Reducing Carbon Footprint of AI Models Without Compromising Performance

Austin Deng, Xingzhi Huang, Michael Lu

Austin Deng
Vandegrift High School
austin12709@gmail.com

Short Resume

- 2 years experience with AI in Competitive Robotics
- 2 years working with Image Classification/Computer Vision models

The problem

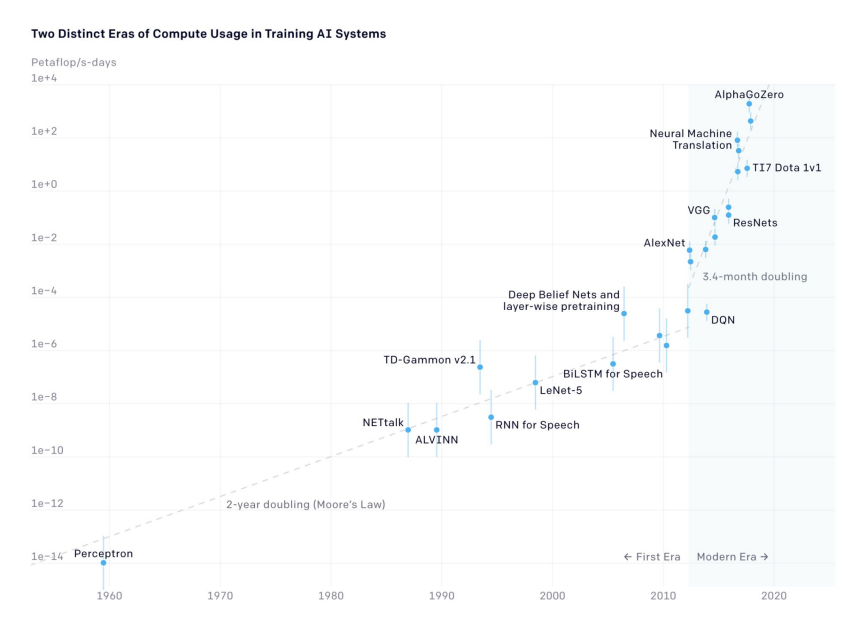
- Starting in the 20th Century, AI has become more and more popular

- AI Models create significant amounts of CO_2
- Environmental impacts of AI have not been fully studied
- Reducing CO_2 Emissions is crucial

Green AI

AI that attempts to reduce the electricity usage of models while maintaining quality

- Most work on Green AI focuses on the training costs of AI
- Computing power required to train AI has grown exponentially



