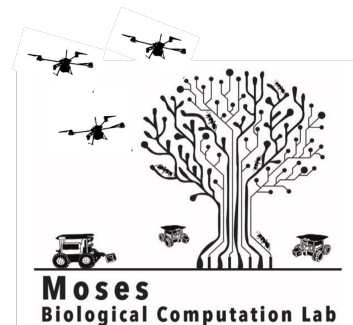


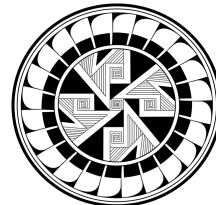
# AI (R)evolution: How did we get here And where are we going?



Sandia INCOSE Enchantment Chapter

Melanie Moses  
UNM Professor of Computer Science  
External Faculty, Santa Fe Institute  
<https://moseslab.cs.unm.edu/>

May 8, 2024



# What I'll talk about today

- My evolving perspective on AI
- How generative AI works (and doesn't)
- Is scale is all we need? (No)
- Overview of AI@UNM & Sandia/LANL partnerships
- The future of AI for New Mexico & the public good

## Southwestern Indian Polytechnic Institute (SIPI) 1<sup>st</sup> place Swarmathon 2018

# My evolving thoughts on AI

- I study bio-inspired robotics and complex systems
- I think of AI as a complex adaptive system that we are integrating into our complex socio-technical system
- This is exciting and scary!
  - AI will help solve scientific problems
  - Current AI encodes biases, promotes surveillance & concentrates power
  - AI codes even better than it writes
  - How will humans co-evolve with AI?



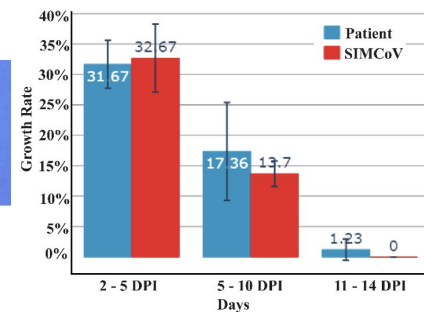
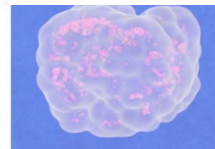
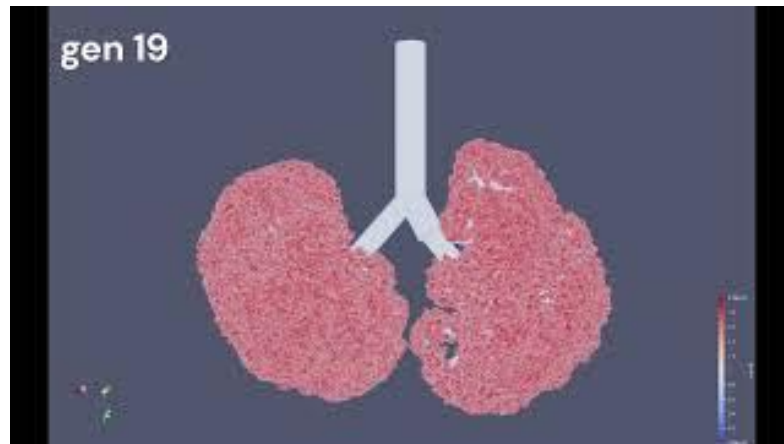
# My evolving thoughts on AI

- I study bio-inspired robotics and complex systems
- I think of AI as a complex adaptive system that we are integrating into our complex socio-technical system
- This is exciting and scary!
  - AI will help solve scientific problems
  - Current AI encodes biases, promotes surveillance & concentrates power
  - AI codes even better than it writes
  - How will humans co-evolve with AI?



# My evolving thoughts on AI

- I study bio-inspired robotics and complex systems
- I think of AI as a complex adaptive system that we are integrating into our complex socio-technical system
- This is exciting and scary!
  - AI will help solve scientific problems
  - Current AI encodes biases, promotes surveillance & concentrates power
  - AI codes even better than it writes
  - How will humans co-evolve with AI?



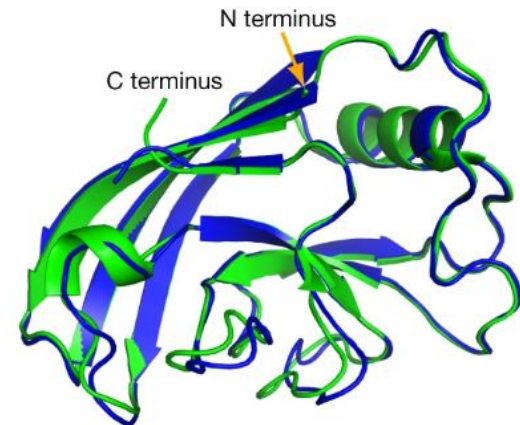
Akil Andrews



Humayra Tasnim

# My evolving thoughts on AI

- I study bio-inspired robotics and complex systems
- I think of AI as a complex adaptive system that we are integrating into our complex socio-technical system
- This is exciting and scary!
  - AI will help solve scientific problems
  - Current AI encodes biases, promotes surveillance & concentrates power
  - AI codes even better than it writes
  - How will humans co-evolve with AI?



Highly accurate protein structure prediction with AlphaFold, *Nature* 2021

AI-assisted science is potentially revolutionary!  
Biomedicine: Antibiotics, therapeutics, diagnoses, vaccines ...

Climate & energy: materials for batteries, biodegradable plastics, fusion(!?)

# My evolving thoughts on AI

- I study bio-inspired robotics and complex systems
- I think of AI as a complex adaptive system that we are integrating into our complex socio-technical system
- This is exciting and scary!
  - AI will help solve scientific problems
  - Current AI encodes biases, promotes surveillance & concentrates power
  - AI codes even better than it writes
  - How will humans co-evolve with AI?



[www.dailymaverick.co.za/article/2021-04-16-this-weekend-were-watching-coded-bias/](http://www.dailymaverick.co.za/article/2021-04-16-this-weekend-were-watching-coded-bias/)

<https://www.codedbias.com/>

States' Automated Systems Are Trapping Citizens in Bureaucratic Nightmares With Their Lives on the Line



# My evolving thoughts on AI

- I study bio-inspired robotics and complex systems
- I think of AI as a complex adaptive system that we are integrating into our complex socio-technical system
- This is exciting and scary!
  - AI will help solve scientific problems
  - Current AI encodes biases, promotes surveillance & concentrates power & threatens jobs, particularly of creatives
  - AI codes even better than it writes
  - How will humans co-evolve with AI?

## Finance worker pays out \$25 million after video call with deepfake 'chief financial officer'



By Heather Chen and Kathleen Magramo, CNN

2 minute read · Published 2:31 AM EST, Sun February 4, 2024



Authorities are increasingly concerned at the damaging potential posed by artificial intelligence technology. boonchai wedmakawand/Moment RF/Getty Images

<https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html>



# My evolving thoughts on AI

- I study bio-inspired robotics and complex systems
- I think of AI as a complex adaptive system that we are integrating into our complex socio-technical system
- This is exciting and scary!
  - AI will help solve scientific problems
  - Current AI encodes biases, promotes surveillance & concentrates power & threatens jobs, particularly of creatives
  - AI codes even better than it writes
  - How will humans co-evolve with AI?

Here's a Python code using 'matplotlib' to display the logistic map:

```
python Copy code  
  
import numpy as np  
import matplotlib.pyplot as plt  
  
def logistic_map(r, x):  
    return r * x * (1 - x)  
  
def plot_logistic_map():  
    r = np.linspace(2.5, 4.0, 10000) # Values of r  
    x = 1e-5 * np.ones_like(r) # Initial condition for x  
  
    iterations = 1000  
    last = 100 # We'll only display the last few iterations to see the steady state  
  
    for i in range(iterations):  
        x = logistic_map(r, x)  
        # Only display the bifurcation diagram for the last few iterations  
        if i >= (iterations - last):  
            plt.plot(r, x, 'k', alpha=0.25)  
  
    plt.xlim(2.5, 4)  
    plt.ylim(-0.1, 1.1)  
    plt.xlabel('r')  
    plt.ylabel('x')  
    plt.title('Bifurcation diagram of the logistic map')  
    plt.show()  
  
plot_logistic_map()
```

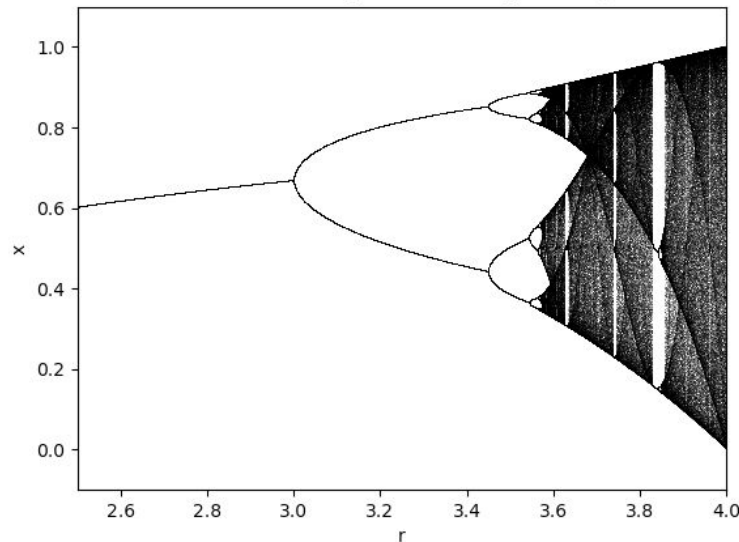
To see the bifurcation diagram:

1. Install the necessary libraries: 'pip install matplotlib'
2. Run the above code.

From ChatGPT  
do my students'  
homework

To Students: Use  
ChatGPT to do your  
homework

Bifurcation diagram of the logistic map



# My evolving thoughts on AI

- I study bio-inspired robotics and complex systems
- I think of AI as a complex adaptive system that we are integrating into our complex socio-technical system
- This is exciting and scary!
  - AI will help solve scientific problems
  - Current AI encodes biases, promotes surveillance & concentrates power
  - AI codes even better than it writes
  - How will humans co-evolve with AI?



