Sustainable AI: Rethinking the AI Revolution

Brendan Conroy

Department of Science and Mathematics

University of Minnesota Morris

Spring 2024

note: all sources of figures, images, and text are included at the end of presentation

Outline

Background

What is ***Artificial Intelligence*** (AI)?

- Machines performing tasks of human level intelligence

What is ***Machine Learning*** (ML)?

- A machine learning ***model*** learns patterns from data by adjusting ***parameters*** to achieve targeted output
- You can think of a ***model*** as a giant, giant, equation with millions of variables, input, and output.

What are ***Neural Networks***?

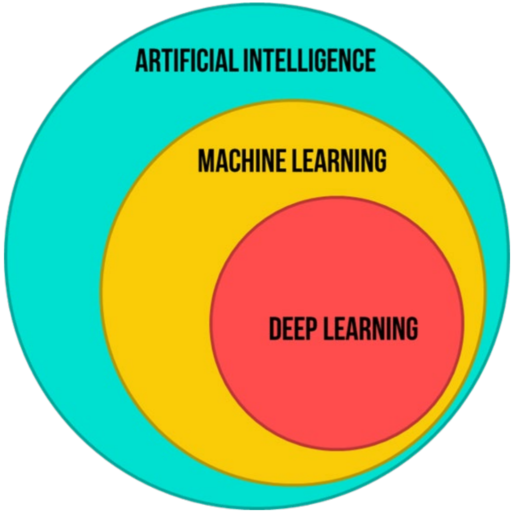
- A form of ML including a network of nodes

What is ***Deep Learning***?

- Refers to Neural Networks with more than one layer.

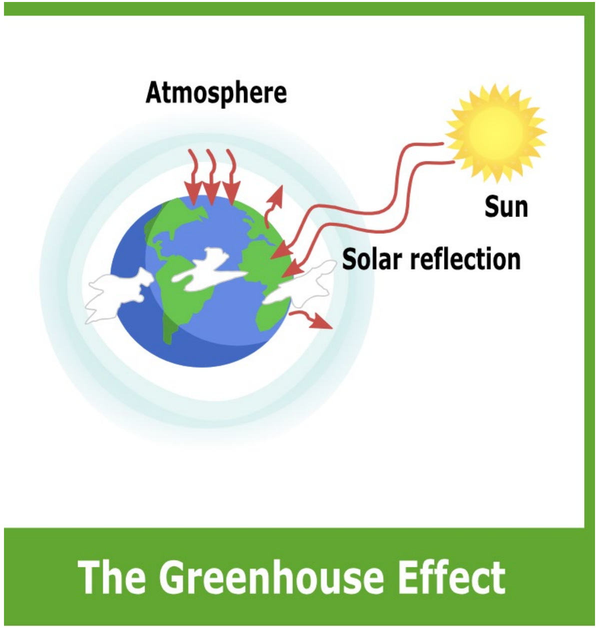
What is ***Green or Sustainable AI?***

- A new way of thinking that focuses on the carbon footprint of AI.