*computational power increased 300,000x over the last 10 years

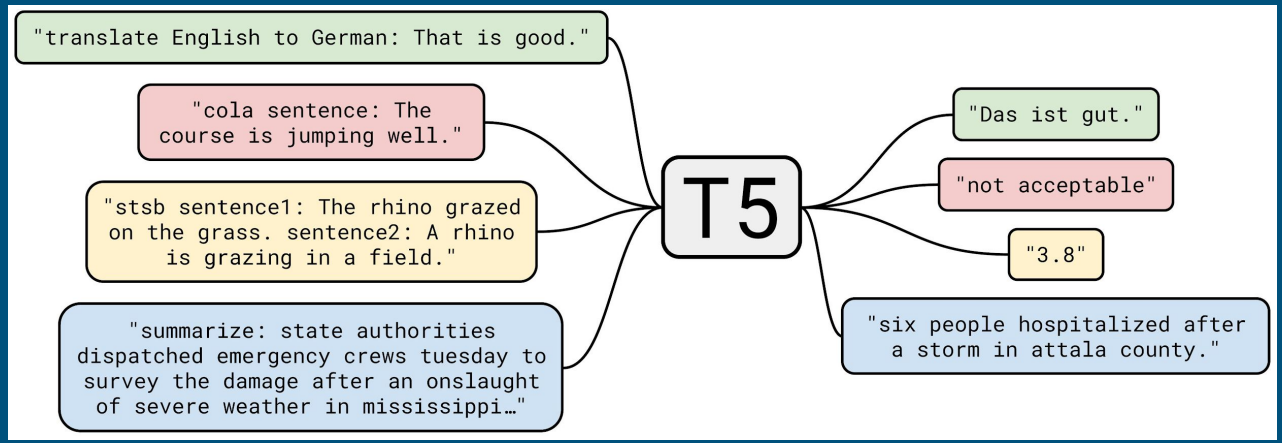
Introduction

- Focus on electricity usage of running AI models
- Deployed wide variety of categories from Huggingface
  - Text to Text Summary
  - Text to Image Generation
  - Image Classification
  - Optical Character Recognition
- Used cloud computing to deploy AI models

Text to Text Summary

T5 Models

- Basic Summary Model–Many summarization models are based on the T5 model
  - T5-one-line-summary
  - T5-base-finetuned-summary-model



T5-one-line-summary vs. T5-base-finetuned-summary-model

Prompt:

President FDR proposed 15 important bills to the Congress in his first 100 days of Presidency. These bills were a part of the New Deal program and 3 Rs: relief, recover, reform. Because of the New Deal Program, the United States recovered faster from the Great Depression.

Results:

T5-one-line-summary:

FDR's New Deal Program

T5-base-finetuned-summary-model:

FDR proposed 15 important bills to the Congress in his first 100 days of Presidency

Text to Image Generation

Stable Diffusion models

- Compvis v1-4
- Runway v1-5
- StabilityAI v2-1

Generally, earlier stable diffusion models produce lower quality images.

Image comparison

Prompt: a photo of a beautiful desert landscape at night

Compvis v1-4

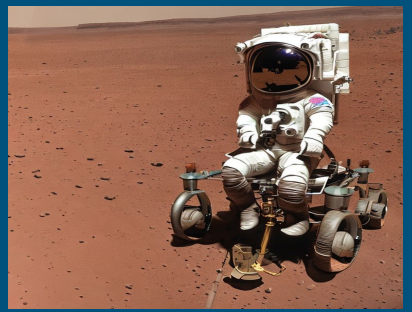
Runway v1-5

Stability v2-1

Methodology

- Generated 10 images with varying prompts
  - Measured electricity usage for all 10 prompts
- Not able to measure quality
  - No reliable objective way to measure “correctness” of the model

Runway v1-5



Stability v2-1

Prompt: a photo of an astronaut riding a horse on mars

Image to Text Classification

Image Classification Models

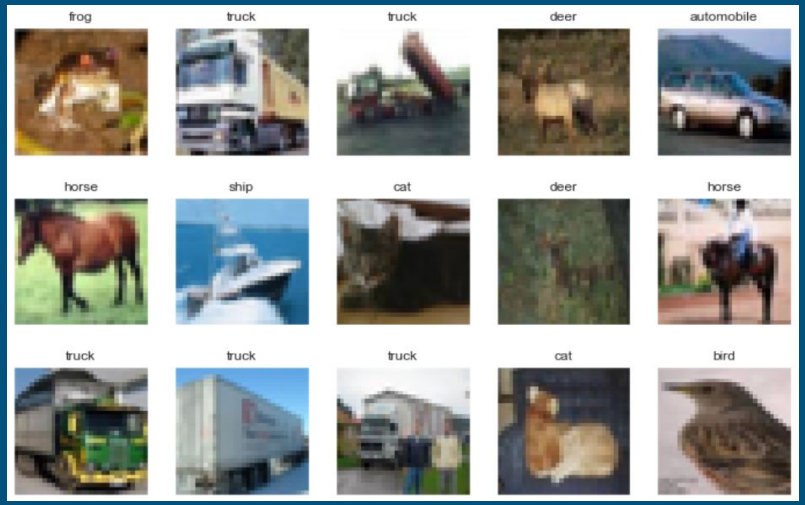
- ViT
  - Google-vit-base-patch16-224
  - Google-vit-base-patch16-384
- CvT
  - Microsoft-cvt-13
- ResNet
  - Microsoft-resnet-50