Generate DALL-E prompts ...

GPT4: A market scene in 2150 where humans buy tech enhancements for their bodies, while robots buy organic parts to feel more human.

# How does AI work? Traditional algorithms vs. AI

Traditional computer programs are like recipes: step by step instructions to achieve a specified outcome.

## A simple cake recipe

1. Preheat oven to 350 degrees F
  2. Mix 2 cups sugar and 4 sticks butter
  3. Add 8 eggs
  4. Mix in 3 cups flour
- ...
- Bake in greased pan for 50 minutes



# How does AI work? Traditional algorithms vs. AI

Traditional computer programs are like recipes: step by step instructions to achieve a specified outcome.

Traditional algorithms can fail, but we usually understand why

Generative AI is different: we usually can't explain its predictions or its mistakes

## A simple cake recipe

1. Preheat oven to 350 degrees F
2. Mix 2 cups sugar and 4 sticks butter
3. Add 8 eggs
4. Mix in 3 cups flour

...

Bake in greased pan for 50 minutes



Unless you are  
above 5000 feet

Context matters!

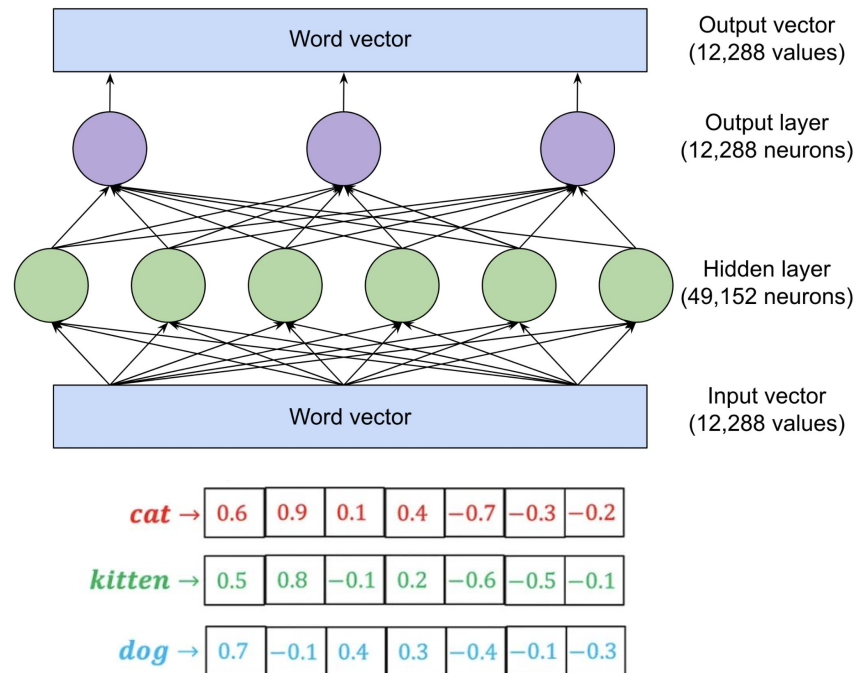
# How does generative AI work?

Large Language Models (LLM, like ChatGPT) use neural networks to find patterns & correlations in data. Those are used to make predictions.

Billions of nodes trained on billions of texts predict what word is likely to come next

Words are vectors (a list of 1000's of numbers) grouped by similarity

Neural networks are pattern matching machines

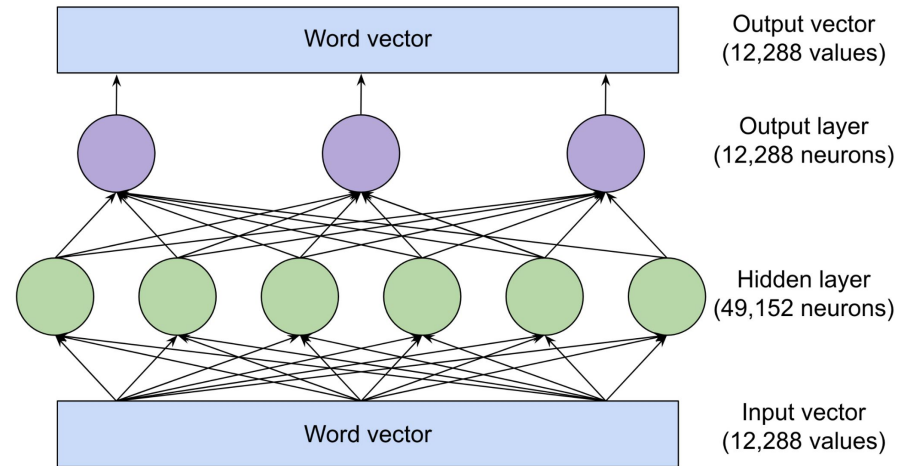
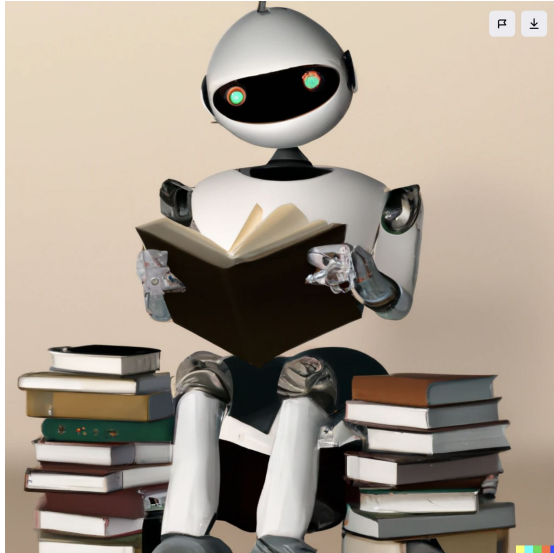


# LLMs learn by predicting what word comes next

Roses are red, violets are xxx

Cherries are red, grapes are xxx

Green means go, red means xxx



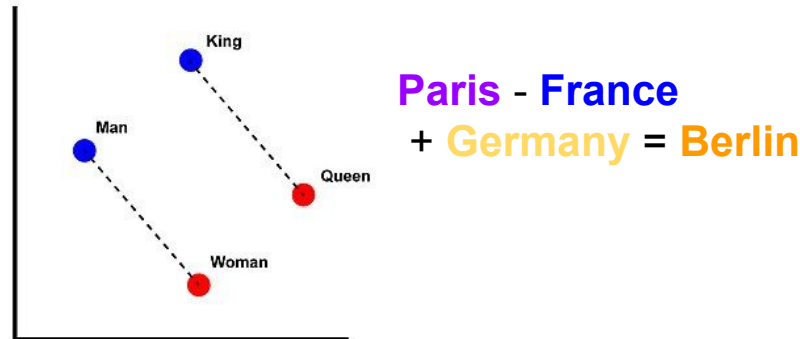
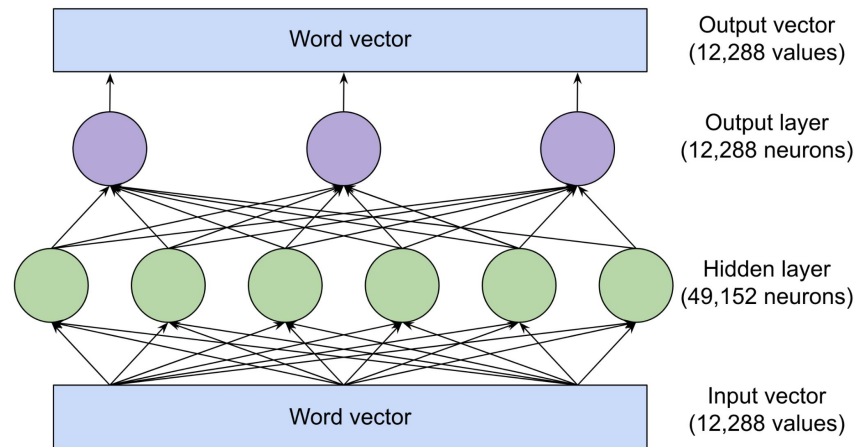
# How does generative AI work?

Large Language Models (LLM, like ChatGPT) use neural networks to find patterns. Those data can be used to make predictions

Billions of nodes trained on billions of texts predict what word is likely to come next

Words are vectors (a list of 1000's of numbers) grouped by similarity

Neural networks are pattern matching machines



# This is tricky!

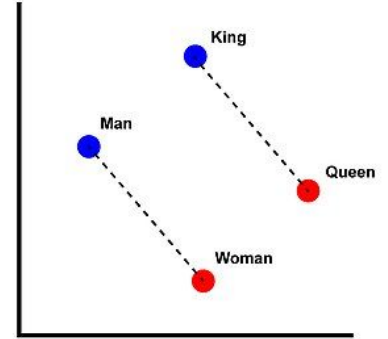
Transformers: “attention mechanisms” disambiguate multiple possible meanings:

Bank: so many meanings!

The doctor examined the patient with her stethoscope (pronouns: confusing!)

Sometimes LLMs guess wrong. The guesses are usually plausible.

It is particularly difficult to recognize plausible falsehoods (BS not hallucinations!)



