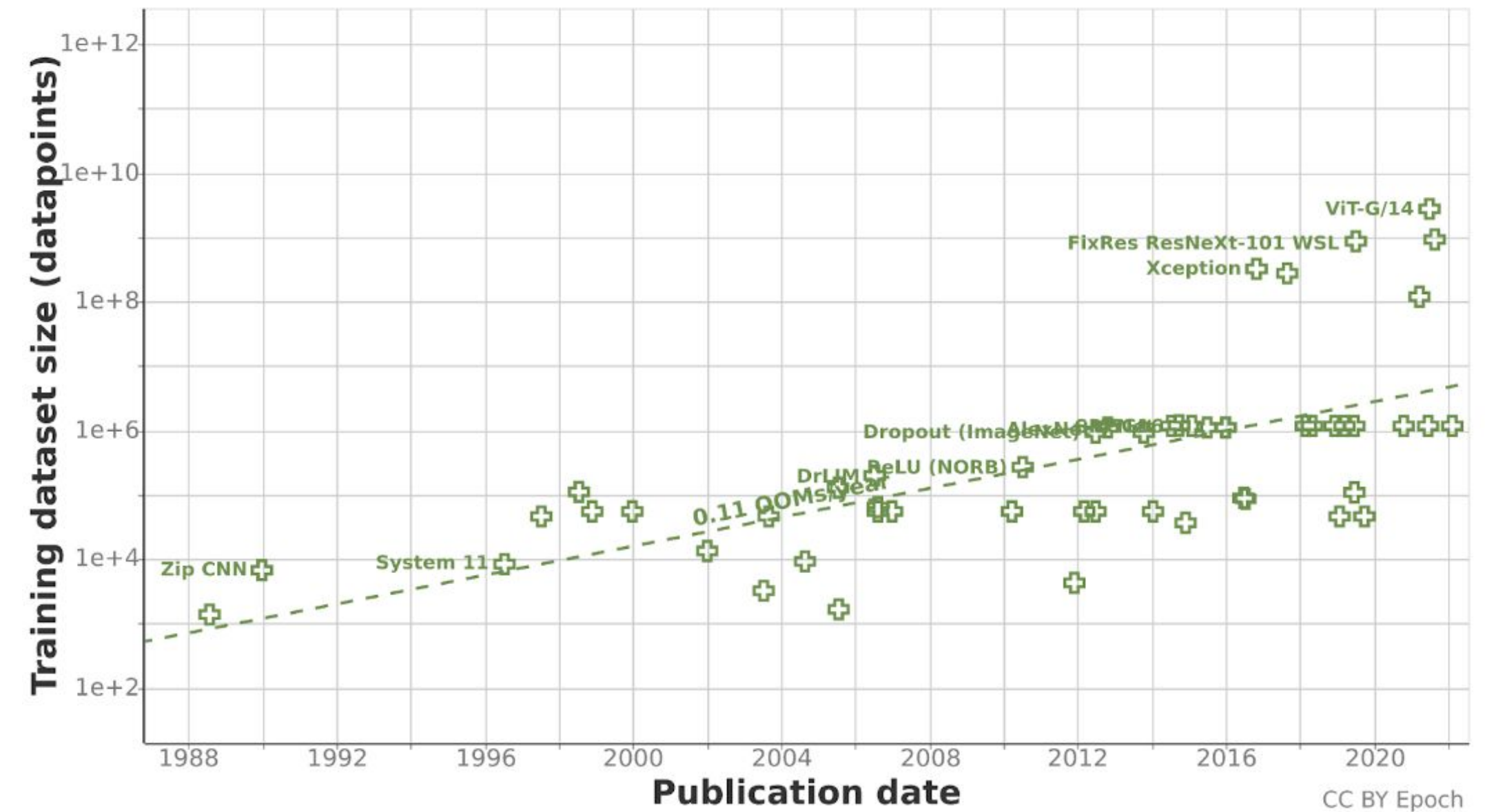
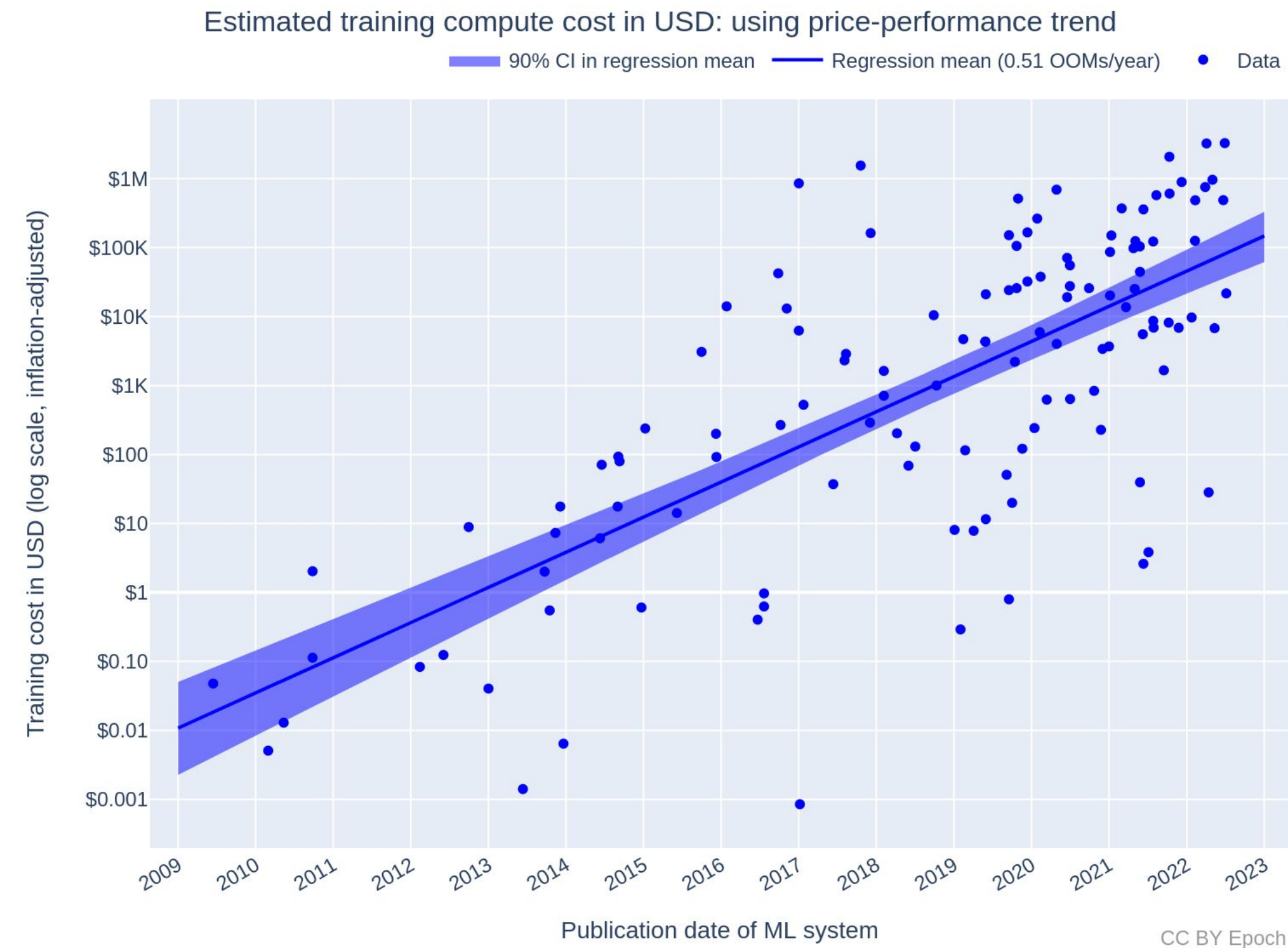


Figure 2: Evolution of vision datasets. A significant number of models is concentrated near $6e4$ and $1e6$, which are the sizes of MNIST and ImageNet, respectively.

CC BY Epoch

<https://epochai.org/blog/trends-in-training-dataset-sizes>

Growth in Compute cost (\$)



<https://epochai.org/blog/trends-in-the-dollar-training-cost-of-machine-learning-systems>

Major approaches used while doing AI

Model Centric & Data Centric Approach

Model-Centric Approach

This involves designing empirical tests around the model to improve the performance. This consists of finding the right model architecture and training procedure among a huge space of possibilities.

Data-centric approach

This consists of systematically changing/enhancing the datasets to improve the accuracy of your AI system. This is usually overlooked and data collection is treated as a one off task.

<https://towardsdatascience.com/from-model-centric-to-data-centric-artificial-intelligence-77e423f3f593#:~:text=Data%2Dcentric%20approach,as%20a%20one%20off%20task>

Community's Bias towards Model Centric Approach

The steel sheets defect detection was one of the examples brought during the session — assuming a series of images from steel sheets we want to develop the best model to detect these defects that can happen during the process of steel sheets manufacturing. There are 39 different defects that we want to be able to identify. By developing a computer vision model with well-tuned hyperparameters, it was able to reach a **76.2% accuracy baseline system**, but the goal is to achieve **90% accuracy**. *How can this be done?*

Steel Sheets Detection Challenge

<https://www.youtube.com/watch?v=06-AZXmwHjo&t=148s>

Difference in Results

Knowing that the baseline model was already good, the task to have it improved to achieve 90% accuracy sound almost impossible — for the model-centric, the improvements based on Network Architecture search and using the state-of-the-art architectures, whereas, for the data-driven, the approach taken was to identify inconsistencies and clean noisy labels. The results were mind-blowing:

Steel sheets defects detection	Baseline	Model-centric	Data-centric
<i>Accuracy</i>	76.2%	+0% (76.2%)	+16.9% (93.1%)