Example

Google-vit-384	trailer truck
Google-vit-224	trailer truck
Microsoft-cvt	pickup truck
Microsoft-resnet	minivan

Methodology

- CIFAR-10
  - 60000 32x32 images
  - Quick evaluation time
  - Low resolution
- Electricity
  - Classified 2500 CIFAR-10 images
- Accuracy
  - Classified 100 CIFAR-10 images



Optical Character Recognition

TROCR models

- TROCR uses transformers to convert images to text
  - microsoft/trocr-large-printed
  - microsoft/trocr-base-printed

Example

Output: HOME

Output: WORDS

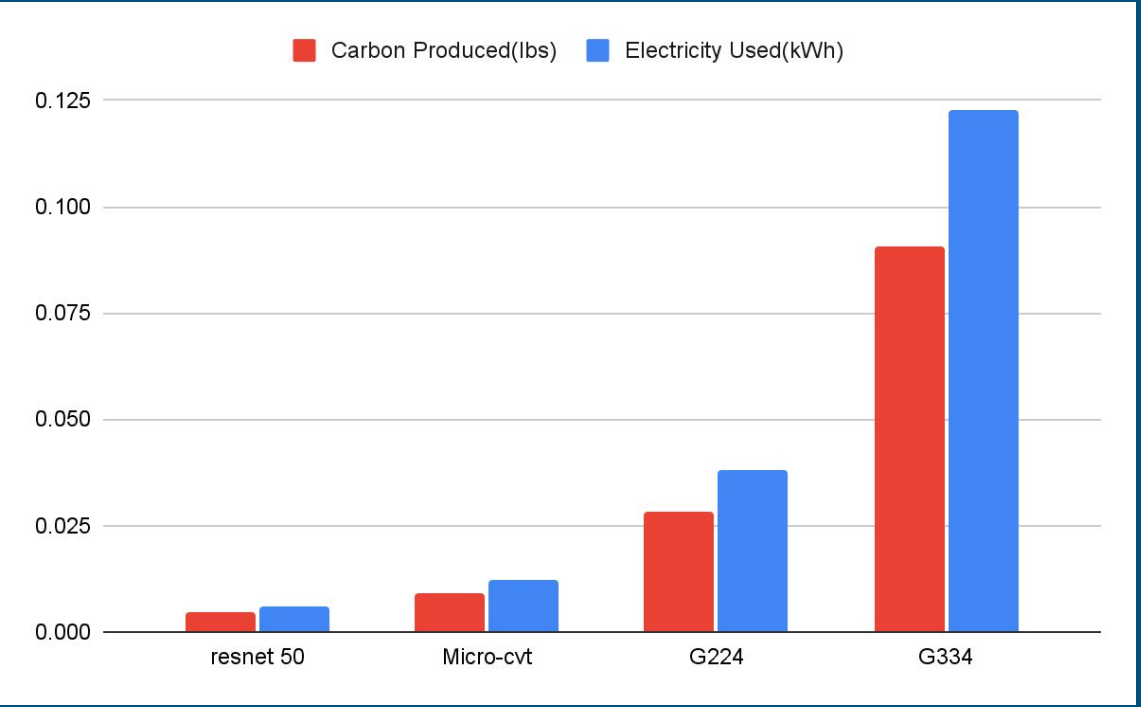
Methodology

- Measured Electricity usage with 10 images
- Determined accuracy using custom images and research paper