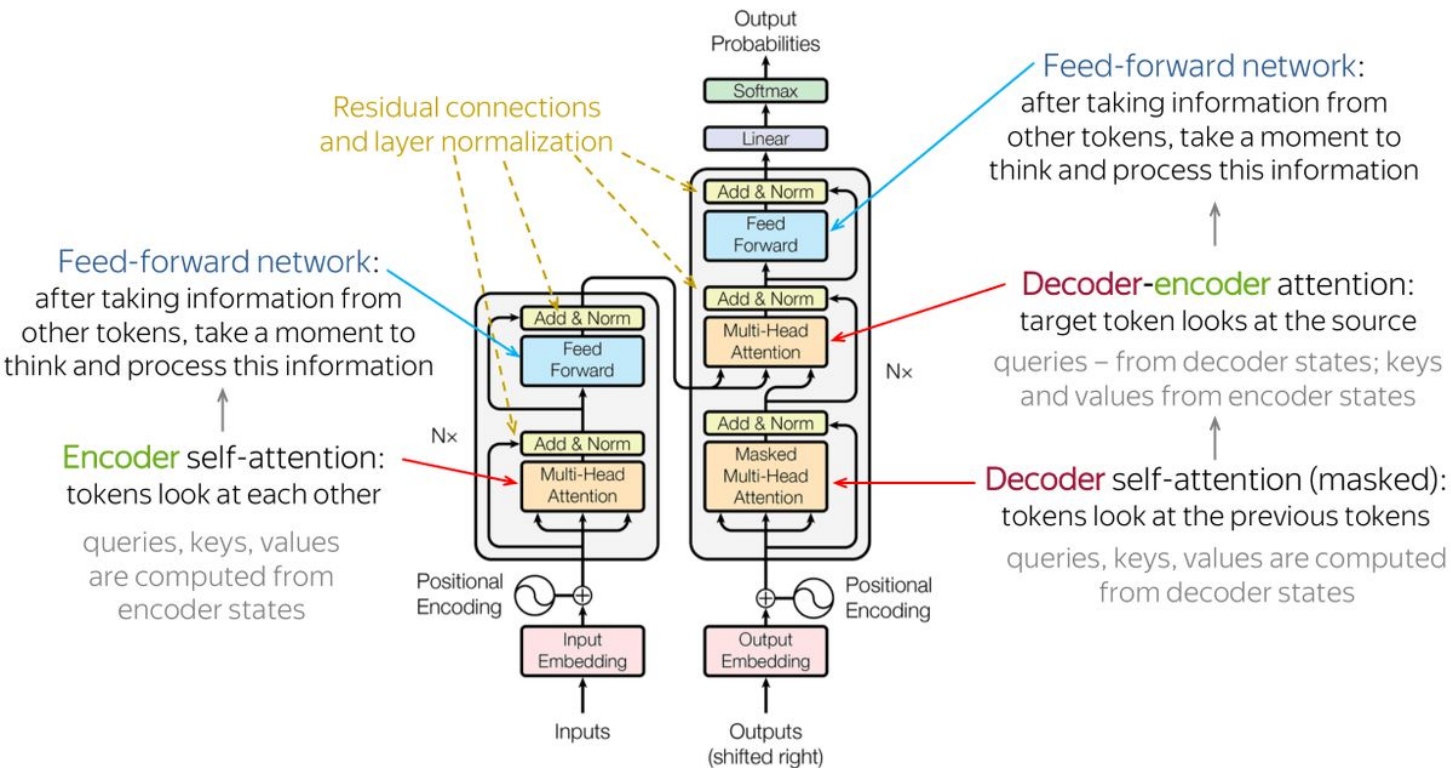
**Paris - France**  
+ **Germany** = **Berlin**

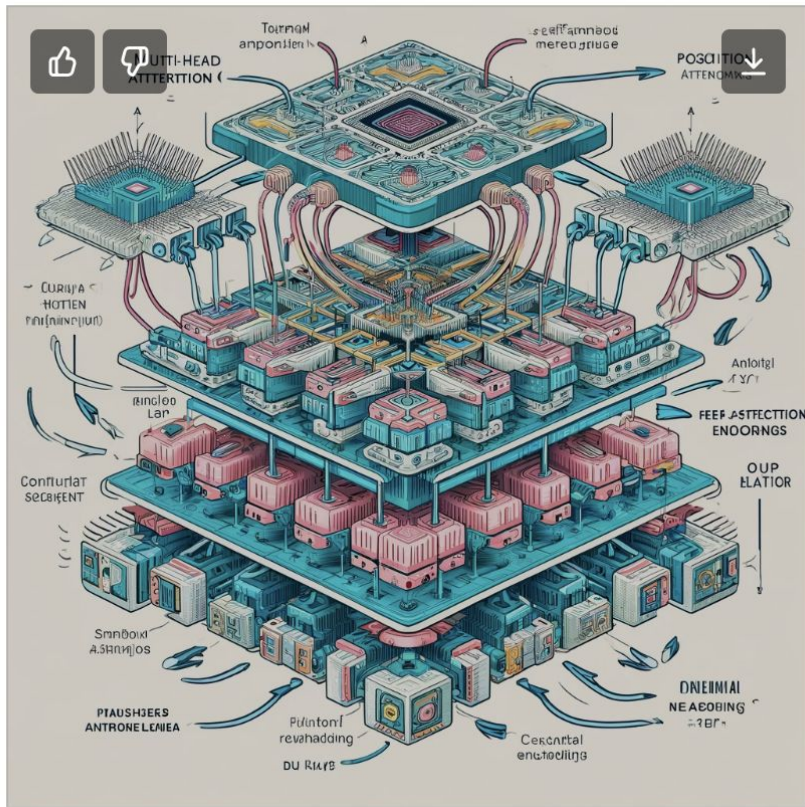




The concepts behind transformers and LLMs - Great Article by [@rasbt](#) - "Understanding Encoder and Decoder LLMs"

# Generative Pre-trained Transformer (GPT)





Transformers, a core architecture in modern Large Language Models (LLMs), utilize a mechanism called self-attention to process sequential data, such as text, in a highly parallelizable manner. This architecture differs significantly from previous sequential models like RNNs or LSTMs, which processed data step-by-step.

#### Core Components

- 1. Self-Attention Mechanism:** At the heart of the transformer is the self-attention mechanism. It allows the model to weigh the relevance of all other tokens in the sequence for each token. This is crucial for understanding context and dependencies, regardless of distance within the sequence.
- 2. Multi-Head Attention:** This is an extension of the attention mechanism where the model runs through multiple attention processes in parallel, allowing it to selectively focus on different parts of the sequence and capture a diverse range of relationships.
- 3. Positional Encoding:** Since transformers do not inherently process sequential data as sequences, positional encodings are added to the input embeddings to provide information about the position of tokens in the sequence.
- 4. Feedforward Neural Networks:** Each layer in a transformer contains a feedforward neural network which applies the same operations to different positions separately and identically. This layer can adaptively change the representation of each token based on the context provided by the attention mechanism.
- 5. Layer Normalization and Residual Connections:** These components help in stabilizing the learning process. Residual connections help in mitigating the vanishing gradient problem by allowing gradients to flow through the network directly.

#### Working Principle

The input sequence is first converted into vectors through embeddings. These embeddings are then processed by the multi-head attention mechanism, which allows the model to focus on different parts of the input sequence for better contextual understanding. The output of this stage is a set of attention-weighted features, which are then passed through a feedforward neural network for further processing. This process repeats across multiple layers, with each layer potentially capturing more complex relationships.

The transformer architecture's ability to handle all tokens of the input simultaneously and its reliance on attention to determine input dependencies makes it exceptionally powerful for tasks involving large sequences of data, leading to significant improvements in tasks like translation, text generation, and more.

Here's a detailed diagram illustrating the transformer architecture used in large language models. This visualization covers key components like multi-head attention, feedforward neural networks, and positional encodings, showing how input tokens are processed to produce contextual embeddings. This should provide a clear visual aid to understand the concepts I described earlier.

Write a haiku about the promise of AI

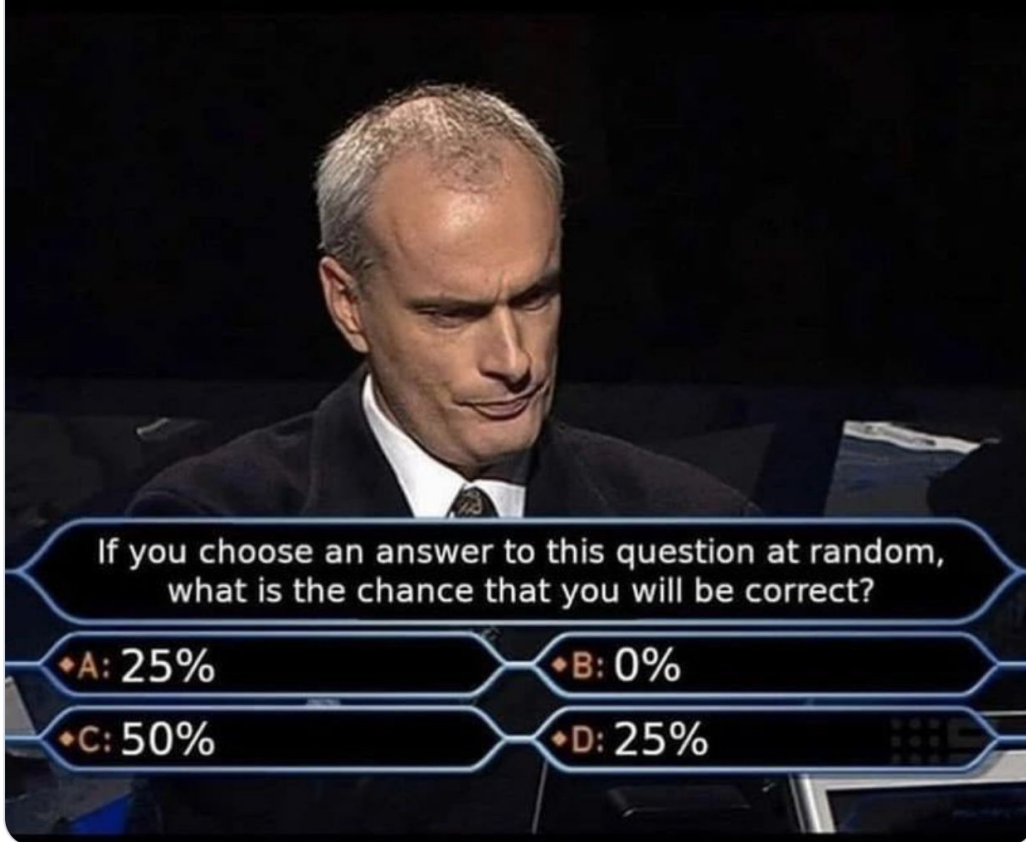
A haiku is a Japanese style of poetry with 5,7,5 syllables in each line

Silent mind, awake,  
AI's promise, boundless skies,  
Humanity's guide.