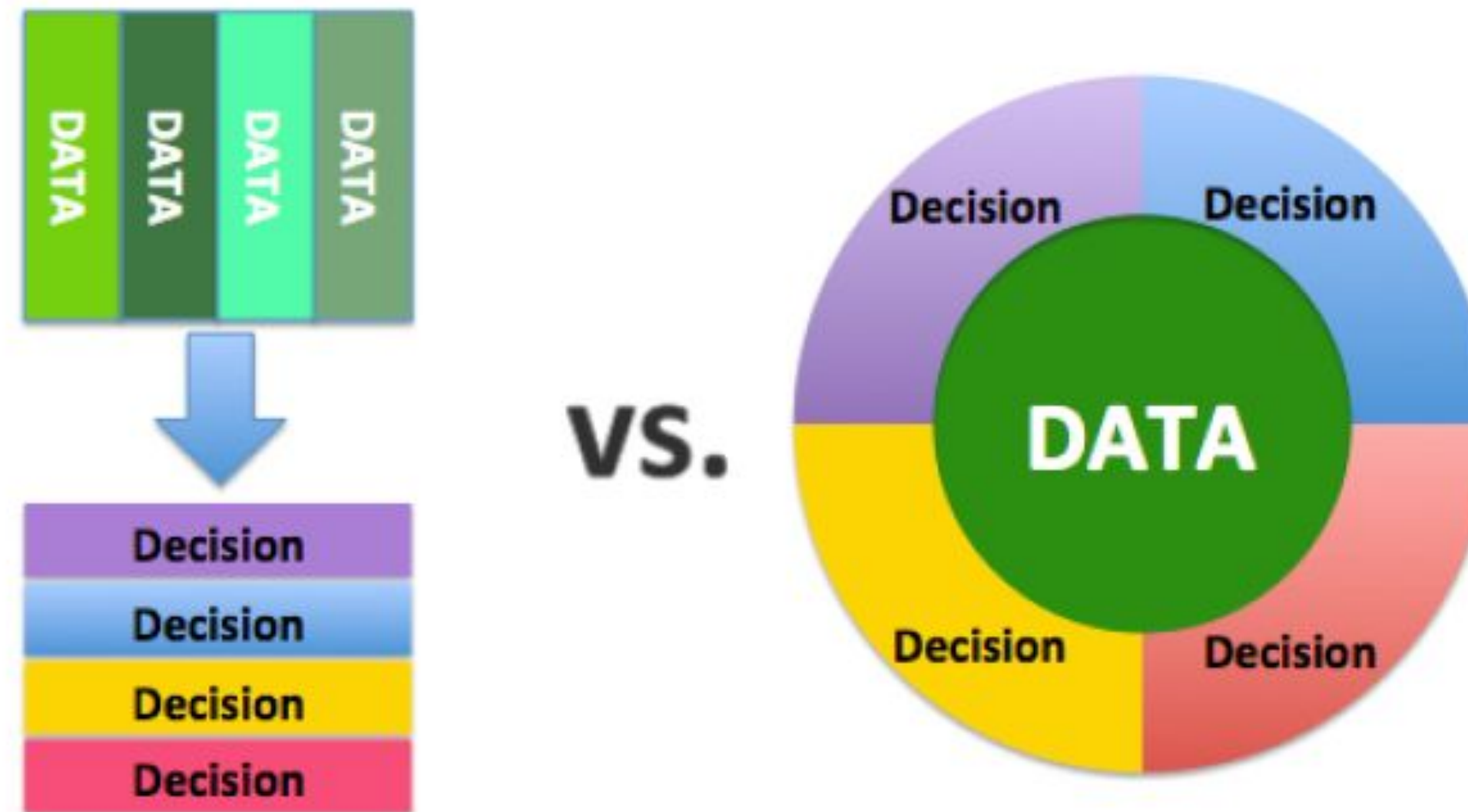
Importance of Data Centric Approaches

	Steel defect detection	Solar panel	Surface inspection
Baseline	76.2%	75.68%	85.05%
Model-centric	+0% (76.2%)	+0.04% (75.72%)	+0.00% (85.05%)
Data-centric	+16.9% (93.1%)	+3.06% (78.74%)	+0.4% (85.45%)

<https://www.youtube.com/watch?v=06-AZXmwHjo&t=324s>

Beware of the Trade Off

Data-Driven vs. Data-Centric









<https://neptune.ai/blog/data-centric-vs-model-centric-machine-learning>

But..

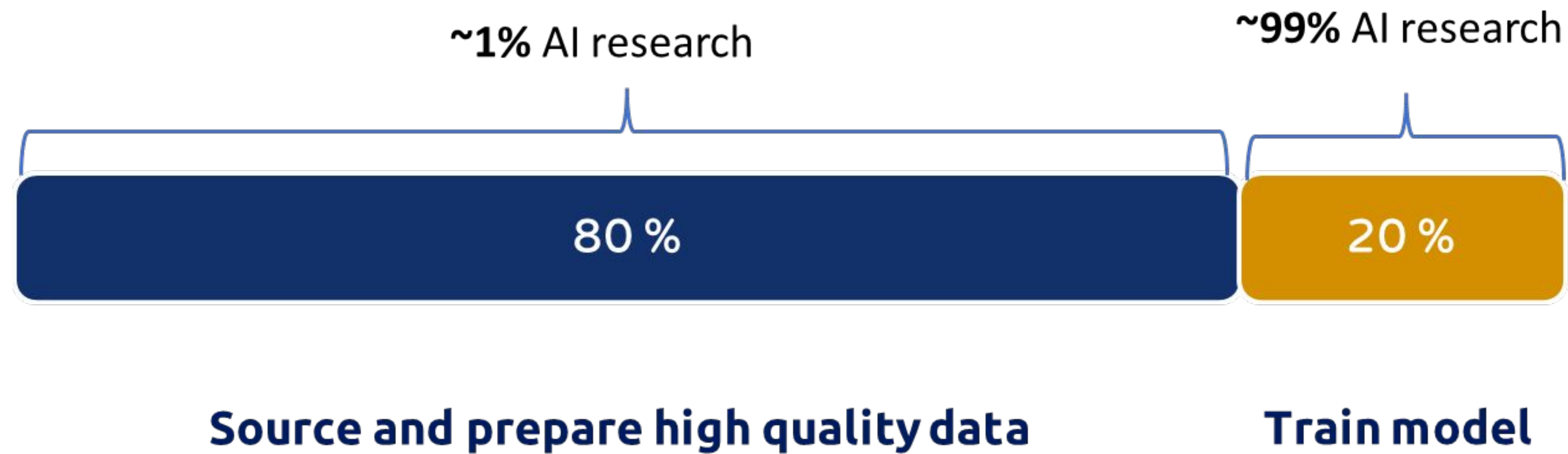
Check for Data Quality

Several factors contribute to the quality of data, including:

 Accuracy	 Completeness	 Relevancy
 Validity	 Timeliness	 Consistency

<https://www.lotame.com/why-is-data-quality-important/>

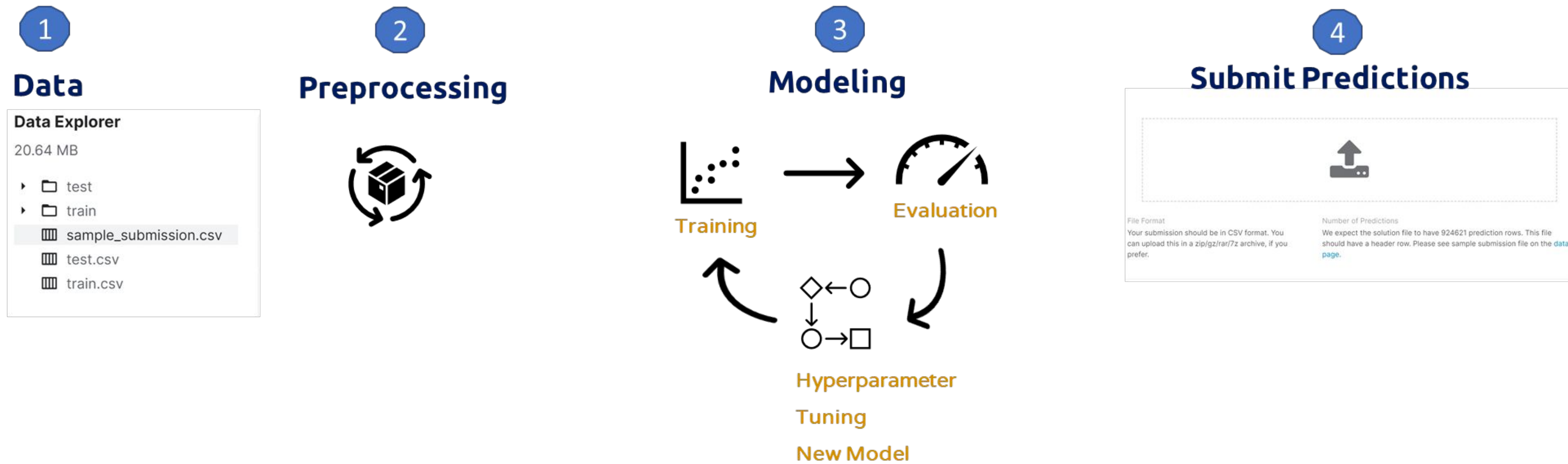
Need for Data Centric Approaches



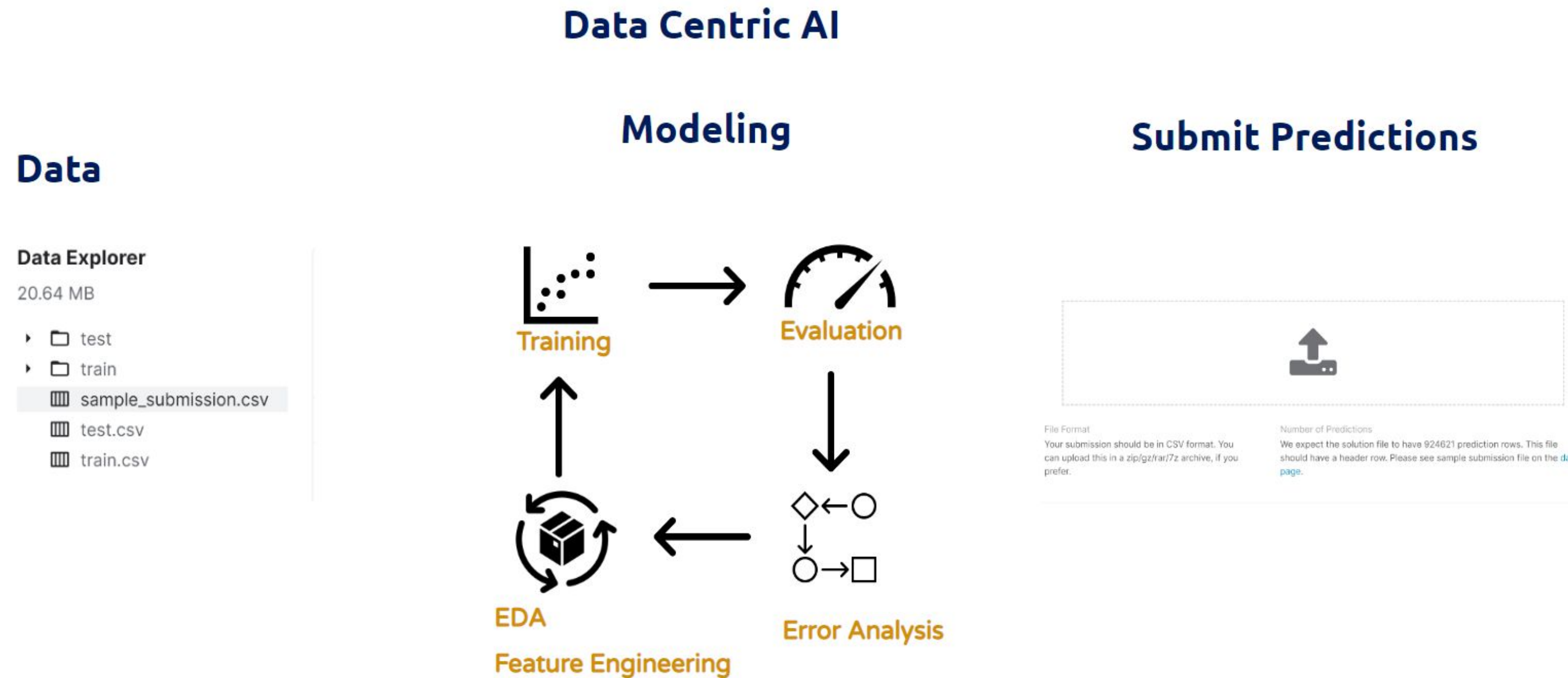
<https://www.youtube.com/watch?v=06-AZXmwHjo&t=1835s>