Saving Carbon with Small AI Models

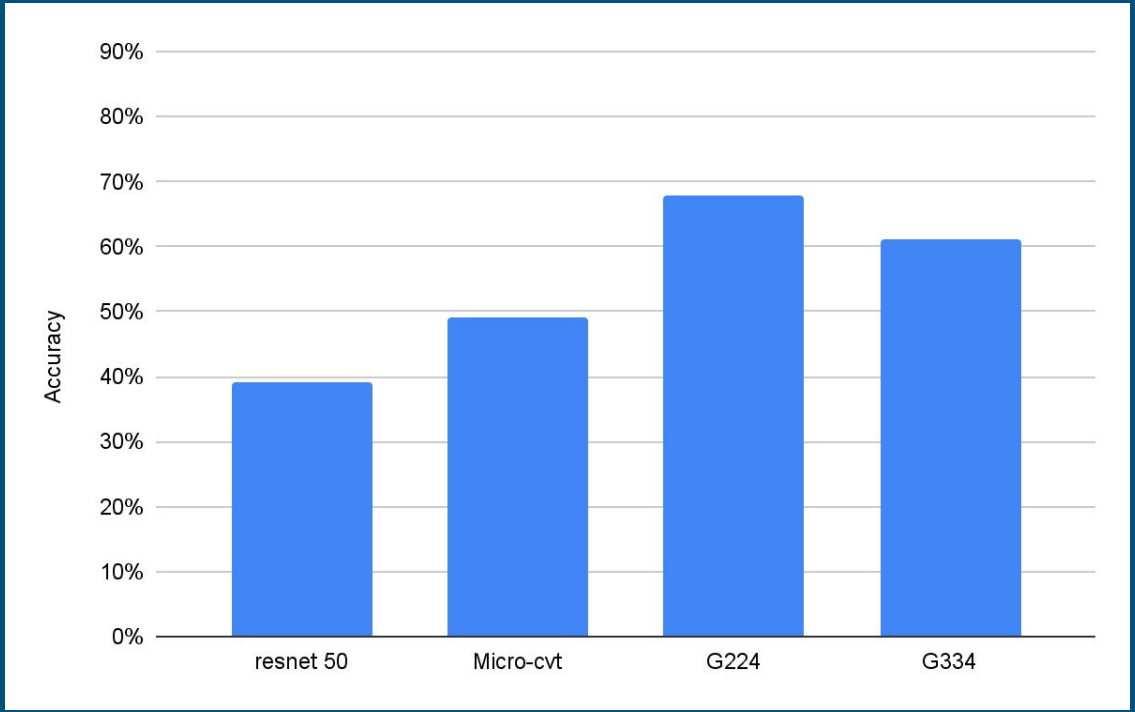
Image Classification Models

Running the models through 2500 of the CIFAR-10 image dataset

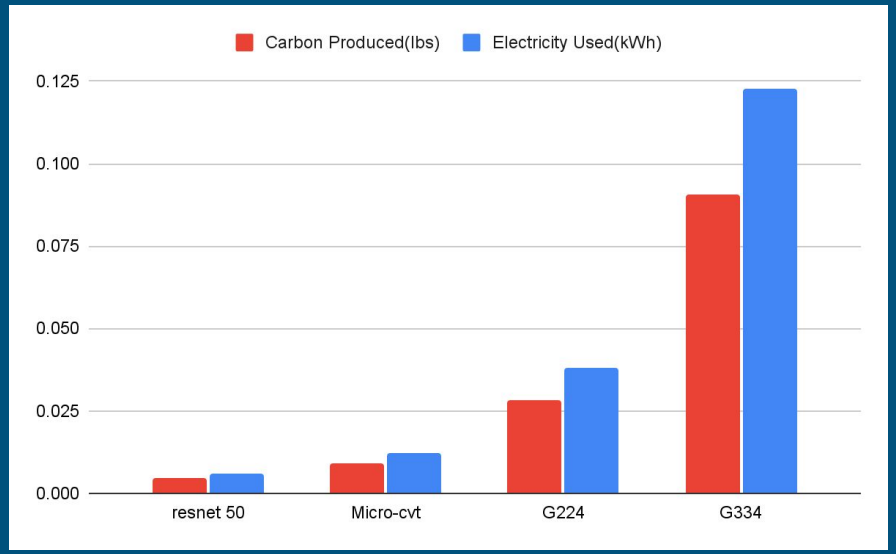


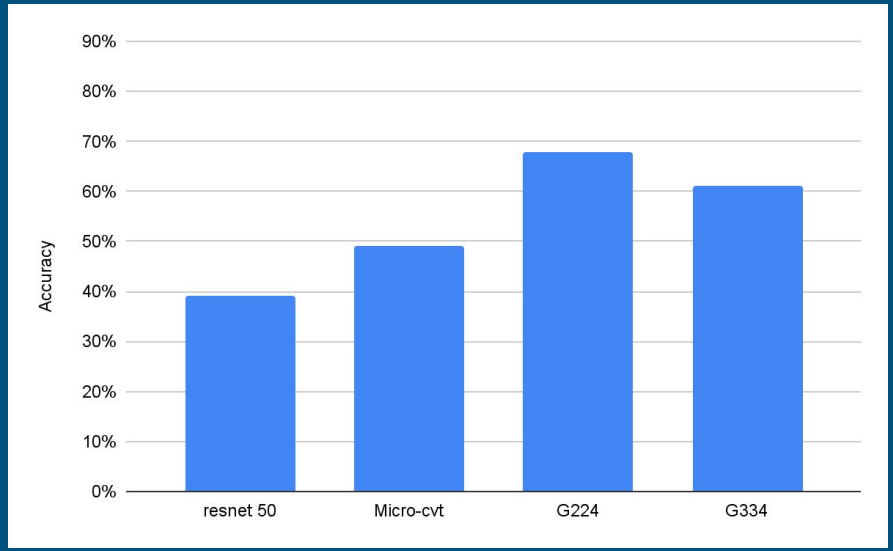
Accuracy of Image Classification

- CIFAR-10 dataset