<https://chat.chatgptdemo.net/>



**Bojan Tunguz**  @tunguz · Mar 3



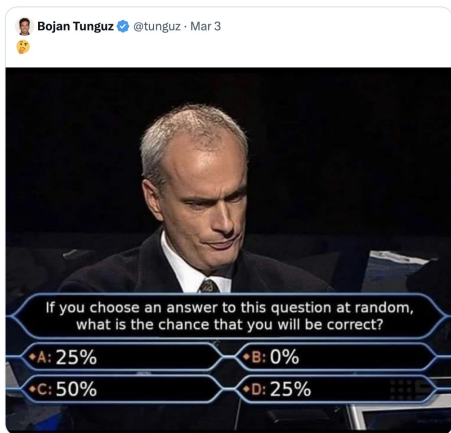


## ChatGPT

This is a classic self-referential question that leads to a paradox if you try to answer it directly, as the percentages provided conflict with each other. Let's break down the options:

1. If we assume A) 25% is correct, then there are two options (A and D) that suggest a 25% chance of being correct. However, this would mean that there's actually a 50% chance of picking one of the two correct answers by chance, which contradicts the initial assumption.
2. If B) 0% is assumed correct, it implies that none of the answers provided would be correct, including B) itself, which creates a contradiction.
3. Assuming C) 50% is correct would mean there are two options out of four that are correct, which is not the case as we see no duplicate percentage values other than what A and D suggest (25%), but this itself leads to a contradiction as explained in point 1.
4. D) 25% faces the same issue as A) 25%, making the assumption of a 25% chance of being correct impossible due to the presence of two answers indicating 25%.

Thus, the question is designed to be paradoxical and doesn't have a straightforward answer within the confines of its own logic. It serves as a playful brain teaser rather than a question with a definitive answer.



# Scale is all you need?

## Moore's Law:

Exponential growth

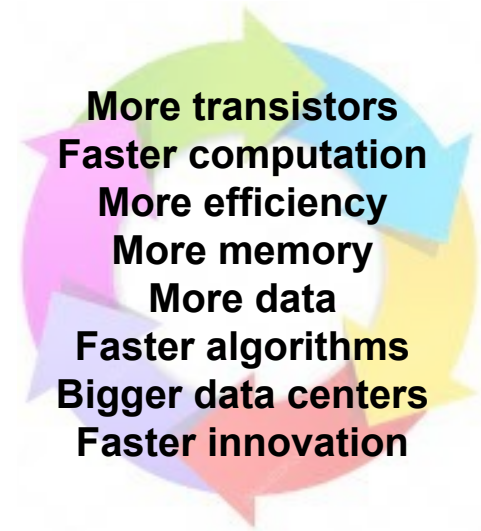
Computation doubles every ~2 years

1970:1000 transistors → 2020: 50 BILLION transistors

Computers are now MILLIONS of times faster

Problems with “Intelligence” at scale

- Climate change<sup>1</sup>
- Little context, many errors<sup>2</sup>
- Large corporations dominate



<sup>1</sup> On the Dangers of Stochastic Parrots, Bender et al 2021

<sup>2</sup> Birhane et al preprint 2023; [arxiv.org/abs/2306.13141](https://arxiv.org/abs/2306.13141)

# How did we get here?

- **Scale**
- A few clever tricks in neural networks
  - Encoders, Transformers
- Human feedback by low paid workers (Kenyan workers paid \$2/day)
- Marketing & economic incentives have concentrated power in a few huge companies

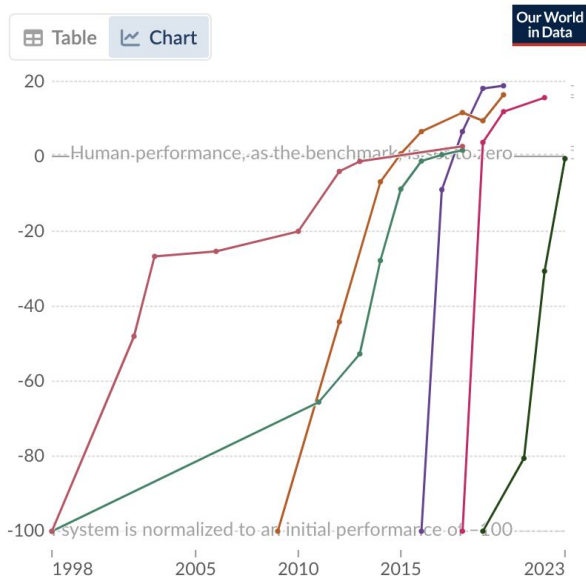
# Where are we going?

- No one knows!
- No exponential lasts forever; all scaling hits constraints
- Feedback comes from LLM users
- Will smaller models become viable - can we democratize AI & use it for good?
- How can we regulate AI & prevent harm?

# Scale is all you need?

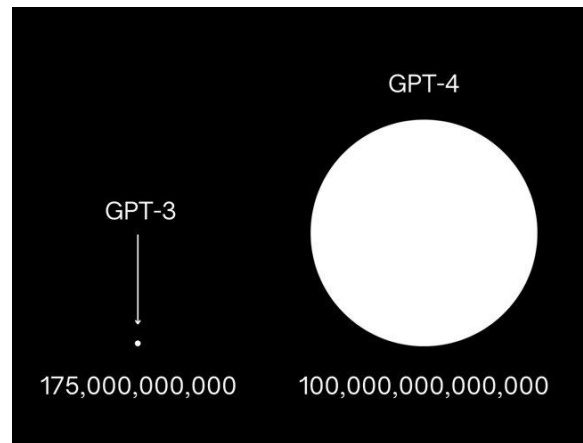
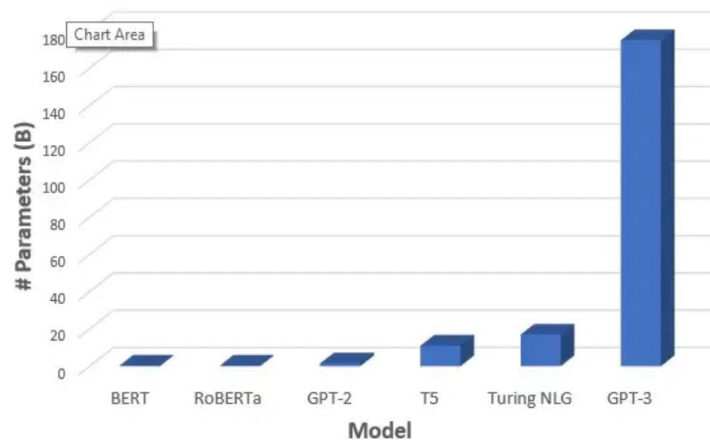
## Test scores of AI systems on various capabilities relative to human performance

Within each domain, the initial performance of the AI is set to -100. Human baseline, set to zero. When the AI's performance crosses the zero line, it's humans.