Model Centric AI Approaches in Kaggle Competitions

Model Centric AI



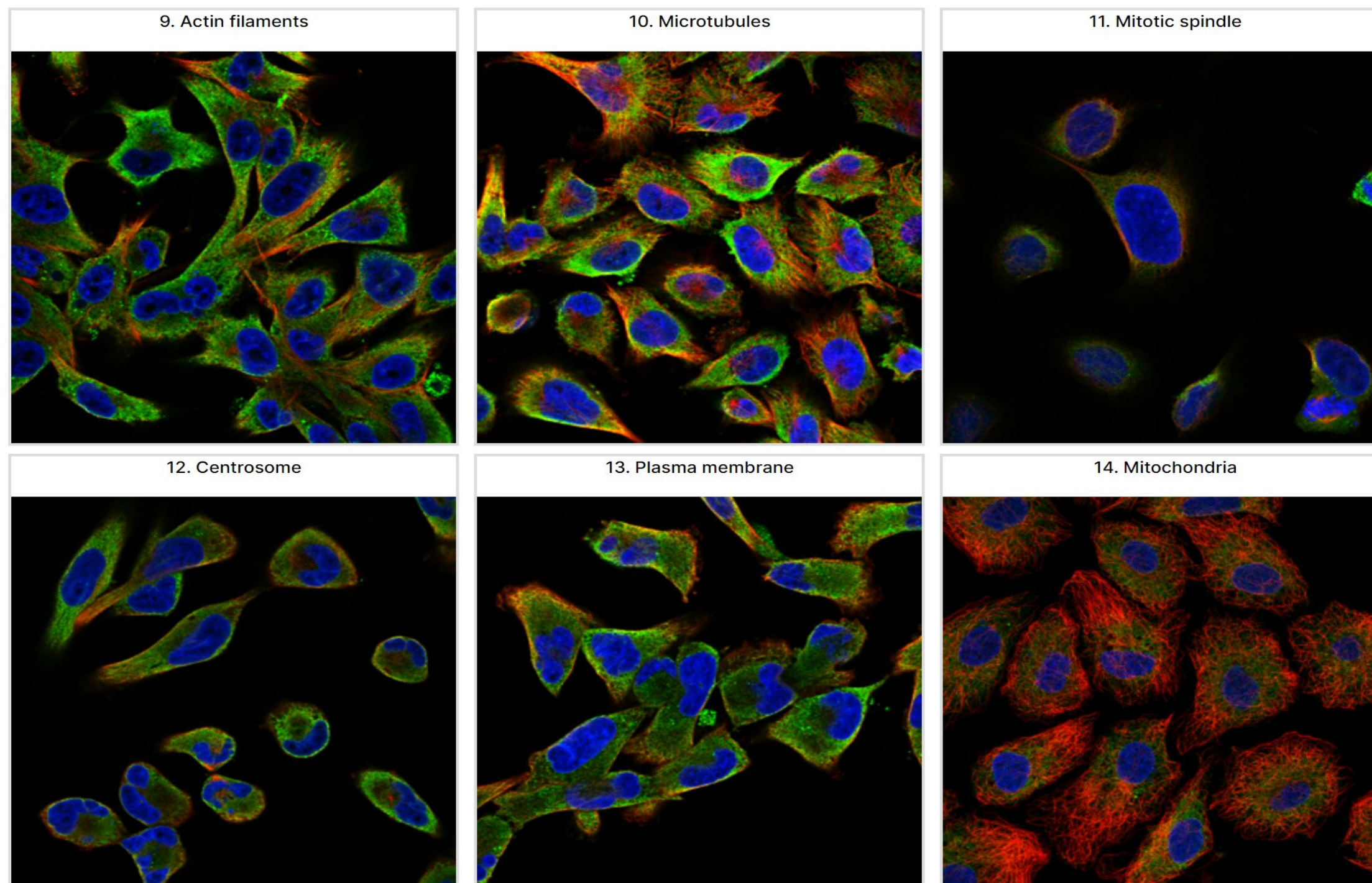
Data Centric AI Approaches in Kaggle Competitions



Human Protein Atlas - Single Cell Classification

[Kaggle Competition]

Data



Task

Segment the cells in the images
and predict the labels of those segmented cells

Challenge

The labels you will get for training are *Image* level
labels while the task is to predict *cell* level labels

Kaggle Competition Solution Approach

Featured Code Competition

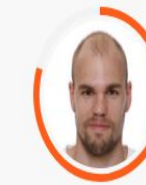
Human Protein Atlas - Single Cell Classification

Find individual human cell differences in microscope images

\$25,000 Prize Money

Human Protein Atlas · 757 teams · a year ago

#	△	Team	Members	Score	Entries	Last	Code
1	▲ 4	bestfitting		0.56670	480	1y	
2	—	[red.ai]		0.55328	459	1y	<>
3	—	MPWARE & ZFTurbo & Dieter		0.54995	500	1y	<>
4	▲ 2	MILIMED		0.54389	258	1y	



CroDoc

Topic Author

4th place

4th Place Solution: MILIMED

Posted in [hpa-single-cell-image-classification](#) a year ago

15

We are a very diverse team of computer scientists and medical doctor/students. It was our great pleasure to participate in this demanding challenge. Hope some of you find this solution useful and/or interesting.

Solution overview

1. Segmentation → HPA-Cell-Segmentation
2. Dataset → 512×512 cell images (20% removed)
3. Parallelization → speed-up → 3h left for inference
4. Manual Labeling → smaller classes & validation (soft labels)
5. Pseudo-Labeling → negative labeling (& positive for mitotic spindle)
6. EfficientNetB0 Ensemble + semi-balanced data sampling
7. Fine-tuning → on manually labeled & non-labeled validation data
8. Cell/Image Weighting → final confidence = $0.7 * \text{cell_confidence} + 0.3 * \text{image_confidence}$

Use of Data effectively for AI, ML

Our projections predict that we will have exhausted the stock of low-quality language data by 2030 to 2050, high-quality language data before 2026, and vision data by 2030 to 2060. This might slow down ML progress.

All of our conclusions rely on the unrealistic assumptions that current trends in ML data usage and production will continue and that there will be no major innovations in data efficiency.

Relaxing these and other assumptions would be promising future work.

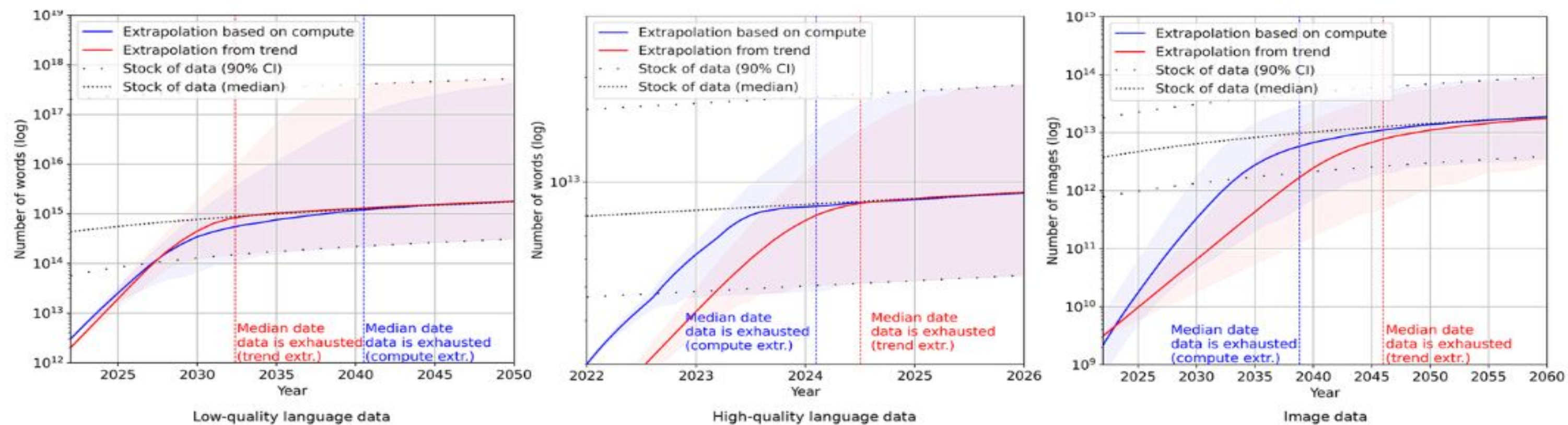
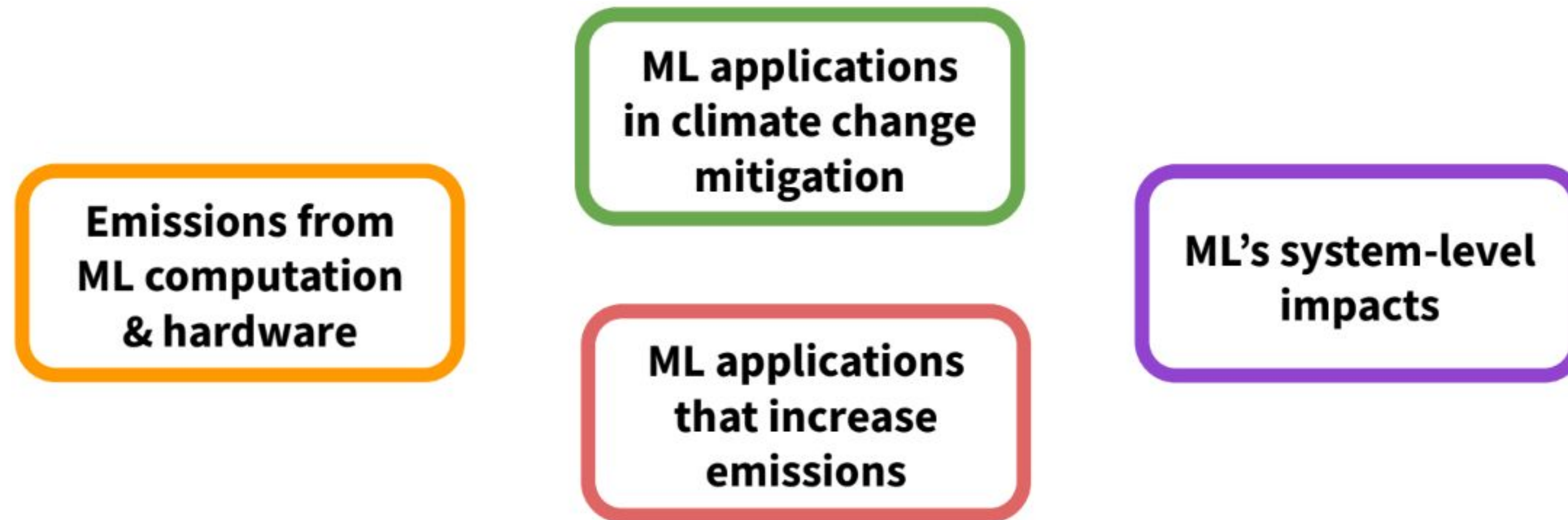