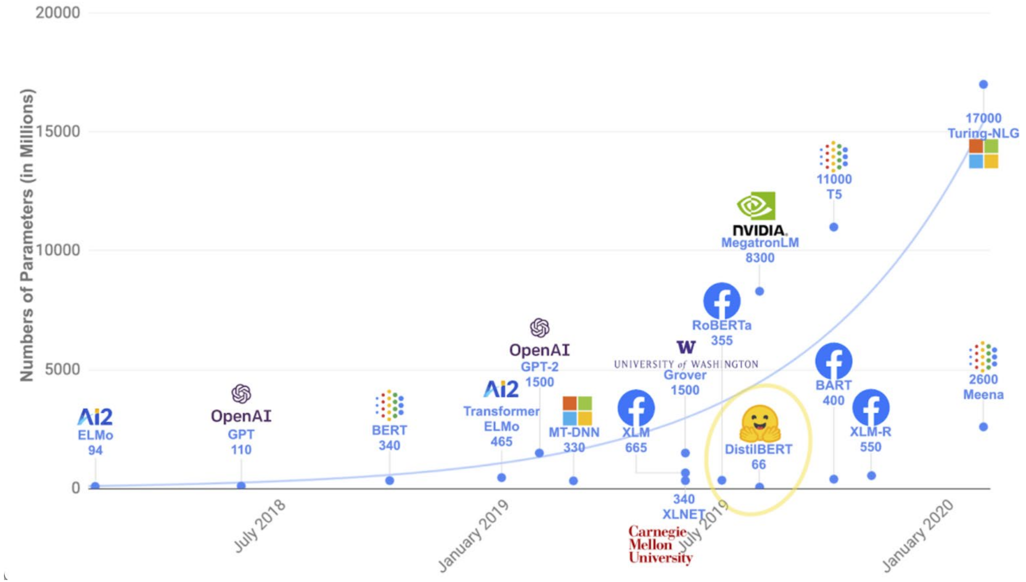
More Background

- What is carbon footprint?
  - Total amount of greenhouse gas emissions generated
- What is the greenhouse effect?
  - Greenhouse gases trap heat in our atmosphere



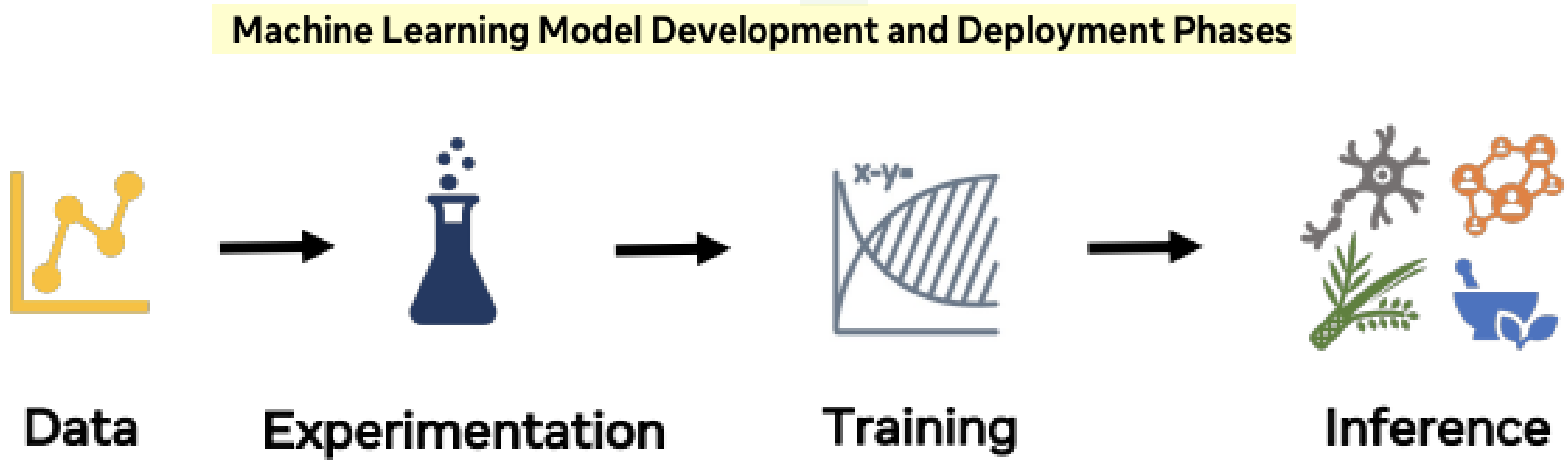
Motivation:

Why do we have to change the way we think about AI?



Note: GPT-4 has 1.76 trillion parameters!