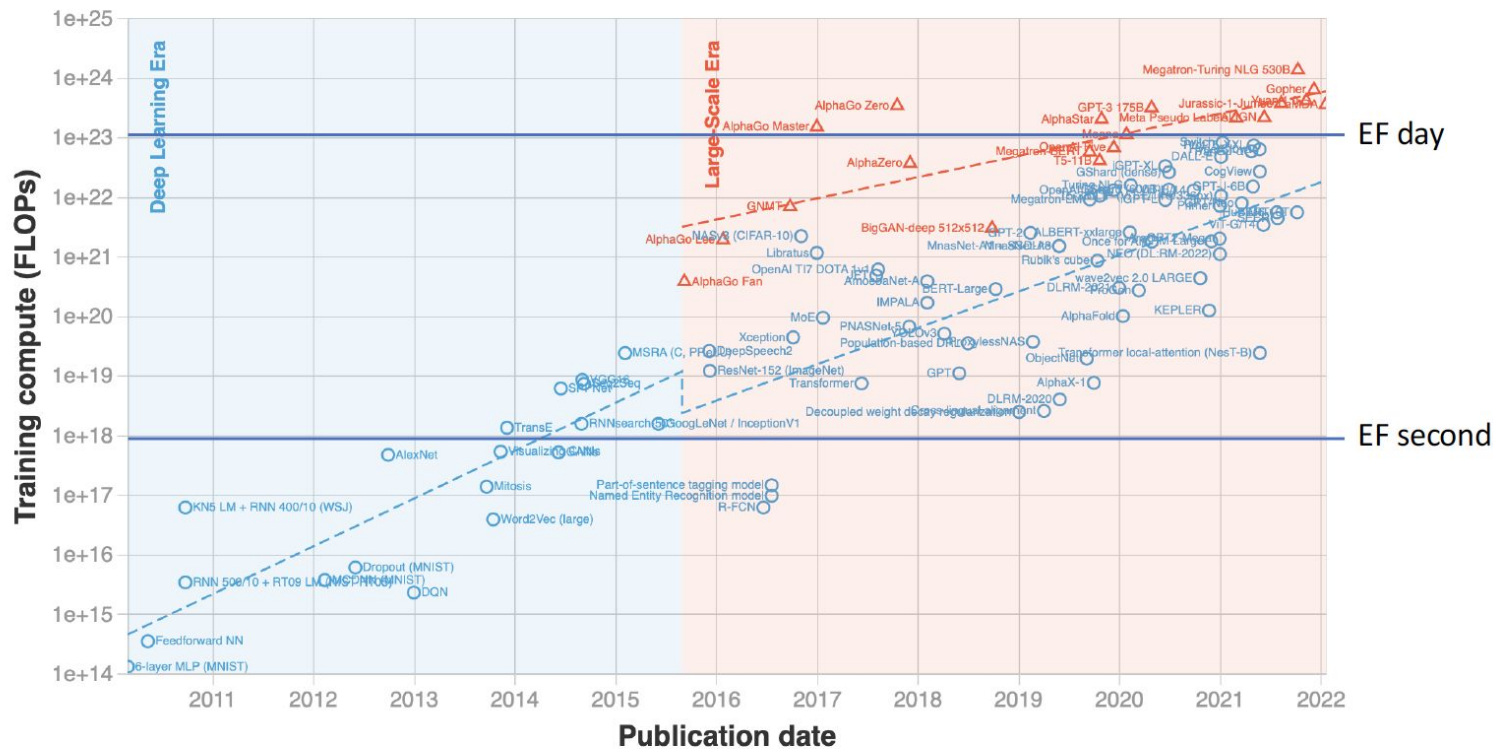
**300 billion trillion calculations to train GPT-3 on 500 billion words.**

It still makes up answers





n = 102



EF day

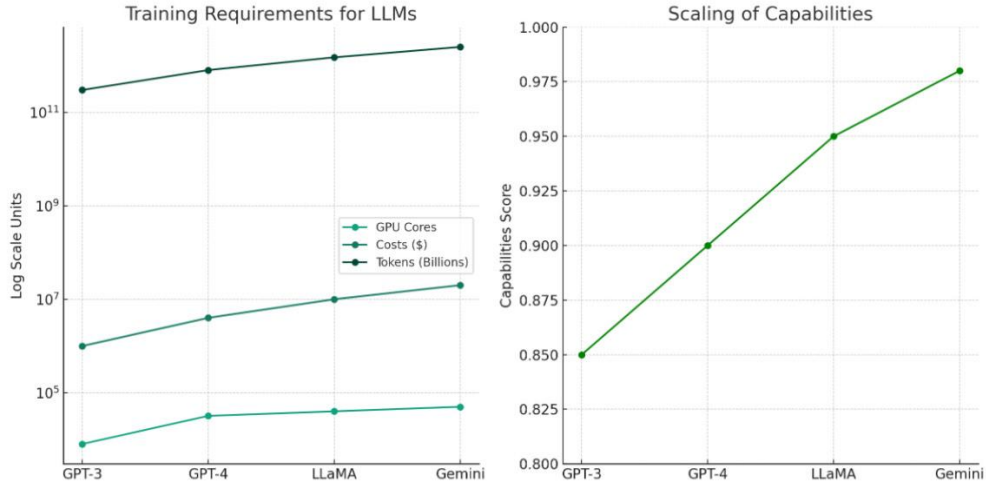
EF day

EF second

EF second

Figure 3: Trends in training compute of  $n=102$  milestone ML systems between 2010 and 2022. Notice the emergence of a possible new trend of large-scale models around 2016. The trend in the remaining models stays the same before and after 2016.

Let's create this visualization. Please hold on while I prepare the figure.



Here's the figure illustrating the scaling of training requirements and capabilities for popular large language models like GPT-3, GPT-4, LLaMA, and Gemini:

WRONG answer from GPT4

# Scale is what they have!



<https://medium.com/@johnadjanohoun/microsoft-and-openai-to-forge-100-billion-supercomputer-for-ais-next-leap-2d5bbc189f6>



ombrulla.com

Saudi Arabia plans  
**\$40 billion**  
investment in AI

Saudi Arabia Goes Big on AI: A \$40 Billion Investment [Visit >](#)

<https://www.linkedin.com/pulse/saudi-arabia-goes-big-ai-40-billion-investment-ombrulla-d3eic>



**Max Tegmark** ✓  
@tegmark

You know that AI is taking off when...Microsoft plans to power data centers with nuclear reactors:

[jobs.careers.microsoft.com/global/en/job/...](https://jobs.careers.microsoft.com/global/en/job/...)

Microsoft, Meta, and Google's parent company, Alphabet, disclosed this week that they had spent more than **\$32 billion combined on data centers and other capital expenses in just the first three months of the year.**

The people who made the biggest fortunes in California's gold rush weren't the miners — they were the people selling the shovels. (NVIDIA)

<https://www.nytimes.com/2024/04/27/technology/ai-big-tech-spending.html?smid=em-share>

# Is scale all you need in information systems?

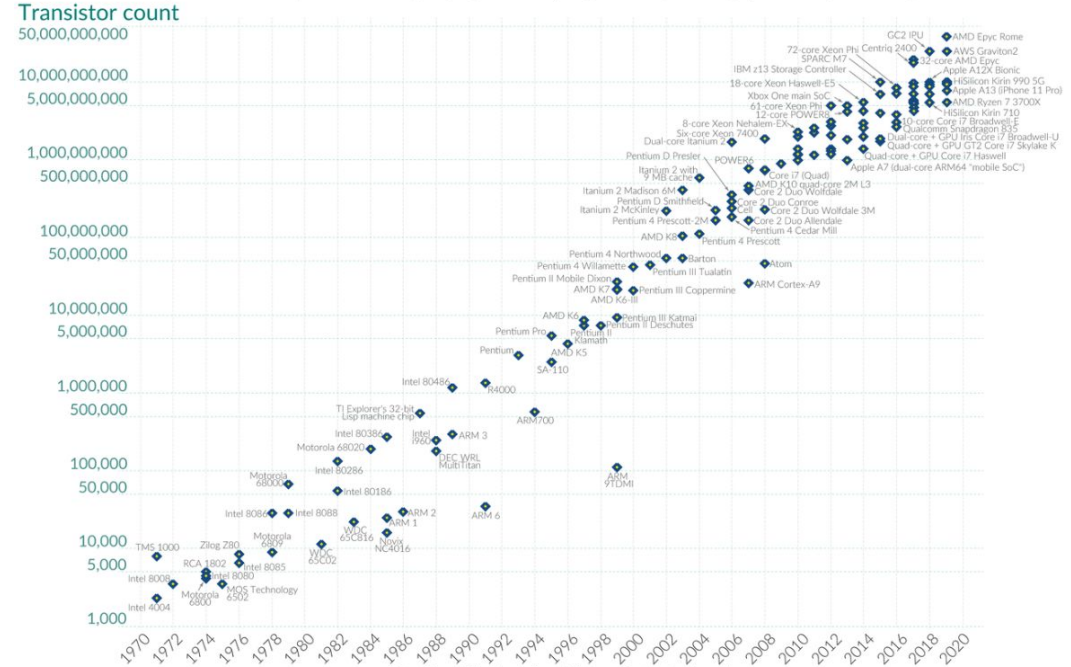
Networks that deliver information rather than energy escape diminishing returns

- Information can be cheaply copied & communicated locally

Moore's Law and the future of AI:

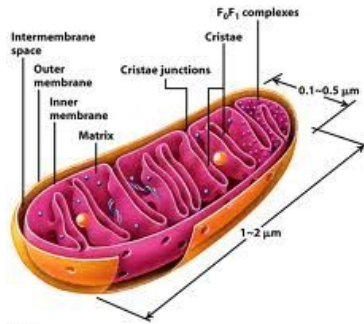
Is scale all you need?

**Moore's Law:** The number of transistors on microchips doubles every two years. Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

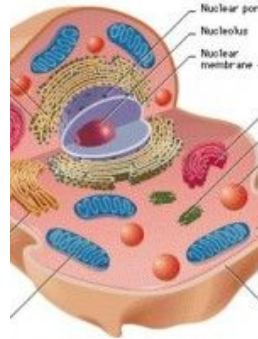


Data source: Wikipedia (wikipedia.org/wiki/Transistor\_count) OurWorldInData.org – Research and data to make progress against the world's largest problems. Licensed under CC-BY by the authors Hannah Ritchie and Max Roser.

