

More Sustainable AI Use and Development: Software Side

AI and Sustainability Symposium

Dr. Emily Hastings

Department of Computer Science

AI Context

- **AI has been studied since 1950s**
- **First chatbot in 1966**
- **First data center in 1945**

A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

Welcome to

```
EEEEEE LL      IIII  ZZZZZZ  AAAAA
EE      LL      II     ZZ     AA  AA
EEEEEE LL      II     ZZZ   AAAAAAA
EE      LL      II     ZZ     AA  AA
EEEEEE LLLLLL  IIII  ZZZZZZ  AA  AA
```

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

```
ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   I think I am.
```

Defining Generative AI

To understand generative artificial intelligence (GenAI), we first need to understand how the technology builds from each of the AI subcategories listed below.

Expert System AI

Programmers teach AI exactly how to solve specific problems by providing precise instructions and steps.

Artificial Intelligence

The theory and methods to build machines that think and act like humans.



Machine Learning

The ability for computers to learn from experience or data without human programming.

Deep Learning

Mimics the human brain using artificial neural networks such as **transformers** to allow computers to perform complex tasks.



Generative AI

Generates new text, audio, images, video or code based on content it has been **pre-trained** on.

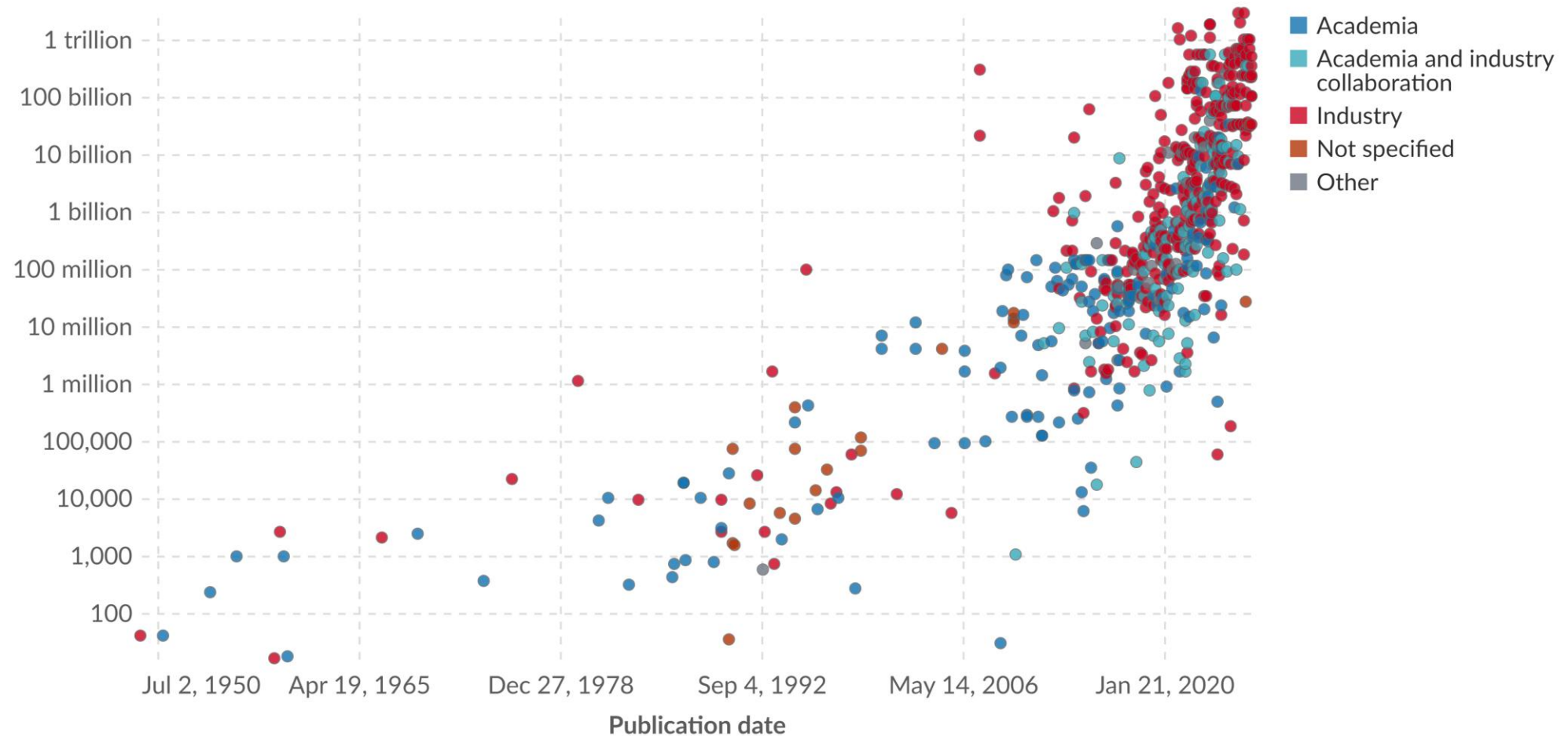


ChatGPT Midjourney Gemini

Parameters in notable artificial intelligence systems

Parameters are variables in an AI system whose values are adjusted during training to establish how input data gets transformed into the desired output; for example, the connection weights in an artificial neural network.