Phases of Model Development



Phase 1 of Model Development: Data Processing

- Raw data collected, and processed prior to training
- Processing may include cleaning, formatting, removing missing values, etc.
- Not typically a big contributor to carbon footprint

Phase 2: Experimentation

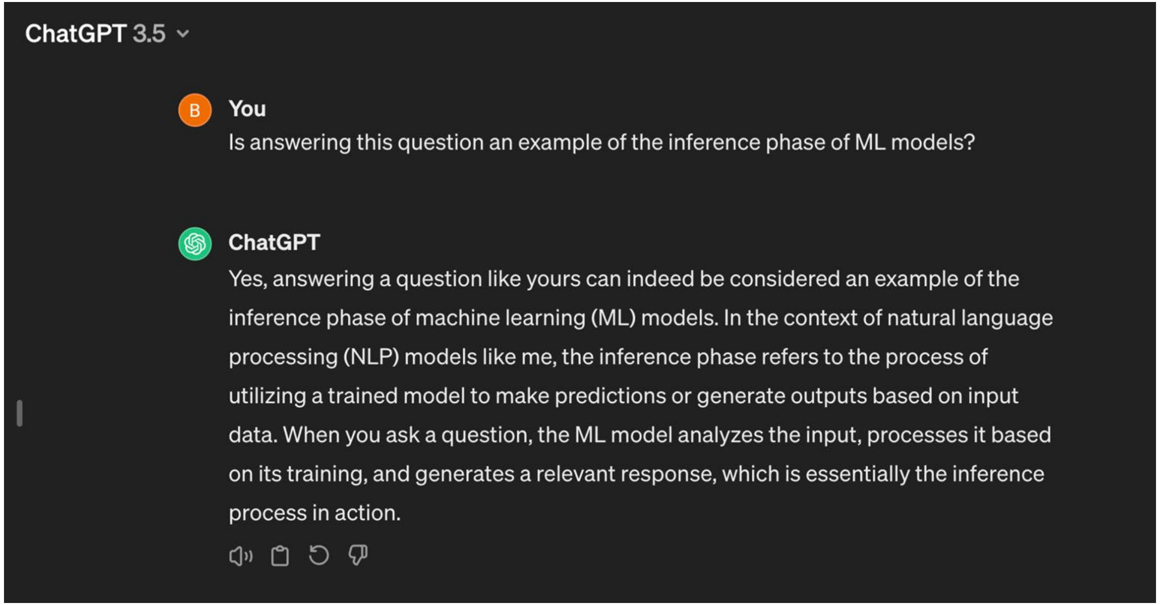
- Determining the most efficient model architecture and hyper-parameters
    - Model architecture Overall structure and design of model
    - Hyper-parameters: Parameters adjusted prior to learning process
- Different model architectures are considered and tested, often simultaneously
- This phase often has a significant demand for power consumption and furthermore a significant effect on carbon footprint

Phase 3: Training

- Processed data fed into selected model architecture
- Where the magic “learning” happens
- Models:
    1. learn patterns and relationships from ***training data***
    2. Adjust parameters to minimize ***loss function***
        a. Loss function: A function that captures the difference between model’s prediction and the actual expected output.
- Training is completed until the model is determined to be ***accurate*** enough
- This phase typically has the most significant power consumption and effect on carbon footprint