# Scaling In Biology: Lessons for Computation



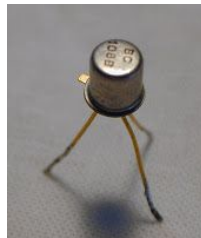
Mitochondrion



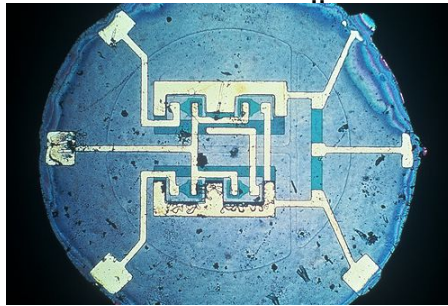
1 to 1000s of Mitochondria per



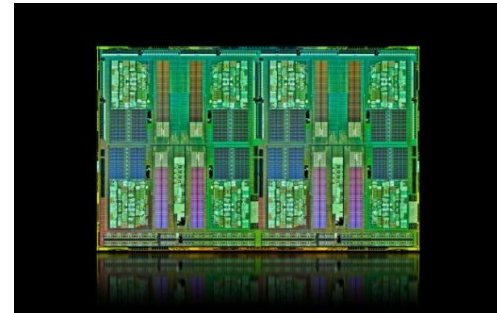
Trillions of mitochondria



Transistor



Integrated circuit



AMD Opteron multi-core chip  
billions of transistors

# Scaling in Biology

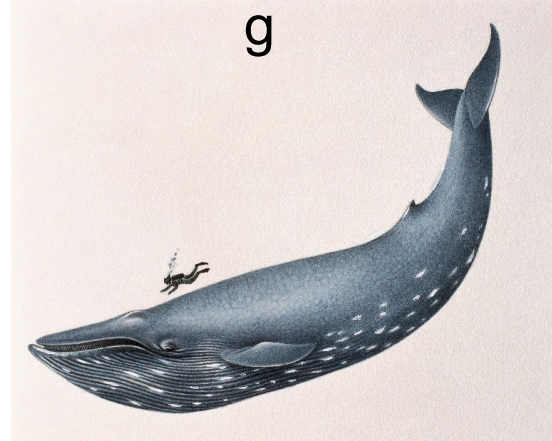
A whale is  
100 000 000 000 000 000 000  
times bigger than an E. coli

$10^{-12}$  g



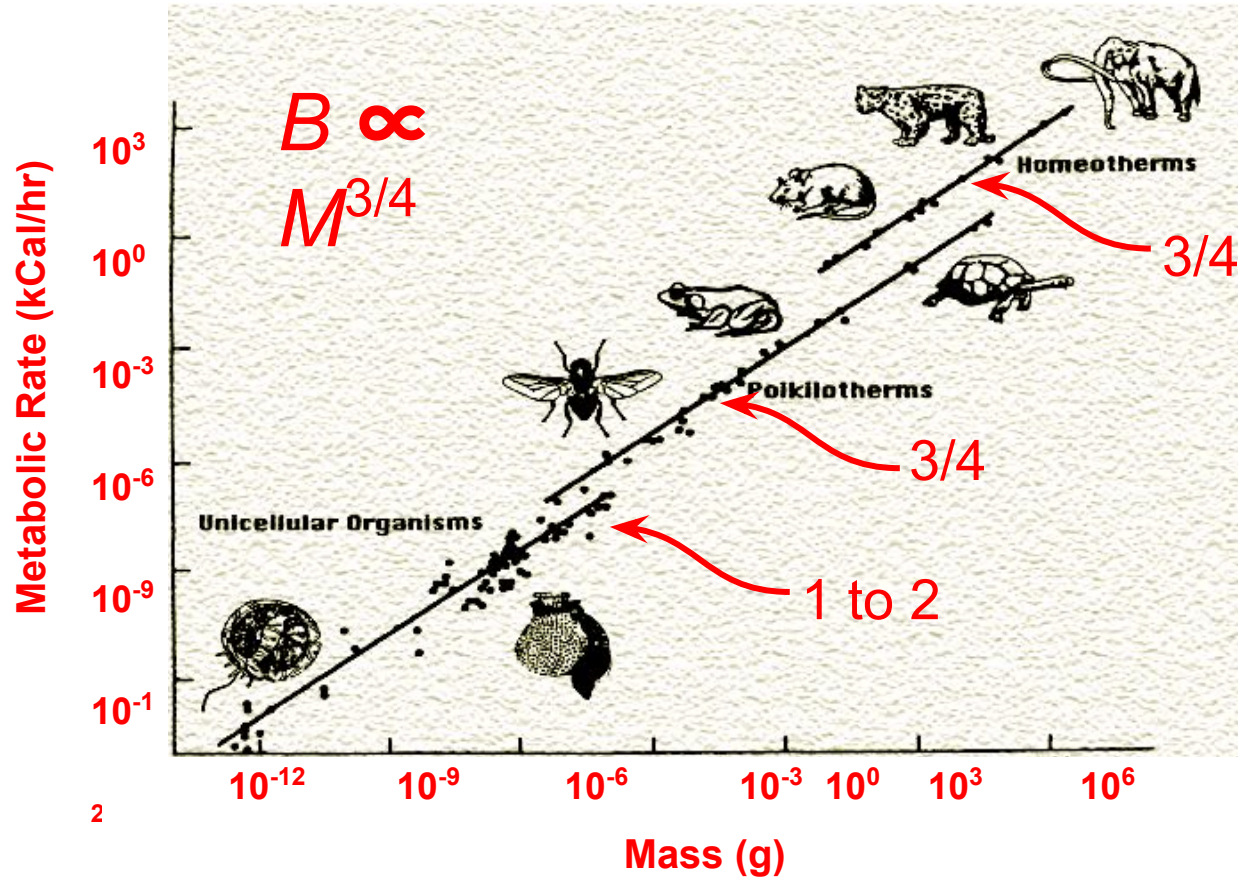
Glaw et al PLoS ONE  
2012

$10^8$   
g



# Metabolic Scaling

A striking universal(?) pattern

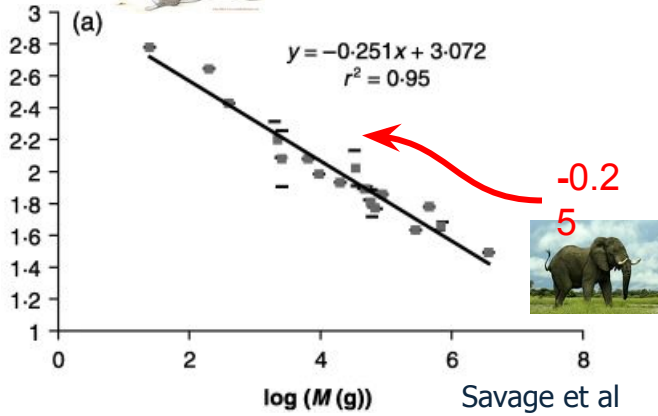




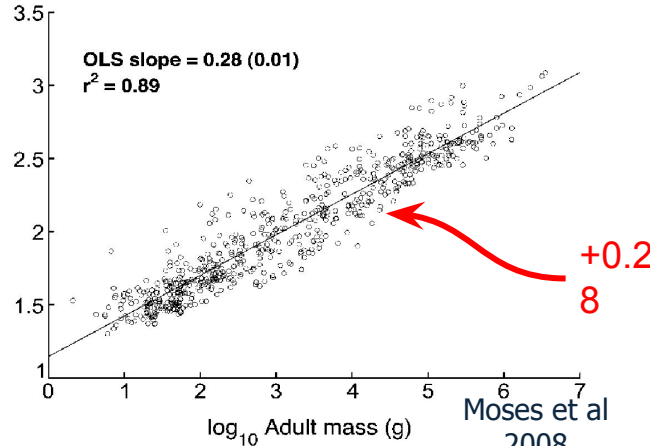
Physiological Rates  $\sim M^{-1/4}$

Physiological Times  $\sim M^{1/4}$

Heart rate in mammals  
(beats/min)



Temperature normalized  
gestation times in fish



Metabolism

rate of energy use  
governs the pace of life

Whole animal:  $B \sim M^{3/4}$

Mass-specific:  $B/M \sim M^{-1/4}$

Mice live fast and die young

- Heart rate, breathing rate  
20x faster
- Gestation times, Lifespans  
20x times shorter
- Invariants
  - 1 Billion heart beats
  - constant fraction to reproduction

# Fractal Networks Generate 3/4 powers

*Centralized* hierarchical, fractal branching

1. Constant branching ratio,

2. Area preserving

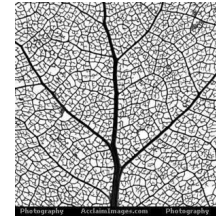
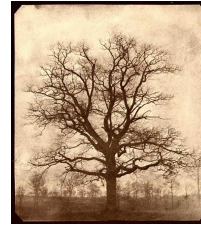
3. Space filling

4. Invariant terminal units

-Capillaries same length, radius & delivery capacity

-Metabolism proportional to # of capillaries

5. Network volume proportional to mass



[West, Brown & Enquist 1997]

# Doubling metabolic rate requires more than doubling network infrastructure

Metabolic rate is proportional to the number of (red) capillaries

To double metabolic rate, double the number of capillaries

Additional (black) network is needed to connect the 2 networks

$$V_{net} = \pi b^k A_{cap} l_{cap} \sum_{i=0}^k b^{i/3}$$

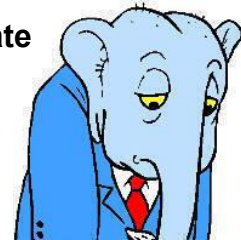
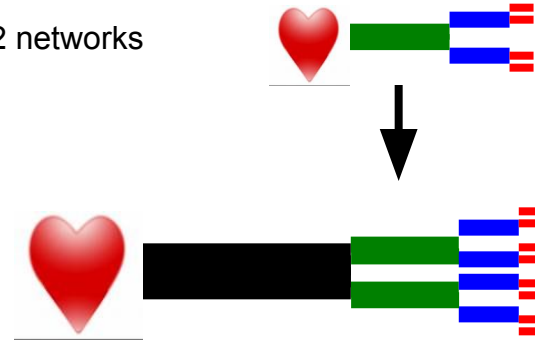
$$V_{net} \propto b^{k^{4/3}}$$

$$V_{net} \propto N_{cap}^{4/3} \propto B^{4/3}$$

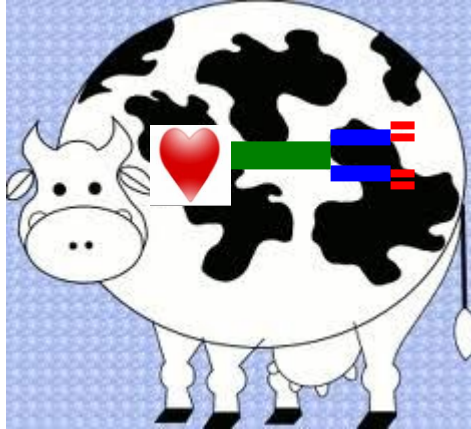
**Diminishing returns: Network size grows faster than network delivery rate**

Assuming Mass and Volume are proportional then

$$B \propto M^{3/4}$$



How do you fit a fractal network into a spherical cow?