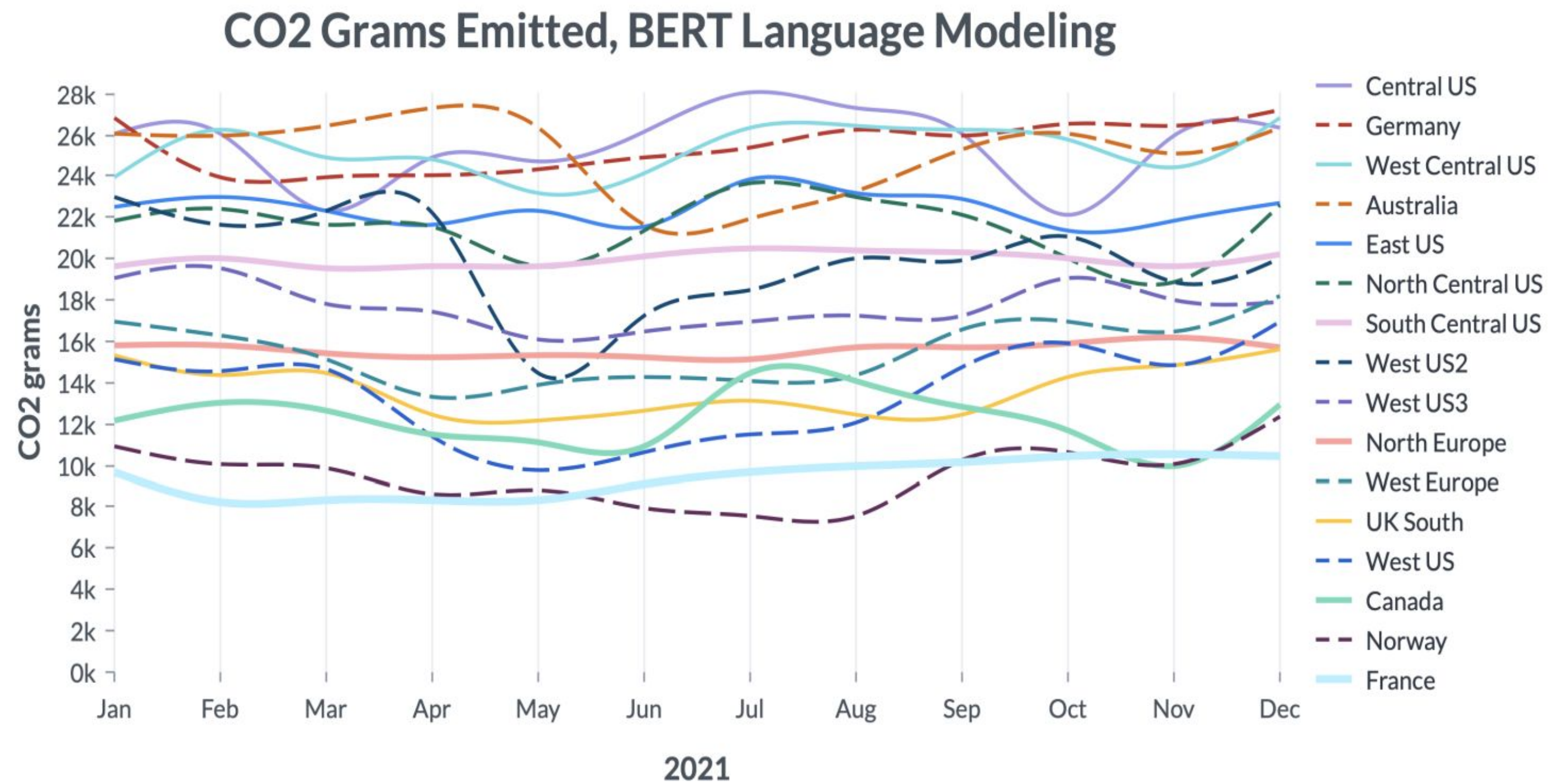
Figure 1: ML data consumption and data production trends for low quality text, high quality text and images.

<https://epochai.org/blog/will-we-run-out-of-ml-data-evidence-from-projecting-dataset>

AI-ML Carbon Footprint



Emissions from AI Cloud Instances

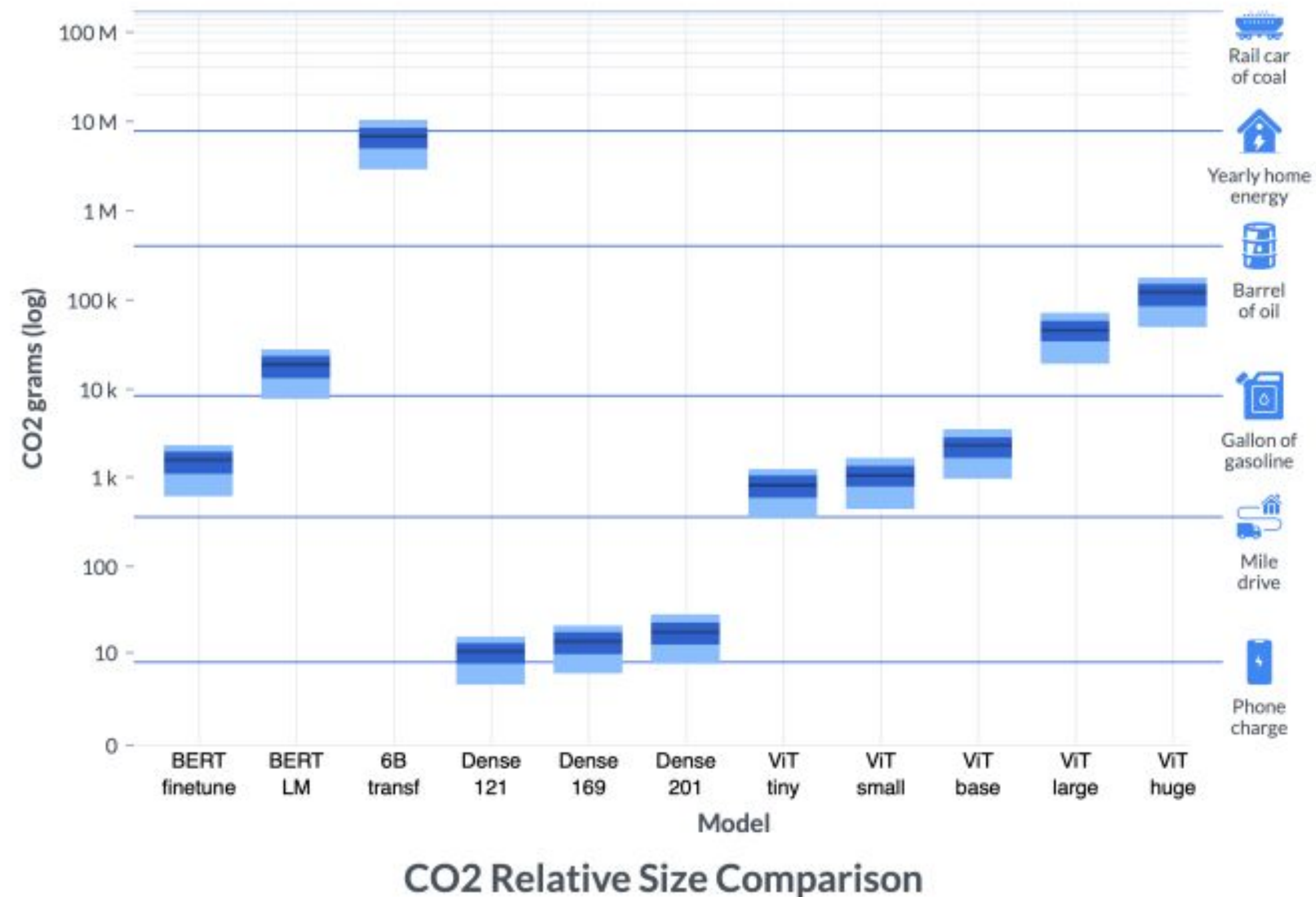


Measuring the Carbon Intensity of AI in Cloud Instances : <https://arxiv.org/pdf/2206.05229.pdf>

Emissions from AI Cloud Instances

Measuring the Carbon Intensity of AI in Cloud Instances

FACCT '22, June 21–24, 2022, Seoul, Republic of Korea



Emissions for the research paper's 11 experiments. For each model they show a vertical blue bar, where the top of the bar is the max, the bottom is the min, and the black line represents the average emissions (across regions and time of year).

First and fourth quartiles are represented by the light blue at the top and bottom of each vertical blue bar. The largest training runs (e.g., 6 billion parameter LM) releases a significant amount of emissions, no matter the region (and recall the 6 billion parameter LM is only trained for 13% of a full run, so a full run would emit about an order of magnitude more emissions than reported here).

The smallest experiments emit very little. Presented on a log scale, with references on the right indicating equivalent sources of emissions per the United States Environmental Protection Agency.

Measuring the Carbon Intensity of AI in Cloud Instances : <https://arxiv.org/pdf/2206.05229.pdf>

Necessity for exploring Sust**AI**nability

Approaching Sustainability as you build AI Systems

One consequence of this increase in computing is the heavy environmental impact of training machine learning models. A recent research paper — [Energy and Policy Considerations for Deep Learning in NLP](#) — notes that an inefficiently trained NLP model using Neural Architecture Search can emit **more than 626,000 pounds of CO₂**. That's about **five times the lifetime emissions of an average American car!**

<https://wandb.ai/amanarora/codecarbon/reports/Tracking-CO2-Emissions-of-Your-Deep-Learning-Models-with-CodeCarbon-and-Weights-Biases--VmlldzoxMzM1NDg3>

Comparison of Certain NLP Models

Model	Hardware	Power (W)	Hours	kWh·PUE	CO ₂ e	Cloud compute cost
Transformer _{base}	P100x8	1415.78	12	27	26	\$41–\$140
Transformer _{big}	P100x8	1515.43	84	201	192	\$289–\$981
ELMo	P100x3	517.66	336	275	262	\$433–\$1472
BERT _{base}	V100x64	12,041.51	79	1507	1438	\$3751–\$12,571
BERT _{base}	TPUv2x16	—	96	—	—	\$2074–\$6912
NAS	P100x8	1515.43	274,120	656,347	626,155	\$942,973–\$3,201,722
NAS	TPUv2x1	—	32,623	—	—	\$44,055–\$146,848
GPT-2	TPUv3x32	—	168	—	—	\$12,902–\$43,008

Table 3: Estimated cost of training a model in terms of CO₂ emissions (lbs) and cloud compute cost (USD).⁷ Power and carbon footprint are omitted for TPUs due to lack of public information on power draw for this hardware.

Relevant Research Paper : <https://arxiv.org/pdf/1906.02243.pdf>

SustAInable Development

UN Brundtland Commission in its report "Our Common Future", published in 1987 defines sustainable development as "**development that meets the needs of the present without compromising the ability of future generations to meet their own needs.**"

It contains within it two key concepts:

- the concept of 'needs', in particular the essential needs of the world's poor, to which overriding priority should be given;
- the idea of limitations imposed by the state of technology and social organization on the environment's ability to meet present and future needs.

In order to achieve this **sustainable development**, objectives are necessary. This is why the United Nations General Assembly defined in 2015 the 17 Sustainable Development Goals (SDGs), adopted by all member countries, to be achieved by 2030.



<https://www.un.org/sustainabledevelopment/blog/2015/12/sustainable-development-goals-kick-off-with-start-of-new-year/>

Sust**AI**nable Future , is it ?

AI has the potential to assist in the achievement of each of the SDGs. This is illustrated by a study by the McKinsey Global Institute, which by November 2018 had identified **as many as 135 cases of AI use worldwide that support the SDGs**. [Whether concrete, partial or simply potential, these cases demonstrate real opportunities for AI applications]

The proposed hypothesis is that

By becoming aware of the potential precariousness of the resources, room for maneuver and other properties intrinsic to the nature of the actor, the business and industry sector, nongovernmental organizations, the scientific and technological community, and local authorities, i.e. the State, are the actors best placed not only to impact generally on the progress towards the SDGs, but also to be the channels through which AI can most ideally, until then, impact beneficially on Sustainable Development and reach the SDGs.

- **In your own words, what solution did you actually build with ERP Air Force?**

“A full multi-sensor IoT monitoring ecosystem boosted by AI for reserve protection and anti-poaching measures.”

- **Where and how was Artificial Intelligence involved, in more details?**

“The pieces where we’ve got AI involved are in the analytics of our camera-traps’ photos, and in our license plates recognition, also leveraging OCR (Optical Character Recognition). And then, the biggest use of AI is in our photographs’ stalls analytics [...] to look for humans.”
We use a combination of OpenCV and Darknet YOLO.”

- **How do you think your solution has a social impact and helps within our societies’ Sustainable Development (Goals)?**

Goal 1 (No poverty); Goal 4 (Quality education); Goal 5 (Gender equality); Goal 8 (Decent work and economic growth); Goal 10 (Reduced inequalities); Goal 11 (Sustainable cities and communities); Goal 15 (Life on land); Goal 17 (Partnerships for the goals)

Economic participation opportunity: “If we don’t do something with ERP Air Force now, to conserve elephants and rhinos, there won’t be economic participation opportunities for these communities in ten years’ time. Nobody will want to go on safari if there’s no animals. Bottom line. People go on safari, they believe in ecotourism, to see wonderful trees and forests and animals. We’re certainly protecting the integrity of these opportunities for the next wave of people in probably ten years’ time.”