- 100 images
- Accuracy of prediction

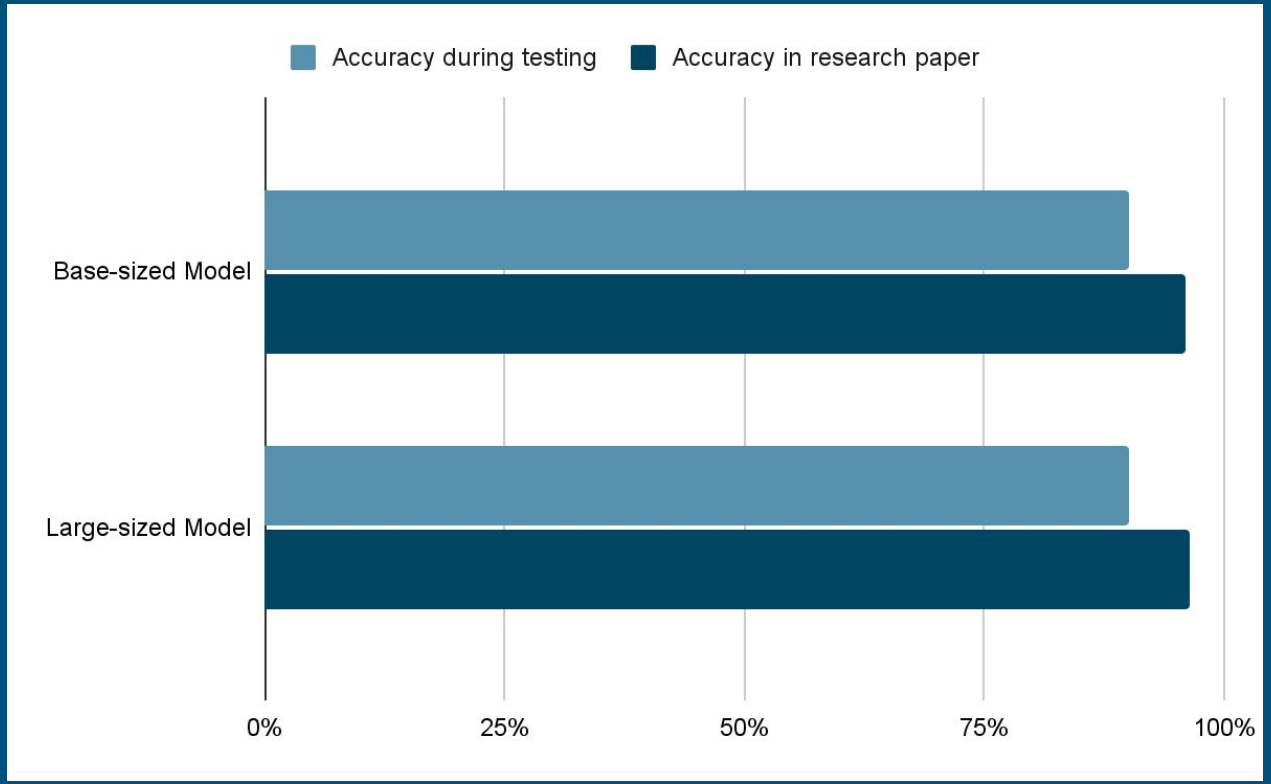


Comparison


Carbon Produced(lbs)
Electricity Used(kWh)
0.125
0.100
0.075
0.050
0.025
0.000
resnet 50
Micro-cvt
G224
G334


Accuracy
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
resnet 50
Micro-cvt
G224
G334