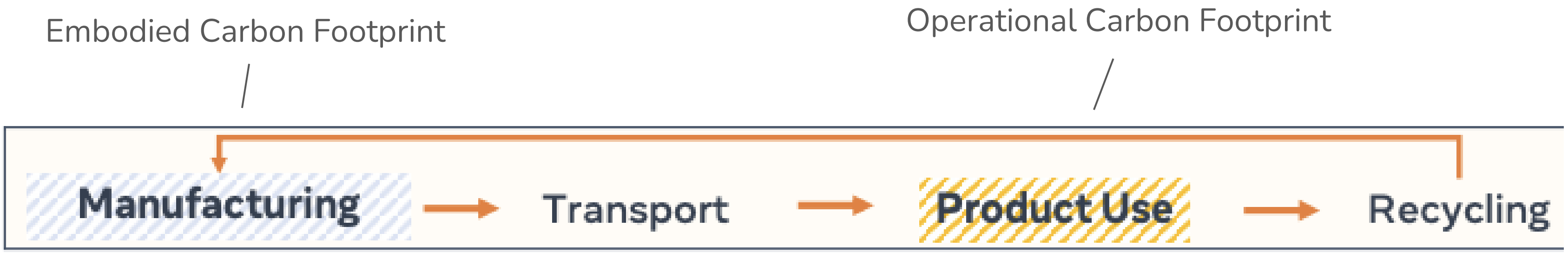
Phase 4: Inference

- Trained model is ready to make predictions on other data at this phase.
- New input is fed into the model, and the model returns some type of prediction or output



System Life Cycle

- Life cycle of ***physical hardware/infrastructure*** that is associated with AI



The Carbon Footprint of AI

=

Operational Carbon Footprint: ***Product use*** emissions

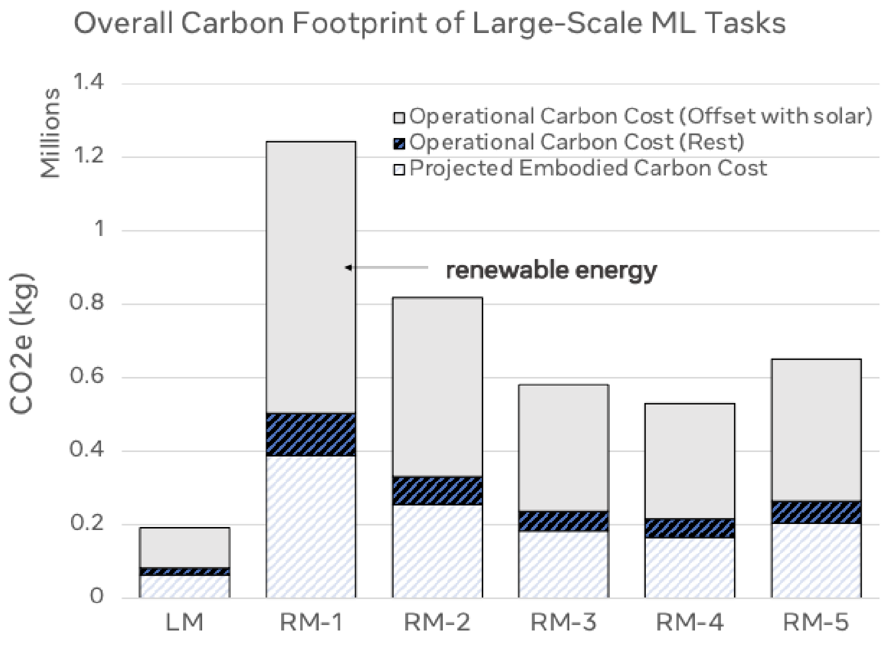
+

Embodied Carbon Footprint: ***Manufacturing*** emissions

The Carbon Footprint of AI: A Case Study

- 5 ML models developed by Meta
- LM: Universal Language Model for text translation
- RM-1-RM-5: Deep learning models developed to recommend and rank Meta products.

- In this case, carbon footprint falls mostly on ***Embodied Carbon Cost***
- Without renewable energy integration, the majority of carbon cost is ***Operational Carbon Cost***



A Green AI Mindset: Overview

1. Accuracy vs Efficiency: A necessary shift in how we define success

1. A holistic approach to capturing AI’s footprint

1. A responsibility to minimize the carbon footprint of AI

Accuracy vs. Efficiency

- Currently, success of a model is defined strictly by it’s ***accuracy***
- Endless pursuit of more accuracy requires bigger models, bigger footprint.
- A new **key** question: At what cost does the model achieve its accuracy?
- ***Efficiency*** needs to be a metric of success alongside ***accuracy***.
- Increased ***model transparency*** is key to being able to analyze the efficiency of models
- ***Successful AI***: achieves new results and accuracy while minimizing computational cost with efficient design principles