Quality education: “We kicked off a lot of programs, with especially schoolkids, on the back of the big ERP Air Force story. Science, technology, engineering, and math are a real problem here in terms of education in Africa. We invite students to see what we do, we show them how we fly the drones and we do demonstrations at high schools. And we’ve actually started sponsoring through one of our other ERP initiatives, “We Code” challenge, where there are smaller drones and we teach kids how to do basic coding and fly these drones. “

Cultural heritage conservation: “Those animals (elephants) have very significant religious value to different cultures in Africa. They have cultural value, they are part of people’s folklore, they are part of people’s learning, their songs, their dances. So if those elephants are gone, that is a huge part of Africa that’s missing. “

Use Case

A use case for Sustainable future

What do farmers want to know?

- ★ When to plant?
- ★ Crop performance
- ★ Potential threats to production (e.g. climate change)
- ★ Actual threats to production (e.g., nearby pest/disease outbreak or weather forecasts)
- ★ Soil moisture, rainfall, temperature, etc.
- ★ Productivity potential (yield gap)
- ★ Suitability of crops (would a different crop or variety grow better?)

What do policymakers want to know?

- ★ Crop performance
- ★ Potential threats to production
- ★ Actual threats to production
- ★ When to intervene
- ★ How to intervene
- ★ Productivity potential
- ★ Suitability of crops
- ★ How suitability will change
- ★ Measure impacts of policies

2nd use case for SustAInable future

Satellite-Enabled Food Security Dashboard

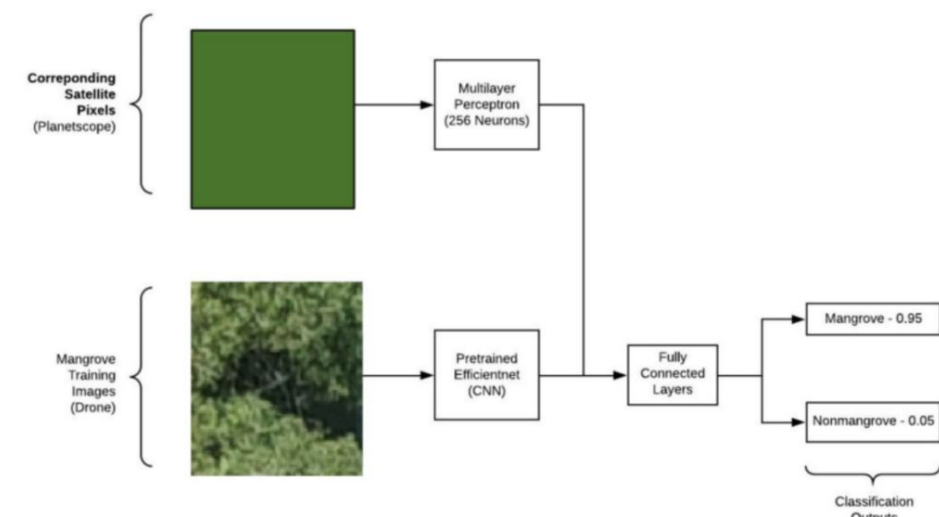
Goal : Create baseline geospatial datasets for measuring and monitoring agricultural production to support policy & efforts to improve food security

How?

- Train machine learning models to predict where crops growing based on Earth observation/satellite data
- Integrate crop maps with other relevant datasets (e.g., socioeconomic and price data)
- Make data available in a public Food Security Dashboard
- Collaborate with stakeholders to ensure products & Dashboard serve community decisions and actions
 - End users: Farmers, Dept of Ag, county council, community organizations, etc.
 - Develop of policies and practices that result in more equitable access to food for the residents
 - Boost agricultural production for food crops, including native crops

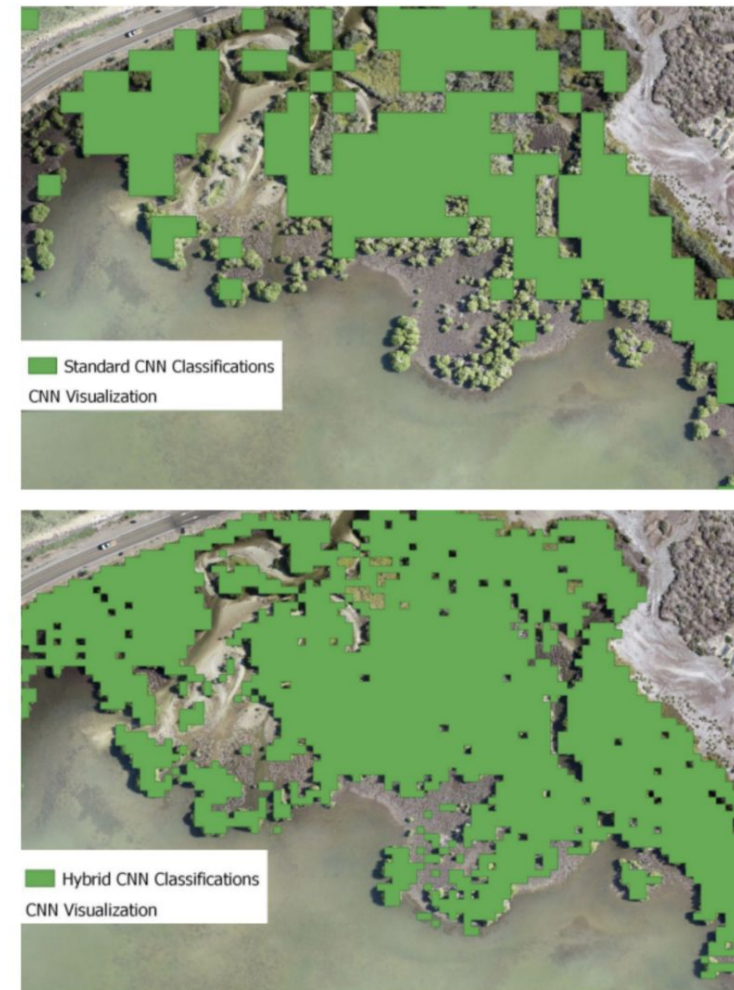
A few other use-cases

Mangrove classification

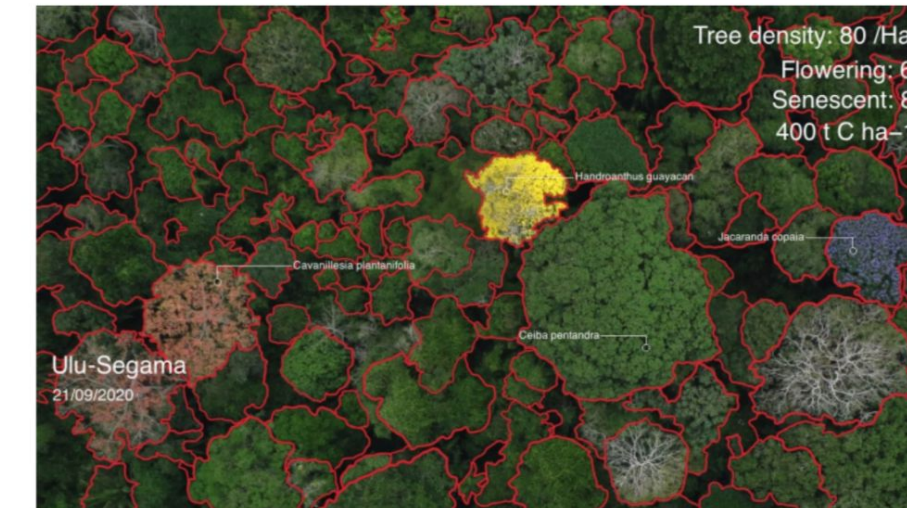


Fusing different data resolutions improves classification accuracy
Challenges: Labeled Data

Mangrove Ecosystem Detection using Mixed-Resolution Imagery with a Hybrid-Convolutional Neural Network
 Hicks et al., CCAI ICML20



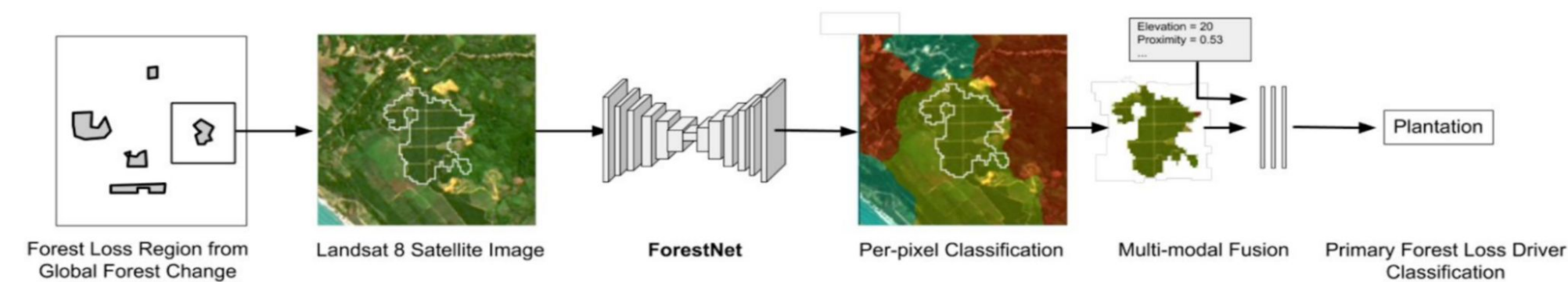
Drone-Based Biomass Estimation



Challenges:
 Limited labeled data for model training



Deforestation driver classification

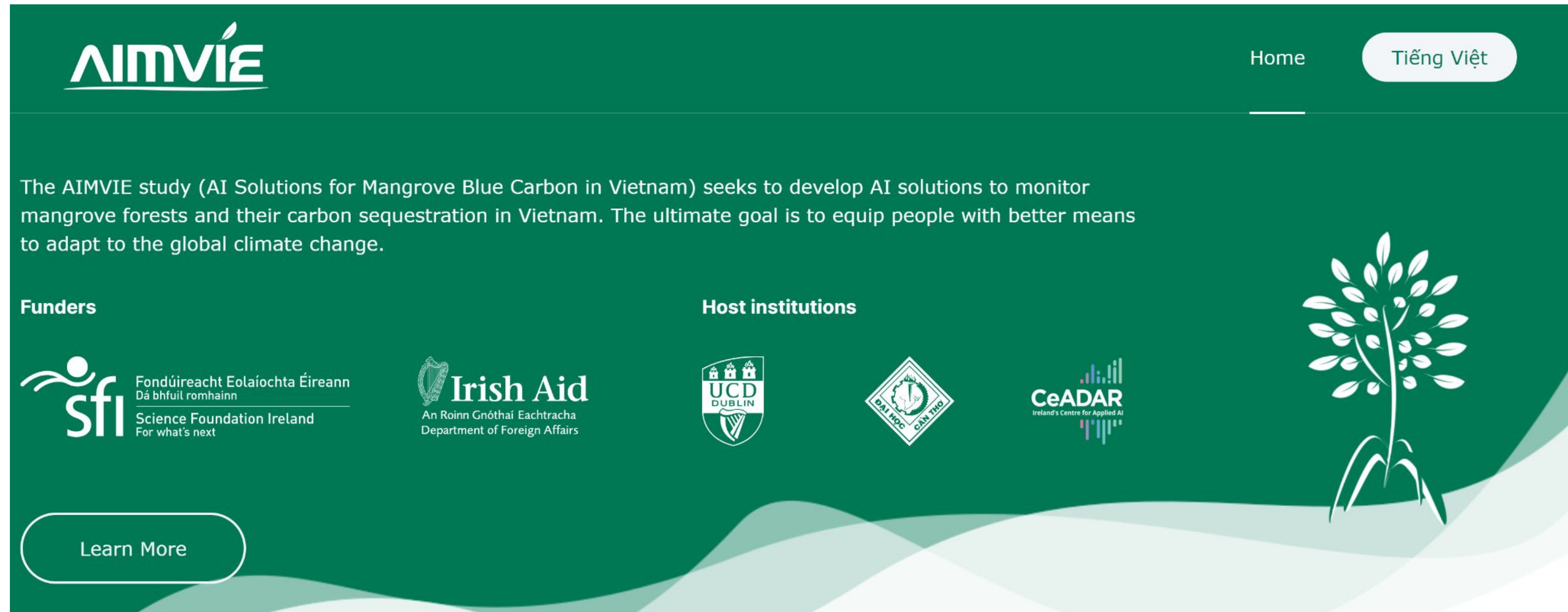


Fusing multi-modal data sources improve classification accuracy
Challenges: Labeled Data