Accuracy of the TrOCR image-to-text Models

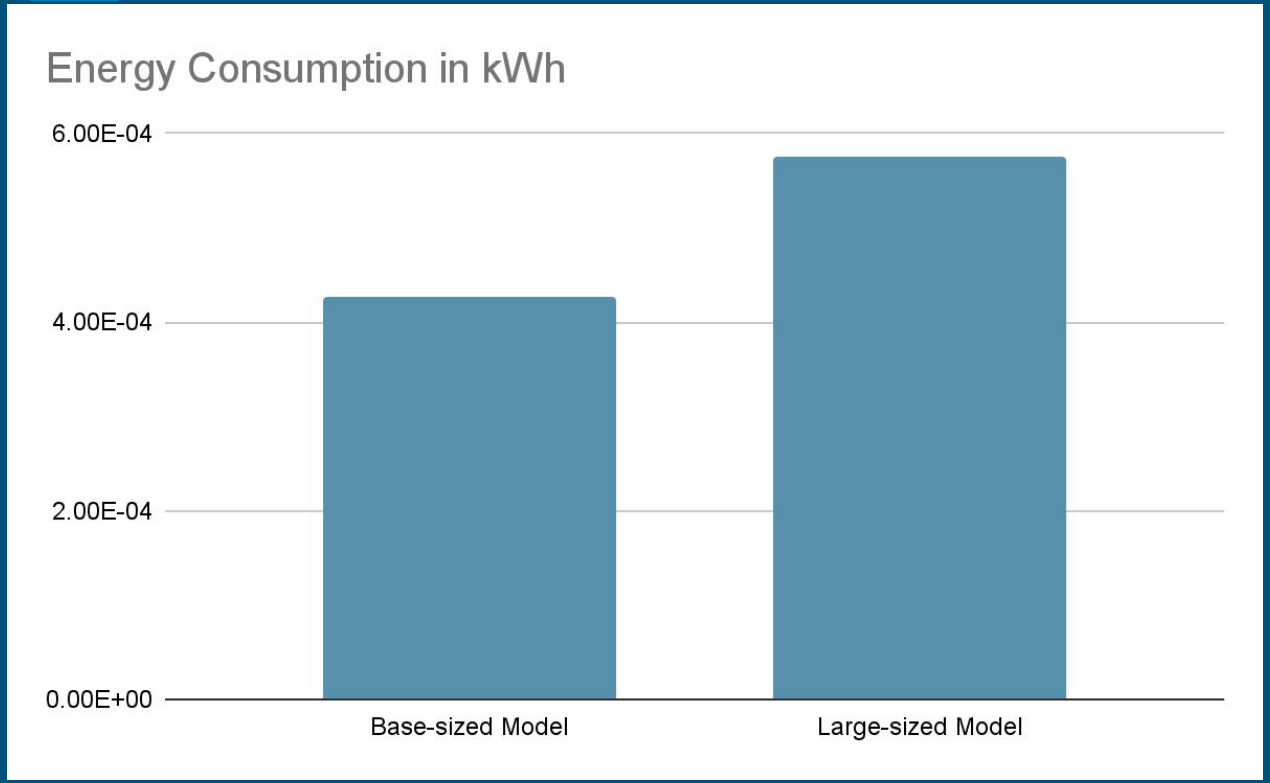


- Experimentation shows both large-sized model and base-sized model about the same accuracy

- original paper for the models shows both have comparable accuracy

Model	Recall
CRNN	28.71
Tesseract OCR	57.50
H&H Lab	96.35
MSOLab	94.77
CLOVA OCR	94.3
TrOCR$_{SMALL}$	95.89