Number of parameters (plotted on a logarithmic axis)



Data source: Epoch AI (2025)

OurWorldinData.org/artificial-intelligence | CC BY

Note: Parameters are estimated based on published results in the AI literature and come with some uncertainty. The authors expect the estimates to be correct within a factor of 10.

Two Phases:

- **Training**

- Consumes 7-8 times more energy than typical computing workload

- **Inference**

- A ChatGPT query consumes ~5 times more electricity than a web search

Examples of Potential Harms

Harm to People

- Individual: Harm to a person's civil liberties, rights, physical or psychological safety, or economic opportunity.
- Group/Community: Harm to a group such as discrimination against a population sub-group.
- Societal: Harm to democratic participation or educational access.

Harm to an Organization

- Harm to an organization's business operations.
- Harm to an organization from security breaches or monetary loss.
- Harm to an organization's reputation.

Harm to an Ecosystem

- Harm to interconnected and interdependent elements and resources.
- Harm to the global financial system, supply chain, or interrelated systems.
- Harm to natural resources, the environment, and planet.

For Users

- **You probably already think critically about the things you consume...**
- **Do the same for your AI use.**

CLIMATE CHANGE AND ENERGY

We did the math on AI's energy footprint. Here's the story you haven't heard.

The emissions from individual AI text, image, and video queries seem small—until you add up what the industry isn't tracking and consider where it's heading next.

by **James O'Donnell** and **Casey Crownhart**

May 20, 2025

Let's say you're running a marathon as a charity runner and organizing a fundraiser to support your cause. You ask an AI model 15 questions about the best way to fundraise.

Then you make 10 attempts at an image for your flyer before you get one you are happy with, and three attempts at a five-second video to post on Instagram.

You'd use about 2.9 kilowatt-hours of electricity—enough to ride over 100 miles on an e-bike (or around 10 miles in the average electric vehicle) or run the microwave for over three and a half hours.

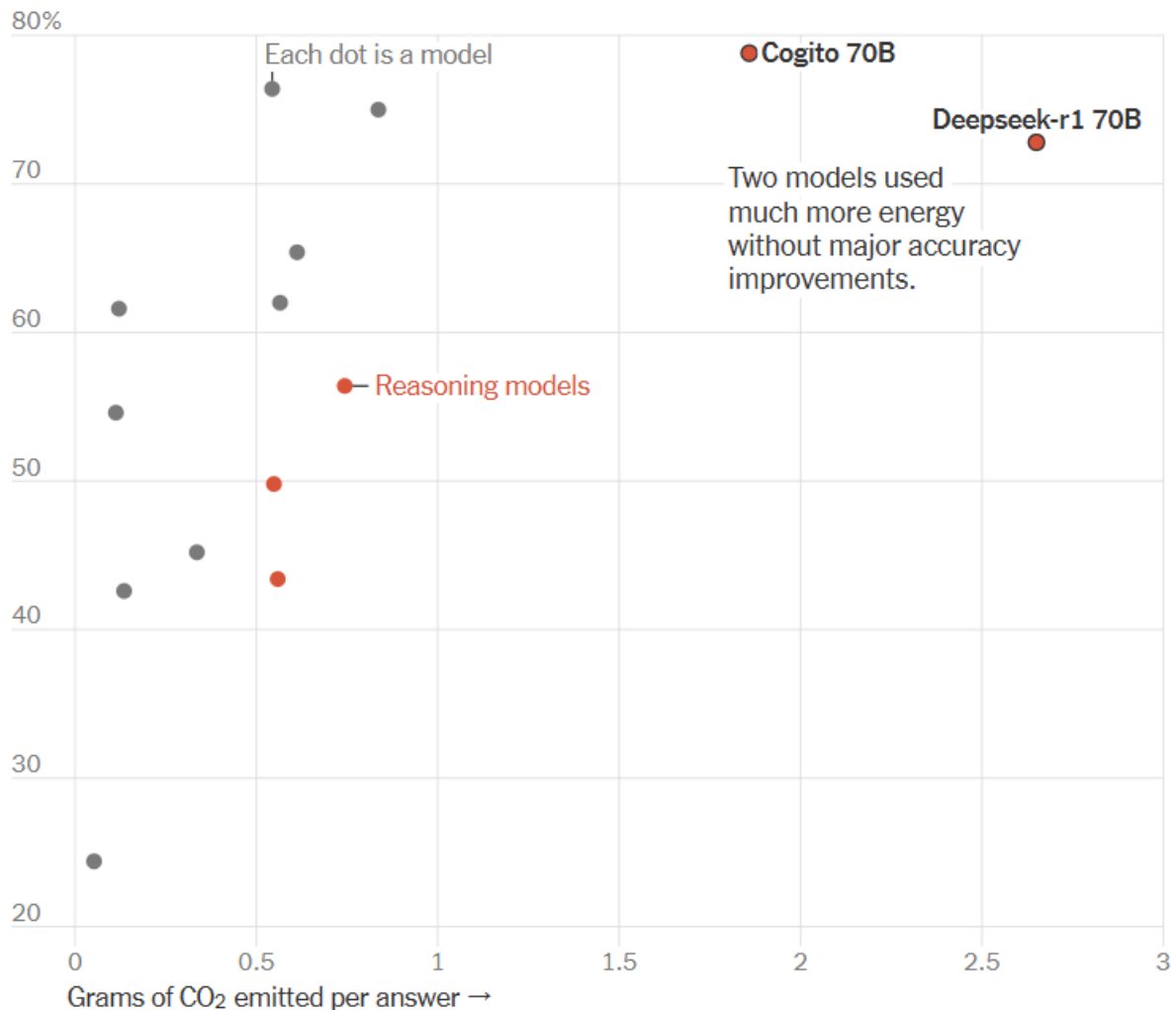
All-Purpose AI, Model Size, and Environmental Impact