ForestNet: Classifying Drivers of Deforestation in Indonesia using Deep Learning on Satellite Imagery
 Irvin *, Sheng * et al. CCAI NeurIPS20

Model	Predictors	Val		Test	
		Acc	F1	Acc	F1
RF	Visible	0.56	0.49	0.49	0.44
RF	Visible + Aux	0.72	0.67	0.67	0.62
CNN	Visible	0.80	0.75	0.78	0.70
CNN + SDA	Visible	0.82	0.79	0.78	0.73
CNN + SDA + PT	Visible	0.83	0.80	0.80	0.74
CNN + SDA + PT	Visible + Aux	0.84	0.81	0.80	0.75

AI for Mangrove Blue Carbon in Vietnam (AIMVIE)



The banner features a dark green background with a white mangrove tree illustration on the right. The AIMVIE logo is in the top left, and navigation links for 'Home' and 'Tiếng Việt' are in the top right. A central text block describes the study's goal. Below this, two columns list funders and host institutions with their respective logos. A 'Learn More' button is located in the bottom left corner.

AIMVIE

Home Tiếng Việt

The AIMVIE study (AI Solutions for Mangrove Blue Carbon in Vietnam) seeks to develop AI solutions to monitor mangrove forests and their carbon sequestration in Vietnam. The ultimate goal is to equip people with better means to adapt to the global climate change.

Funders

Host institutions

sfi Fondúireacht Eolaíochta Éireann
Dá bhfuil romhainn
Science Foundation Ireland
For what's next

Irish Aid
An Roinn Gnóthaí Eachtracha
Department of Foreign Affairs

UCD DUBLIN

ĐẠI HỌC ĐÀ LẠT

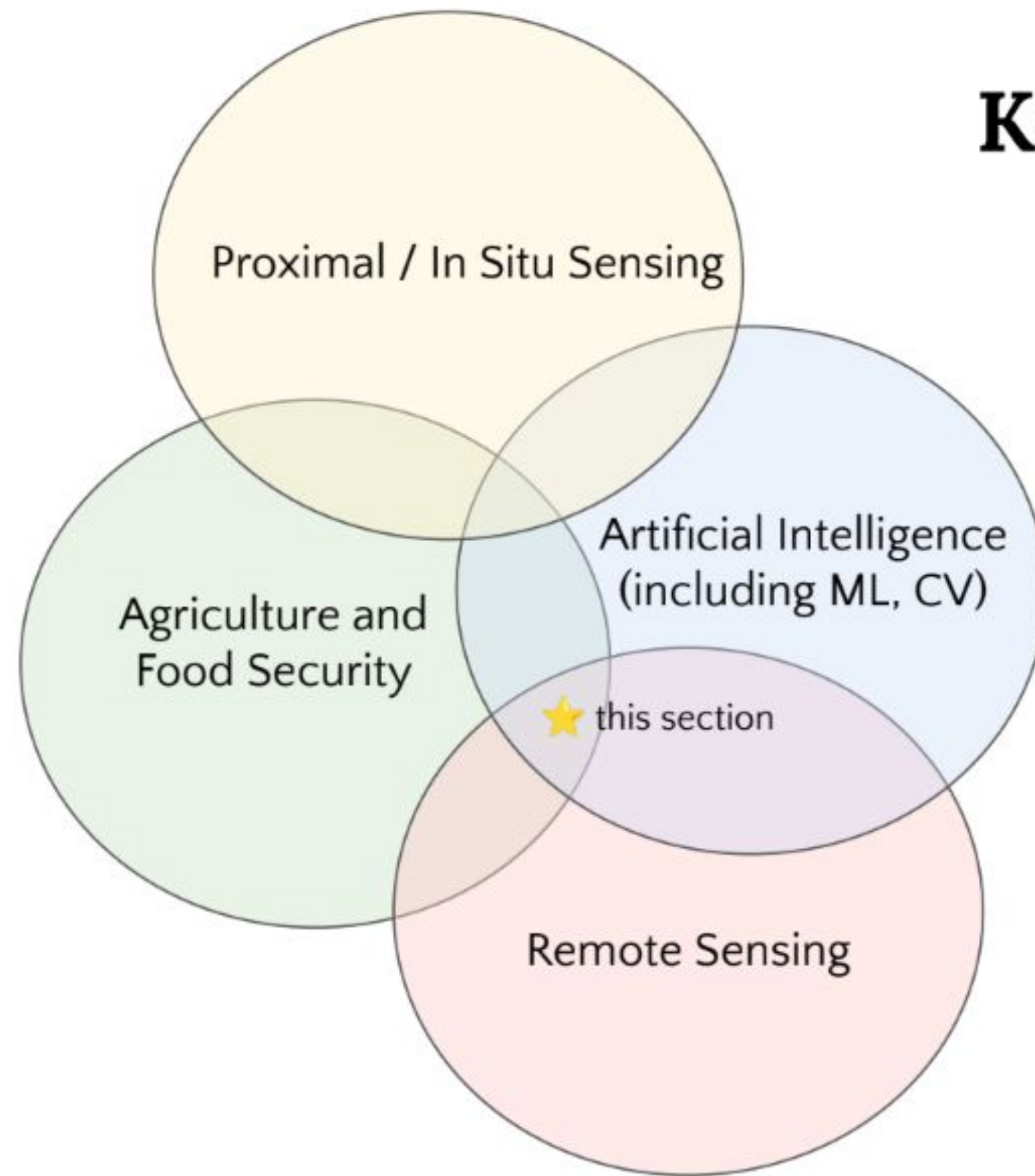
CeADAR
Ireland's Centre for Applied AI

[Learn More](#)

<https://aimvie.org/>

Visit - <https://ayon-roy.netlify.app>

How AI concepts can help ?



Key topics

Crop mapping → Binary classification

Crop type mapping → Multi-class classification

Field boundary delineation → Segmentation

Yield estimation → Regression

Pest and disease detection → OOD detection

Domain adaptation, distribution shift, multi-fidelity data fusion, learning from limited labeled data, etc.

<https://drive.google.com/file/d/1qXflrzdZ4r1MCmdK2RPtxPp9xkl0F8F/view> : Detailed PPT on how AI can be used for Agriculture

Microsoft's Open Source efforts for Sustainability

Demo

<https://replicate.com/bencevans/megadetector>

Important Links

<https://github.com/microsoft/CameraTraps/blob/main/megadetector.md>

<https://agentmorris.github.io/camera-trap-ml-survey/>

Colab Notebook

https://colab.research.google.com/github/microsoft/CameraTraps/blob/main/detection/megadetector_colab.ipynb

How can AI help ?

- Distilling raw data into actionable information
- Optimizing complex systems
- Improving predictions
- Accelerating scientific discovery
- Approximating time-intensive simulations

How AI can help ?

1. Distilling raw data

Role: Distilling raw data into actionable information

Some relevant ML areas: Computer vision, natural language processing for

- ▶ Gathering data on building footprints/heights [M]
- ▶ Evaluating coastal flood risk [A]
- ▶ Parsing corporate disclosures for climate-relevant info [A]

Examples

- ▶ Mapping deforestation and carbon stock [M]

2. Optimizing complex systems

Role: Improving efficient operation of complex, automated systems

Some relevant ML areas: Optimization, control, reinforcement learning

- ▶ Optimizing rail and multimodal transport [M]
- ▶ Demand response in electrical grids [M]
- ▶ Controlling heating/cooling systems efficiently [M]

3. Improving predictions

Role: Forecasts and time series predictions

Some relevant ML areas: Time series analysis, computer vision, Bayesian methods

- ▶ Forecasting electricity demand [M]
- ▶ Predicting crop yield from remote sensing data [A]
- ▶ “Nowcasting” for solar/wind power [M]

4. Accelerating scientific discovery

Role: Suggesting experiments in order to speed up the design process

Some relevant ML areas: Generative models, active learning, reinforcement learning, graph neural networks

- ▶ Algorithms for controlling fusion reactors [M]
- ▶ Identifying candidate materials for batteries, photovoltaics, and energy-related catalysts [M]

5. Approximating simulations

Role: Accelerating time-intensive, often physics-based, simulations

Some relevant ML areas: Physics-informed ML, computer vision, interpretable ML, causal ML

- ▶ Simulating portions of car aerodynamics [M]
- ▶ Speeding up planning models for electrical grids [M]
- ▶ Superresolution of predictions from climate models [A]

Underwater Data Centers for Sustainable Computing

Microsoft finds underwater datacenters are reliable, practical and use energy sustainably



Written by
John Roach
Published
September 14, 2020

Earlier this summer, marine specialists reeled up a shipping-container-size datacenter coated in algae, barnacles and sea anemones from the seafloor off Scotland's Orkney Islands.

The retrieval launched the final phase of a years-long effort that proved the concept of underwater datacenters is feasible, as well as logistically, environmentally and economically practical.

<https://news.microsoft.com/source/features/sustainability/project-natick-underwater-datacenter/>



<https://www.youtube.com/watch?v=IBeepqQBpU>

AI used for Energy Saving

DeepMind's AI cuts energy costs for cooling buildings

Research firm DeepMind has built an AI to optimise cooling systems in buildings. In tests, it reduced energy usage by around 10 per cent



TECHNOLOGY 20 December 2022

By [Jeremy Hsu](#)

<https://www.newscientist.com/article/2352075-deepminds-ai-cuts-energy-costs-for-cooling-buildings/>

How can you approach
Sust**A**Inability?