TrOCR$_{BASE}$	96.37
TrOCR$_{LARGE}$	96.59

Energy Usage of the TrOCR Models



Average Energy Savings when using base-sized model:

25.96%

Saving Carbon on Large AI Models

Data Analysis

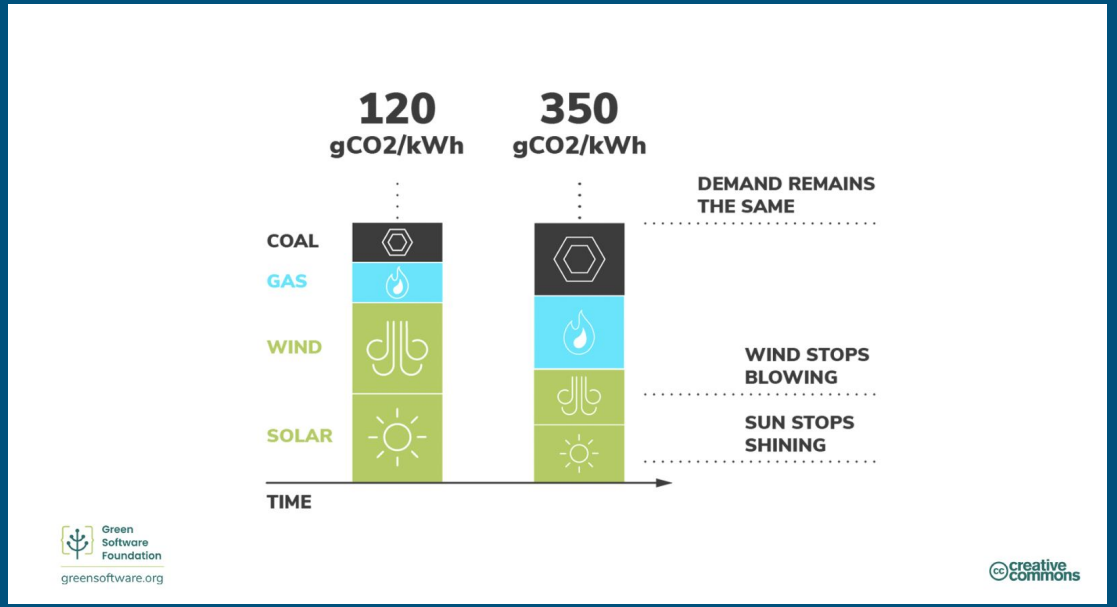
- Compvis v1-4 and runway v1-5 similar accuracy and quality
- Stability v2-1 less accurate, higher quality pictures