# WBE introduced the concept of the service volume

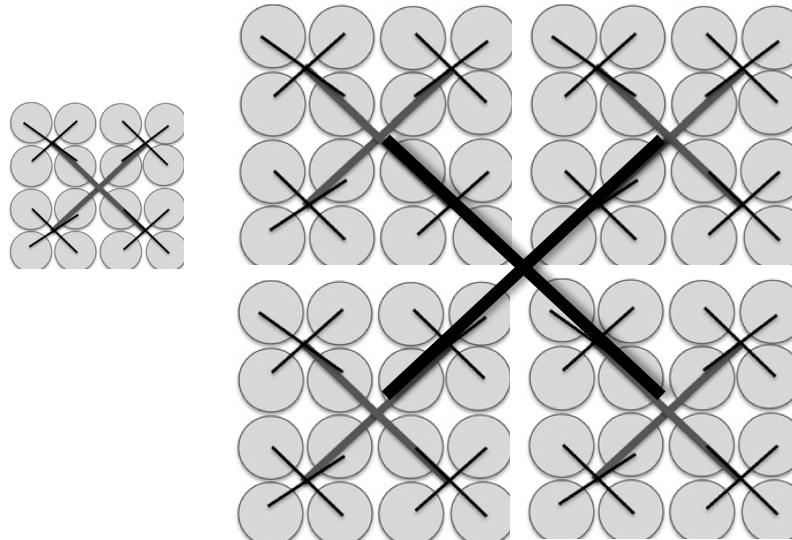
The service volume is the region of tissue supplied by 1 capillary

To accommodate more network infrastructure,  
capillary density is lower in larger animals

$$N_{cap} \propto B \propto V^{3/4}$$

Thus, **the service volume is larger in larger animals**

$$\rho_{cap} \propto N_{cap}/M \propto M^{-1/4}$$



# Microprocessor Scaling: The same, but different

## Network Scaling on computer chips

- a. Information can be copied (cheaply)
- b. Information can be communicated (primarily) locally
- c. Fundamental technology is still improving - process size, materials, packaging, design

[Moses et al 2008, 2016]

# The 1994 DEC Alpha H-tree synchronizes timing

Centralized fractal network delivers a timing signal

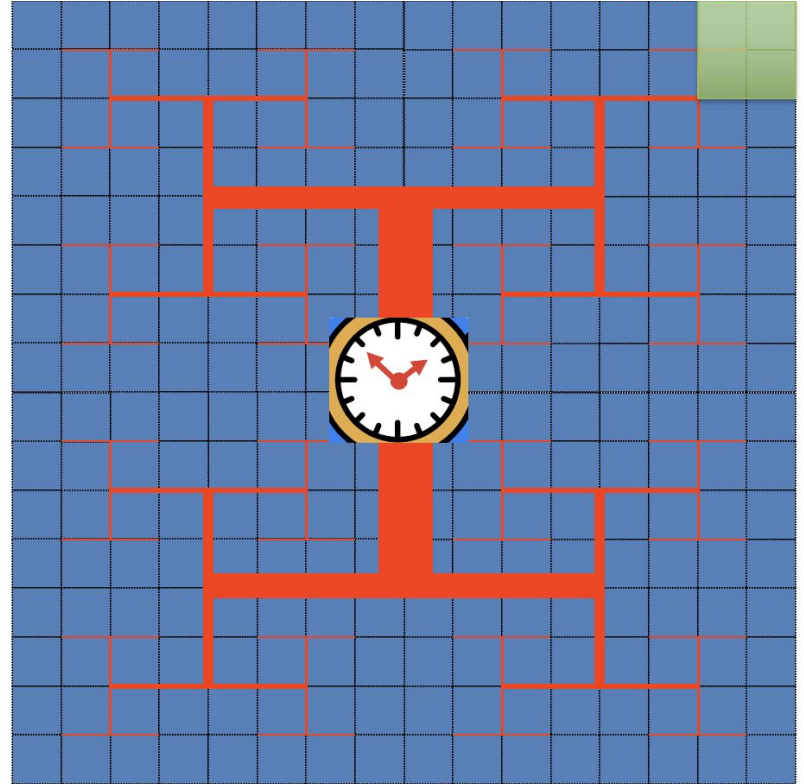
Wire lengths & radii follow 2D scaling predictions:

$$A_{\text{clock-tree}} \sim A_{\text{chip}}^{3/2}$$

Clock speed is limited by the isochronic area

Diminishing Returns:

- Unprecedented speed (300 MHz)
- Clock ultimately consumed 40% of total power of the chip



# The Intel Itanium 2 (2004) eliminates diminishing returns

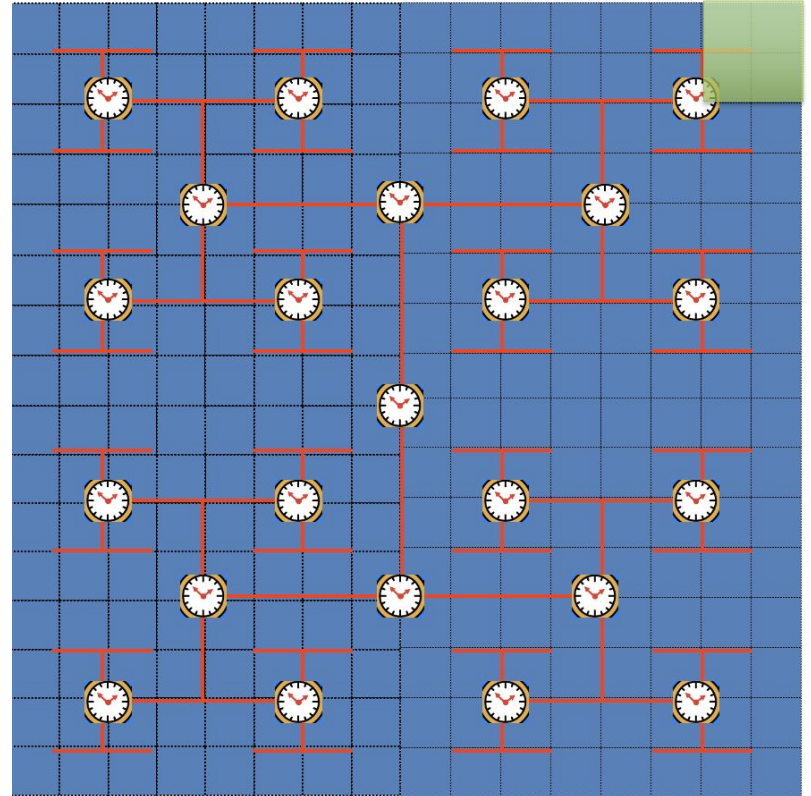
Amplifiers regenerate clock signal at each branch  
No need for thick wires

Network footprint is proportional to chip area:

$$A_{\text{clock-tree}} \sim A_{\text{chip}}^1$$

Clock speed is increased because isochronic area can be reduced

Because information can be copied →  
synchronize more transistors with less power





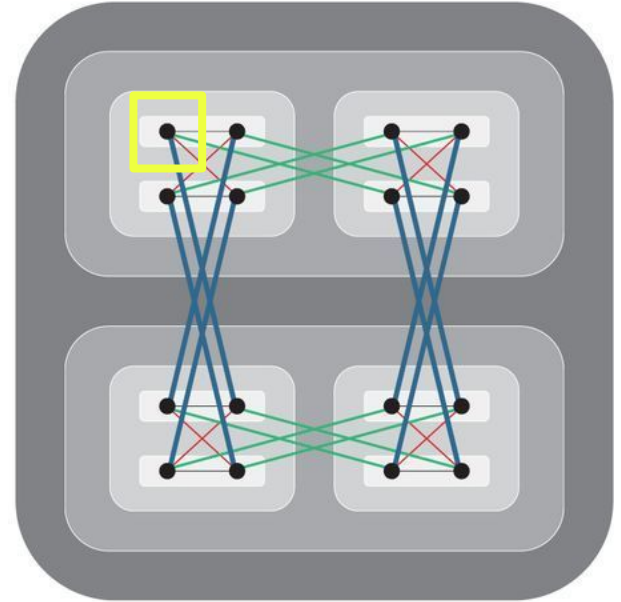
# Rent's Rule: probability of a wire inversely proportional to distance

Rent's Rule was an empirical observation: the probability of a wire crossing a boundary of area  $A$  is  $\propto A^{-1/2}$

Implication: The fraction of all possible connections declines linearly with distance

Again reduces network footprint but allows a completely connected graph

If long distance communication is rare  $\rightarrow$  indefinitely scalable network