Focus on your day to day activities

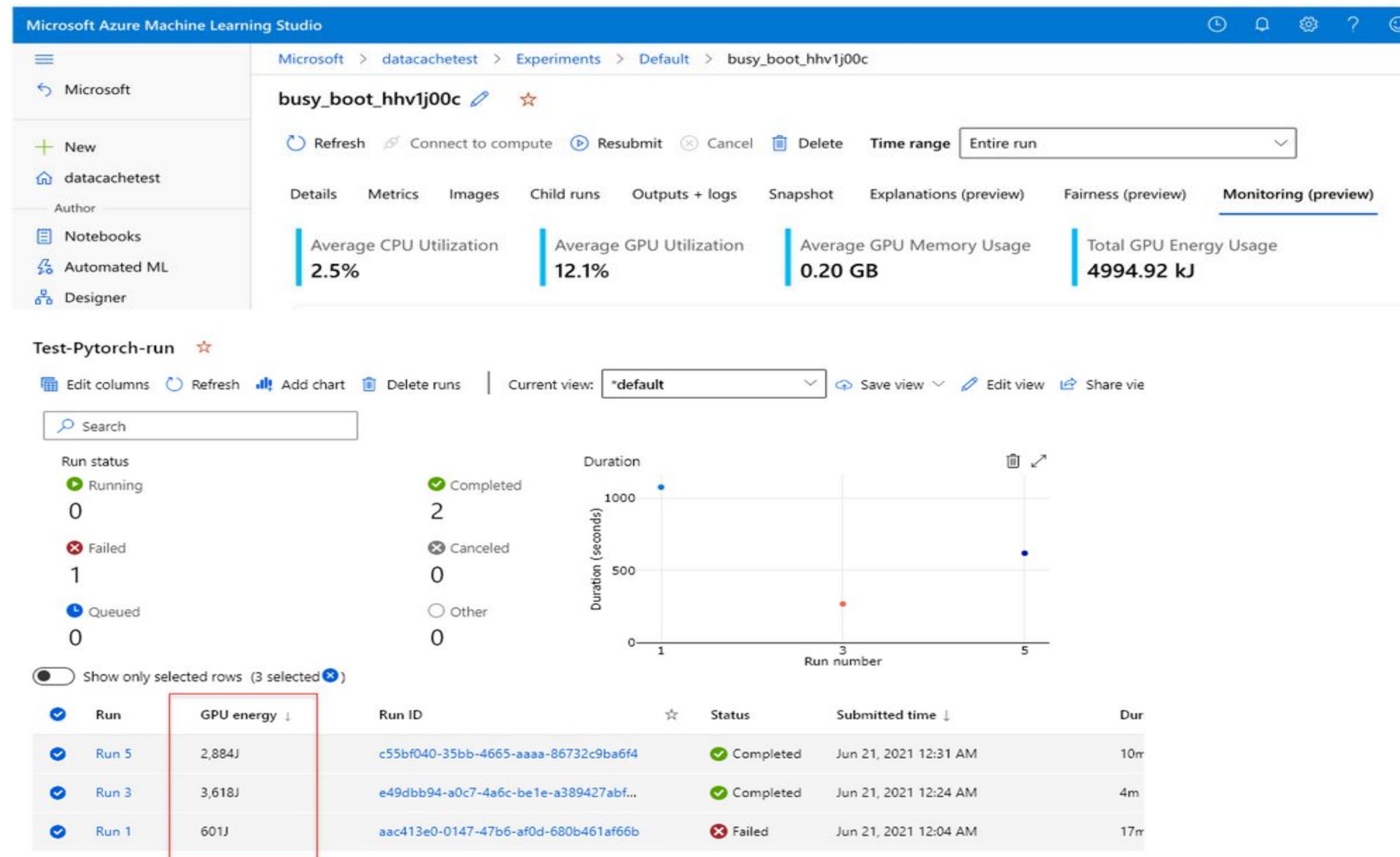
Computing-related

- ▶ Measure your footprint with tools such as ML CO2 Impact, CodeCarbon, Carbontracker, or tools specifically for Azure or Hugging Face
- ▶ Reduce your impacts by choosing more efficient models, and reducing wasteful model retraining & execution

Application-related

- ▶ Quantify and evaluate the application impacts where possible
- ▶ Be transparent about impacts in publications and with stakeholders (quantitatively and qualitatively)
- ▶ Choose what you (or the ML community) works on

Sustainable AI with Azure Machine Learning resource metrics



<https://techcommunity.microsoft.com/t5/green-tech-blog/charting-the-path-towards-sustainable-ai-with-azure-machine/ba-p/2866923>

Visit - <https://ayon-roy.netlify.app>

Tracking CO2 Emissions of Your Deep Learning Models with CodeCarbon + Weights & Biases



ABOUT HOW IT WORKS CALL FOR ACTION TEAM CONTACT US

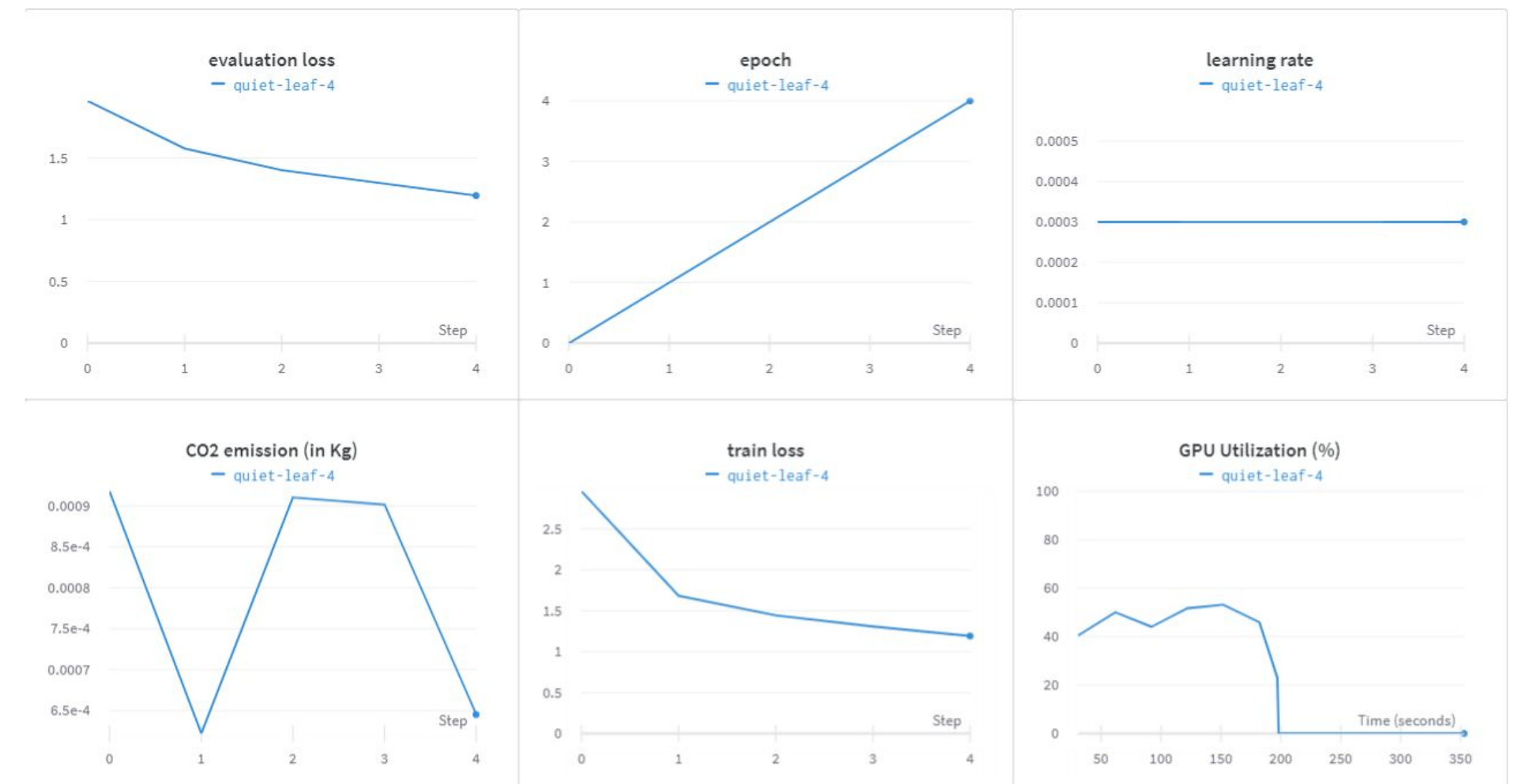
Track and reduce CO2 emissions from your computing

AI can benefit society in many ways but, given the energy needed to support the computing behind AI, these benefits can come at a high environmental price.

CodeCarbon is a lightweight software package that seamlessly integrates into your Python codebase. It estimates the amount of carbon dioxide (CO2) produced by the cloud or personal computing resources used to execute the code.

It then shows developers how they can lessen emissions by optimizing their code or by hosting their cloud infrastructure in geographical regions that use renewable energy sources

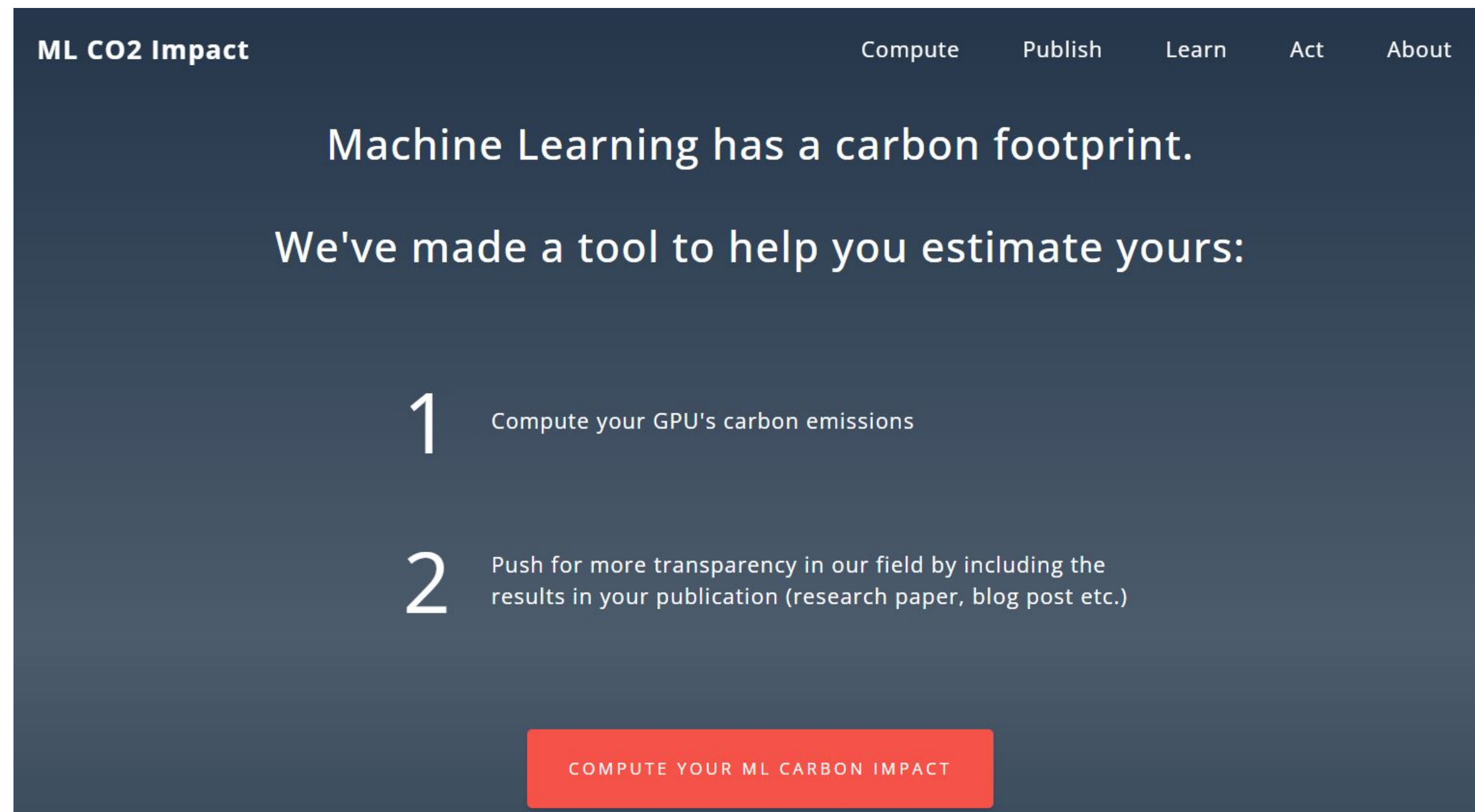
<https://codecarbon.io/>



<https://wandb.ai/amanarora/codecarbon/reports/Tracking-CO2-Emissions-of-Your-Deep-Learning-Models-with-CodeCarbon-and-Weights-Biases--VmlldzoxMzM1NDg3>

Visit - <https://ayon-roy.netlify.app>

A few other interesting ideas to explore



The image shows the landing page for 'ML CO2 Impact'. The header includes the title 'ML CO2 Impact' and navigation links for 'Compute', 'Publish', 'Learn', 'Act', and 'About'. The main text reads: 'Machine Learning has a carbon footprint. We've made a tool to help you estimate yours:'. Below this, there are two numbered steps: '1 Compute your GPU's carbon emissions' and '2 Push for more transparency in our field by including the results in your publication (research paper, blog post etc.)'. At the bottom, there is a red button that says 'COMPUTE YOUR ML CARBON IMPACT'.