Electricity usage

Stability 2-1	0.01438 kWh
Runway 1-5	0.00321 kWh
Compvis	0.00323 kWh

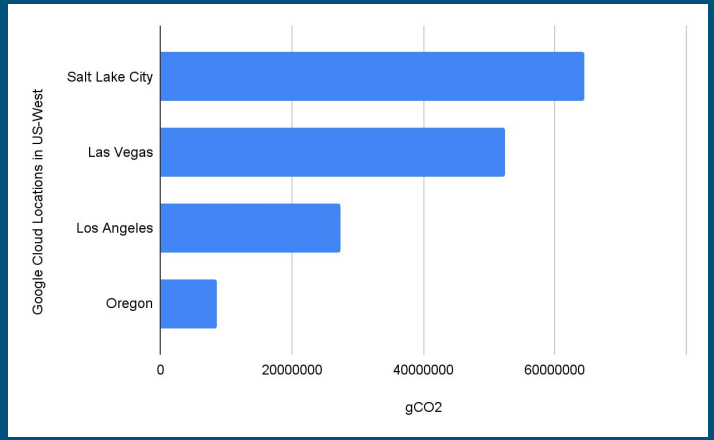
Carbon Intensity: How “Clean” the Energy Is

	A	B
1	Cloud Region/Location	Grid Carbon Intensity(gCO2/kWh)
2	Taiwan	456
3	Hong Kong	360
4	Tokyo	464
5	Mumbai	670
6	Singapore	372
7	Sydney	598
8	Melbourne	521
9	Warsaw	576
10	Finland	127
11	Madrid	121
12	Belgium	110
13	London	172
14	Paris	59
15	Toronto	29
16	São Paulo	129
17	Iowa	394
18	South Carolina	434
19	North Virginia	309
20	Dallas	296
21	Oregon	60
22	Los Angeles	190
23	Salt Lake City	448
24	Las Vegas	365



Carbon usage

- 100M images
- Approximately 143.8 MWh used
- Drastically reduce carbon emissions by switching cloud regions
- Based in SLC, Utah - 64.4 million grams
- Based in Oregon - 8.6 million grams
- 86.6% decrease



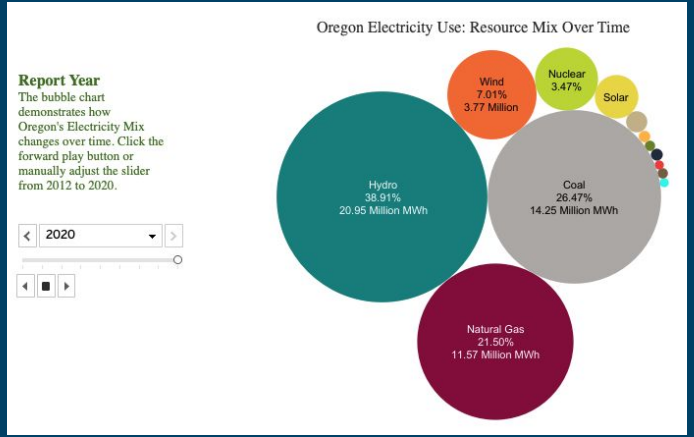
Data Analysis

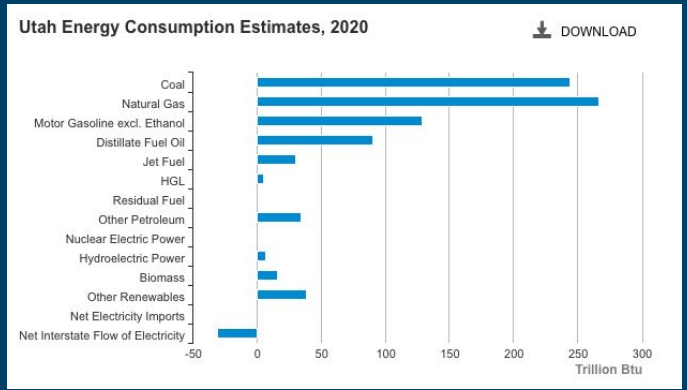
Oregon

- Oregon draws significant Hydro power
  - Consistent energy
- Much lower than U.S average fossil fuel consumption

Utah

- Most renewables come from less reliable sources like wind
- Significant usage of fossil fuels





Conclusions

- Larger models are not necessarily better in some tasks
- Cloud computing can be used to significantly reduce carbon production
- Carbon reduction could become more effective over time