A holistic approach

- Includes both ***operational***, and ***embodied*** carbon footprint in analysis
- the footprint of hardware and infrastructure is considered along the complete ***system life cycle***
- Every ***phase of model development*** is optimized towards efficiency

Responsibility: an ethical lens

- We have a ***responsibility*** to reduce the carbon footprint of AI
- ***Technological ethics lens***: As with any new technology, creators and users have a responsibility to minimize negative effects of the technology on society.
- As an individual not associated with developing these models, this can still apply to you!
- limit AI use to necessary use, vote, and spread awareness !

Additional ways to make AI more green

- ***Floating point operators*** (FPO) provide an estimate of work performed by computation process, and can be used as a metric for efficiency
- Requiring AI work published to report ***FPO*** or other metrics of efficiency.
- Reporting efficiency incentivizes efficiency and helps others to learn about how to develop models more efficiently
- Require reports about the ***experimentation*** phase of development
  - How many model architectures were tested? What was learned?
- Encouraging releasing ***trained models*** to the public, to avoid the carbon cost of having to retrain models

Renewable Energy Integration as a Solution

- As seen in case study earlier, ***renewable energy integration*** can significantly reduce the ***operational carbon*** footprint of AI models.
- An issue: Intermittent nature of renewable energy
- An AI solution: AI can be used as a tool to predict fluctuations in renewable energy production, and make our grids more stable and balanced

Conclusion

- It is critically important that we make a shift in the way that we approach developing and discussing AI.
- This shift needs to include holistically examining the environmental footprint of AI along every phase of development, and every life cycle stage of associated hardware.
- We have a responsibility to manage the negative environmental implications this rapidly growing technology has on our world!

What can you do to help with this issue?

Reduce, Reuse, Recycle!

Acknowledgements

- Thanks to all the CSCI Faculty: Elena Machkasova, Kristin Lamberty, Peter Dolan, Nic McPhee, Kristofer Schlieper, and Wenkai Guan
- Thanks to my family!
- Udit Gupta’s research and work on this topic inspired a lot of my work here
  - His website: https://ugupta.com/
- Thanks to my classmates and peers for supporting me in many ways!
- Thank you all for listening!

Sources

[1] Carole-Jean Wu, Ramya Raghavendra, Udit Gupta, Bilge Acun, Newsha Ardalani, Kiwan Maeng, Gloria Chang, Fiona Aga Behram, James Huang, Charles Bai, Michael Gschwind, Anurag Gupta, Myle Ott, Anastasia Melnikov, Salvatore Candido, David Brooks, Geeta Chauhan, Benjamin Lee, Hsien-Hsin S. Lee, Bugra Akyildiz, Maximilian Balandat, Joe Spisak, Ravi Jain, Mike Rabbat, and Kim Hazelwood. 2022. Sustainable AI: Environmental Implications, Challenges and Opportunities. arXiv:2111.00364 [cs.LG]

(2) Roy Schwartz, Jesse Dodge, Noah A. Smith, and Oren Etzioni. 2020. Green AI. Commun. ACM 63, 12 (December 2020), 54–63. https://doi.org/10.1145/3381831

Sources

https://research.aimultiple.com/gpt/

https://getgenie.ai/gpt-3-vs-gpt-4/

https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.advancinganalytics.co.uk%2Fblog%2F2021%2F12%2F15%2Funderstanding-the-difference-between-ai-ml-and-dl-using-an-incredibly-simple-example&psig=AOvVaw0X8I-DbBvNWsqC2AmJFETh&ust=1712965258134000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhxqFwoTCPC2uP6qu4UDFQAAAAAdAAAAABAE

https://chat.openai.com/

https://commons.wikimedia.org/wiki/File:Meta-Logo.png

Questions?