<https://mlco2.github.io/impact/>

☰ README.md

carbontracker

📦 pypi v1.1.7 python >=3.7 🔄 build passing license MIT

About

carbontracker is a tool for tracking and predicting the energy consumption and carbon footprint of training deep learning models as described in [Anthony et al. \(2020\)](#).

Citation

<https://github.com/lfwa/carbontracker>

 **Hugging Face** 🔍 Search models, datasets, users... 📦 Models 📁 Datasets 📁 Spaces 📄 Docs 📁 Solutions 📄 Pricing

← Back to blog

CO2 Emissions and the 🤗 Hub: Leading the Charge

Published April 22, 2022

Update on GitHub

 [sasha](#)
Sasha Luccioni  [muellerzr](#)
Zachary Mueller  [nateraw](#)
Nate Raw

<https://huggingface.co/blog/carbon-emissions-on-the-hub>

Visit - <https://ayon-roy.netlify.app>

Datasets to start your journey towards Sustain**AI**nability

- **Energy:** CityLearn, OPFLearn, ARPA-E GO, PowerGridworld, L2RPN, BeoBench, Building Data Genome, bbd.labworks.org, COBS, BOPTTEST/ACTB, Open Catalyst
 - **Land use:** TorchGeo, blutjens/awesome-forests, CropHarvest, Radiant ML Hub, LandCoverNet, Agriculture-Vision, chrieke/awesome-satellite-imagery-datasets
 - **Climate & Earth science:** mldata.pangeo.io, ClimateBench, ClimART
- ESGF; Pangeo, Large Ensemble Simulations (look for CMIP5 and CMIP6 data)
- ECMWF data (for historical data and near-term predictions)
- CORDEX (for downscaled climate models)
- Remote sensing data (e.g., ESA-phi lab)
- NASA climate datasets
- NCAR Climate Data Guide
- ClimateBench (for training ML models)
- **Adaptation:** wandb/droughtwatch, Global Flood Database, FloodNet, ITU GEOAI
 - **Biodiversity:** iNat dataset, LifeCLEF, FGVC, iWildCam, Movebank

Additional resources

- Carbon Brief - how do climate models work:
<https://www.carbonbrief.org/qa-how-do-climate-models-work/>
- Applied Machine Learning Tutorial for Earth Scientists
https://github.com/eabarnes1010/ml_tutorial_csu
- Analyse CMIP6 data from the cloud servers
<https://cmip6moap.github.io/resources/loading-data-xarray/>

But...

- ▶ Is AI needed to address the problem?
- ▶ What is the scope & time horizon of the impact?
- ▶ What is the likelihood that a solution can be found?
- ▶ Can a solution feasibly be deployed?
- ▶ What are the potential side effects of deploying the candidate solution?
- ▶ Who are the relevant stakeholders who are involved in or affected by the application?

Key considerations

- AI is not a silver bullet and is only relevant sometimes
- High-impact applications are not always flashy
- Interdisciplinary collaboration
 - ▶ Scoping the right problems
 - ▶ Incorporating relevant domain information
 - ▶ Shaping pathways to impact
- Equity considerations
 - ▶ Empowering diverse stakeholders
 - ▶ Selecting and prioritizing problems
 - ▶ Ensuring data is representative

Responsible Sustain**AI**nability

Mitigating biases in data and models

- ▶ E.g., Buildings data: Housing discrimination, geographic disparities in availability
- ▶ E.g., Weather models: Calibration may be optimized for particular regions

Improving trustworthiness and accountability

- ▶ Safety and robustness: Critical in, e.g., power systems and industrial operations
- ▶ Interpretability and auditability: Critical in, e.g., policymaking contexts centering equity and climate justice
- ▶ Centering diverse stakeholders: E.g., industrial ag vs. smallholder farmers
- ▶ Avoiding centralization: Democratized capacity and compute, digital divide
- ▶ Avoiding digital colonialism: E.g., smart meters, analysis of remote sensing data

Let us come together & do our bit

Green
New Energy

Artificial Intelligence Is Booming—So Is Its Carbon Footprint

Greater transparency on emissions could also bring more scrutiny

OCTOBER 24, 2022

AXA Future Risks Report 2022

Climate change has become the number one concern around the world

IN THE NEWS | ⌚ 1 MINUTE

- Climate change is the number one risk in all geographic areas
- Geopolitical risks rank second, overtaking cyber and pandemic
- Sense of vulnerability to certain risks is increasing and the level of trust is deteriorating

Over the last year, economic and geopolitical challenges have added a new layer of uncertainty to the disruptions caused by the Covid pandemic and the climate crisis.

<https://www.axa.com/en/news/2022-future-risks-report>

<https://www.climatechange.ai/papers> : Great place to look for papers on Climate Change related topics

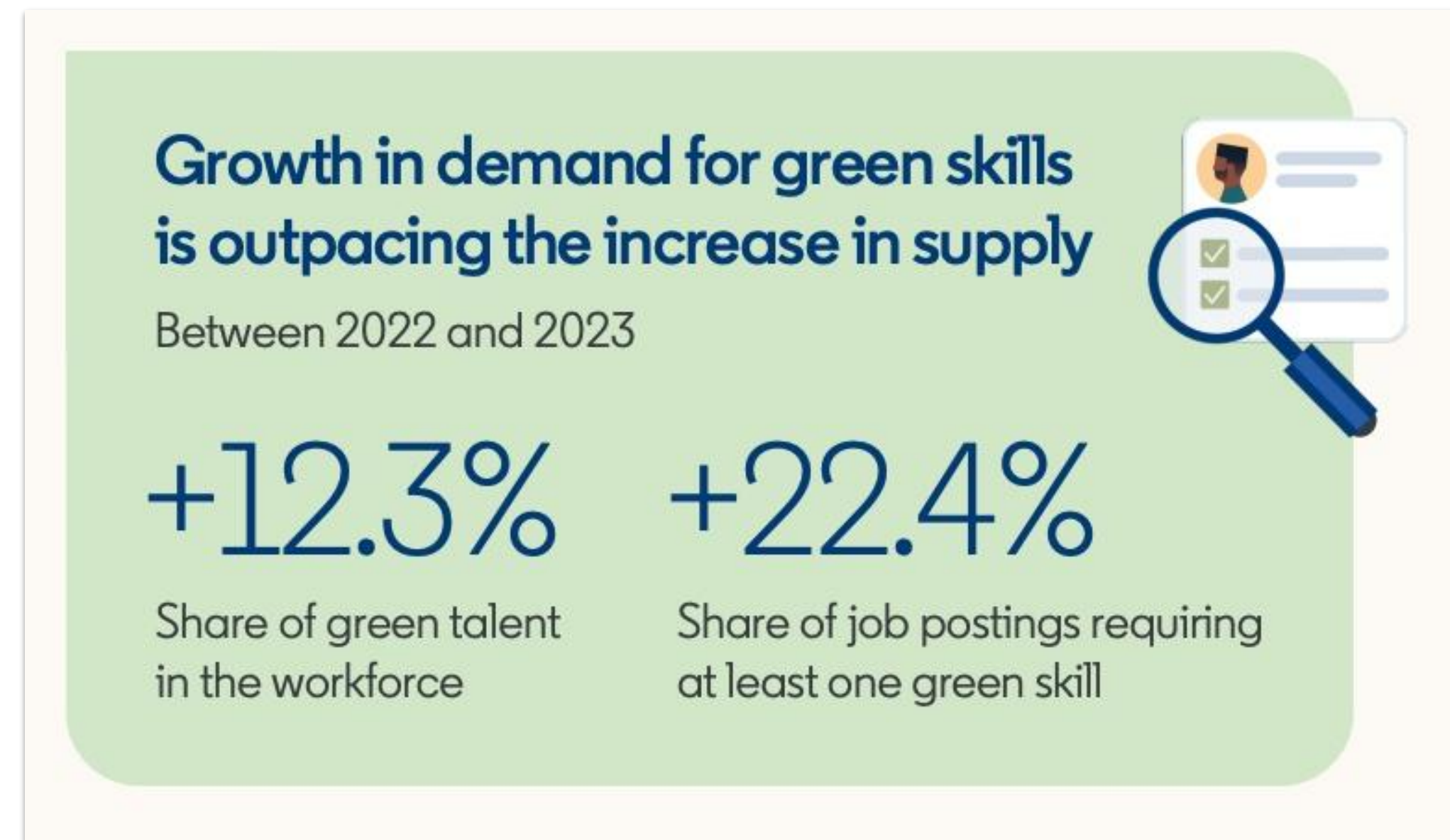
<https://www.bloomberg.com/news/articles/2023-03-09/how-much-energy-do-ai-and-chatgpt-use-no-one-knows-for-sure#xj4y7vzkq>

https://www.dkn-future-earth.org/activities/working_groups/107334/index.php.en : Research group work on Sustainable ML

Visit - <https://ayon-roy.netlify.app>

Why skill yourselves Sustainably ?

The green economy faces large talent bottlenecks



Source: [LinkedIn Global Green Skills Report 2023](#)

Let me answer your Questions now

Finally, it's your time to speak!



“

AI IS A POWERFUL TECHNOLOGY AND A FORCE FOR GOOD, BUT IT'S IMPORTANT TO BE CONSCIOUS OF ITS GROWING ENVIRONMENTAL IMPACT. THE CODE CARBON PROJECT AIMS TO DO JUST THAT, AND I HOPE THAT IT WILL INSPIRE THE AI COMMUNITY TO CALCULATE, DISCLOSE AND REDUCE ITS CARBON FOOTPRINT.

YOSHUA BENGIO

CODE CARBON

Danke Schoen

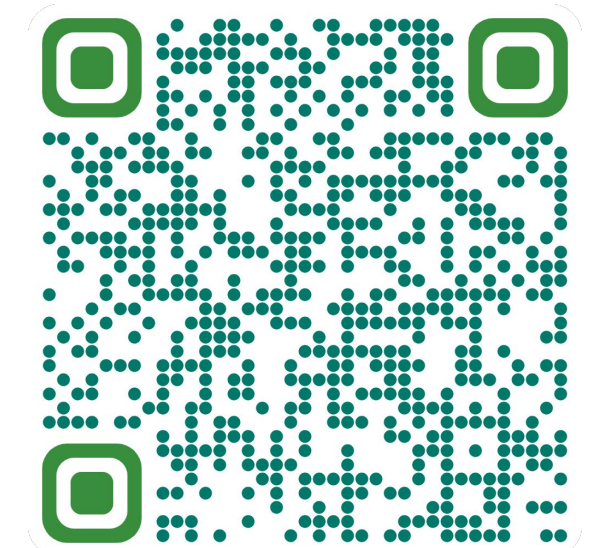
Questions ? Any Feedbacks ? Did you like the talk?
Tell me about it.

If you think I can help you,
connect with me via

Email : ayon-roy@outlook.com

LinkedIn : <https://www.linkedin.com/in/ayon-roy>

Website : <https://AYONROY.ML/>



**Download the
slides**