- **Same technology being used as companion chatbots, enterprise productivity tools, etc.**
 - Interaction styles and risk profiles are different
 - NIST Framework emphasizes type of risk and risk tolerance is very contextual and use case specific
- **Benefits to more tailored, use-case specific models that could be smaller and use less energy:**

Dauner, Maximilian, and Gudrun Socher. "Energy costs of communicating with AI." *Frontiers in Communication* 10 (2025): 1572947.

More energy, diminishing returns on accuracy

Accuracy of A.I. model ↑



Source: Dauner and Socher, 2025 · Note: A.I. models answered 500 free-response questions · By Harry Stevens/The New York Times

<https://www.nytimes.com/2025/06/19/climate/ai-emissions-chatbot-accuracy.html>

Questions about math tended to generate more emissions

Grams of CO₂ emitted per answer

	Abstract algebra	H.S. math	H.S. world history	International law	Philosophy
Average of 14 language models	1.0	1.0	0.6	0.5	0.4
Results by model					
Deepseek-r1 70B REASONING	4.8	4.0	1.7	1.4	1.2
Cogito 70B REASONING	2.3	2.3	1.7	1.6	1.4
Qwen2.5 72B	1.1	0.8	1.0	0.7	0.5
Cogito 8B REASONING	1.1	1.8	0.3	0.3	0.3
Llama 3.3 70B	0.7	0.7	0.7	0.6	0.4
Llama 3.1 70B	0.6	0.6	0.7	0.6	0.3
Deepseek-r1 7B REASONING	1.1	1.0	0.4	0.2	0.2
Deepseek-r1 8B REASONING	1.1	0.9	0.4	0.2	0.2
Cogito 70B	0.6	0.5	0.7	0.5	0.4
Qwen 72B	0.3	0.5	0.5	0.2	0.2
Llama 3.1 8B	0.1	0.2	0.1	0.1	0.1
Qwen2.5 7B	0.2	0.1	0.1	0.1	0.1
Cogito 8B	0.1	0.2	0.1	0.1	0.1
Qwen 7B	<0.1	0.1	0.1	<0.1	<0.1

See fewer models

Source: Dauner and Socher, 2025 · Note: A.I. models answered 100 free-response questions in each category · By Harry Stevens/The New York Times

Another Example

- **Llama 3.1 8B (8 billion parameters)**
 - ~114 joules per response when accounting for cooling, etc.
 - Equivalent of riding 6 ft on e-bike, running microwave for 0.1 sec
- **Llama 3.1 405B (50x larger)**
 - ~6706 joules per response
 - 400 ft on e-bike, microwave for 8 seconds
- **Takeaway: pick the right model for your task**

Dr. Luccioni said that these kinds of “old school” A.I. tools, including classic search engine functions, have been overlooked as generative models have become more widespread. Most of the time, she said, the average person doesn’t need to use an LLM at all.

“We’re reinventing the wheel,” Dr. Luccioni said. People don’t need to use generative A.I as a calculator, she said. “Use a calculator as a calculator.”

-Sasha Luccioni, the A.I. and climate lead at Hugging Face, quoted in *NYT*