# Chip Scaling MINIMIZES Energy x Time

$$E_{\text{sys}} = E_{\text{net}} + E_{\text{nodes}}$$

$$T_{\text{sys}} = \min_{D_r, D_w, D_l} (E_{\text{nodes}} \times T_{\text{sys}})$$

$$E_{\text{net}} \propto N^{1-1/D_l} \sum_{i=0}^H \lambda^{i(1/D_l + D_w - 1)}$$

$$T_{\text{net}} \propto R_0 C_0 \propto \frac{l_0^2}{r_0^2} \propto N^0$$

$$E_{\text{net}} \propto N^{1-1/D_l} \sum_{i=0}^H \lambda^{i(1/D_l + D_w - 1)}$$

$$T_{\text{node}} \propto N^{-1/D_l}$$

$D_l = 2$  (area-filling in 2D chips)

To minimize the  $E_{\text{sys}} \times T_{\text{sys}}$ :

$$D_r = 2$$

$$D_w \leq 2$$

	general	energy - time minimization
--	---------	----------------------------

mammals

$E_{\text{net}}$	$I_0 u_0 N^{2/D_l - 1}$	$N^{1/12}$
$E_{\text{node}}$	$N$	$N$
$T_{\text{net}}$	$u_0^{-1} N^{1-2/D_l}$	$N^0$
$T_{\text{node}}$	$u_0^{-1} N^{1-2/D_l}$	$N^0$
$E_{\text{sys}} \times T_{\text{sys}}$	$I_0 + u_0^{-1} N^{2-2/D_l}$	$N^{1/12} + N$

computers

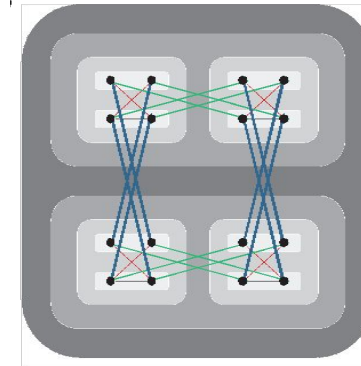
$E_{\text{net}}$	$N^{1-1/D_l}$	$N^{1/2}$
$E_{\text{node}}$	$N^{1-1/D_l}$	$N^{1/2}$
$T_{\text{net}}$	$N^0$	$N^0$
$T_{\text{node}}$	$N^{-1/D_l}$	$N^{-1/2}$
$E_{\text{sys}} \times T_{\text{sys}}$	$N^{1-1/D_l} + N^{1-1/D_l}$	$N^{1/2} + N^{1/2}$

Power

$$P = \frac{E_{\text{sys}}}{T_{\text{sys}}} \propto N^{1/2}$$

Throughput

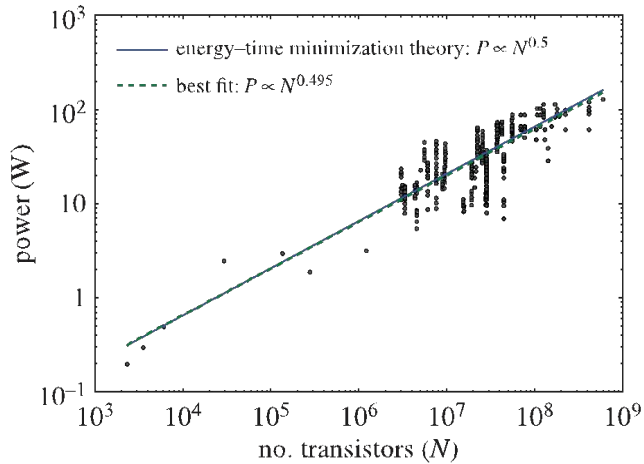
$$Tp \propto \frac{N}{T_{\text{sys}}} \propto N$$



# Scaling Theory for Computation Predicts

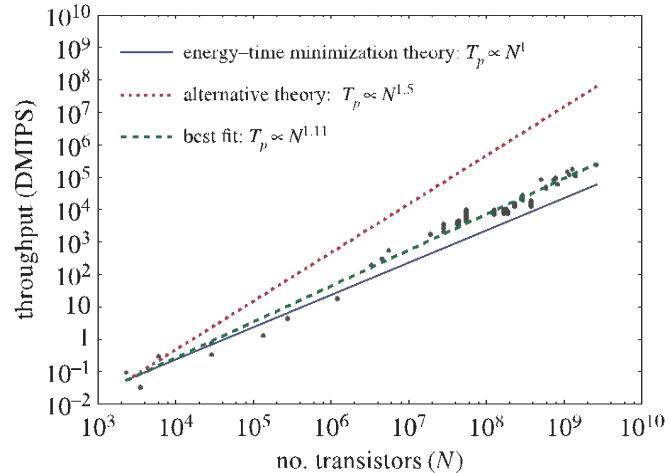
Power

$$P = \frac{E_{\text{sys}}}{T_{\text{sys}}} \propto N^{1/2}$$



Throughput

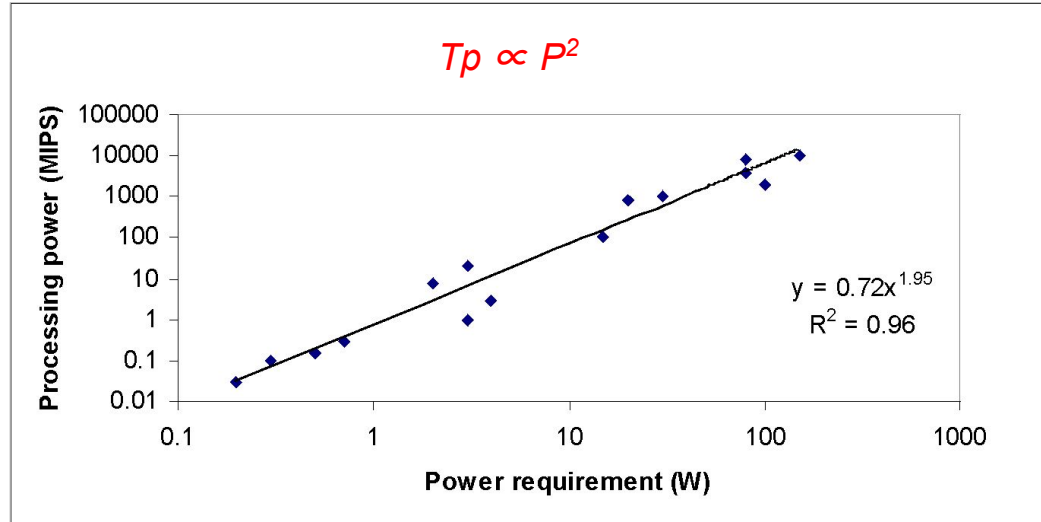
$$Tp \propto \frac{N}{T_{\text{sys}}} \propto N$$



Hegemony of the network

# Power scaling: Increasing returns

Thousand-fold increase in power, Million-fold increase in MIPS



**1970**

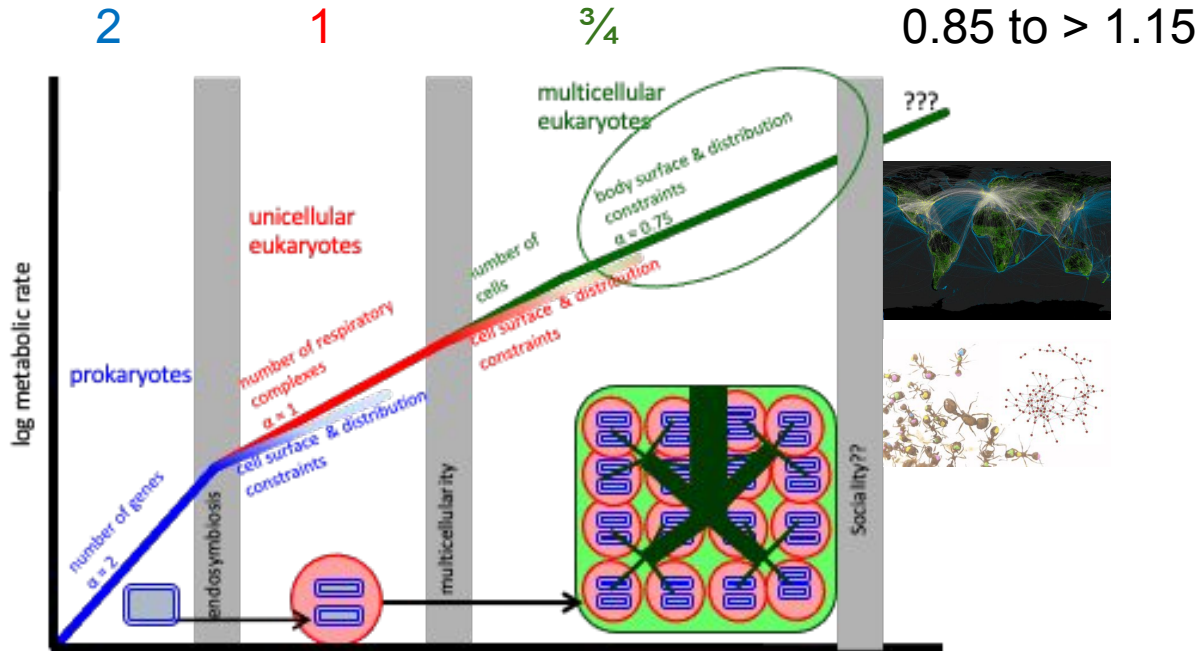
**2005**

In 1970, 100 Watts powered 15 MIPS. In 2005, 6700 MIPS

Transistors perform computations  
Power consumption is dominated by wires

# Scaling constraints necessitate evolutionary transitions

Scaling exponents change with each transition

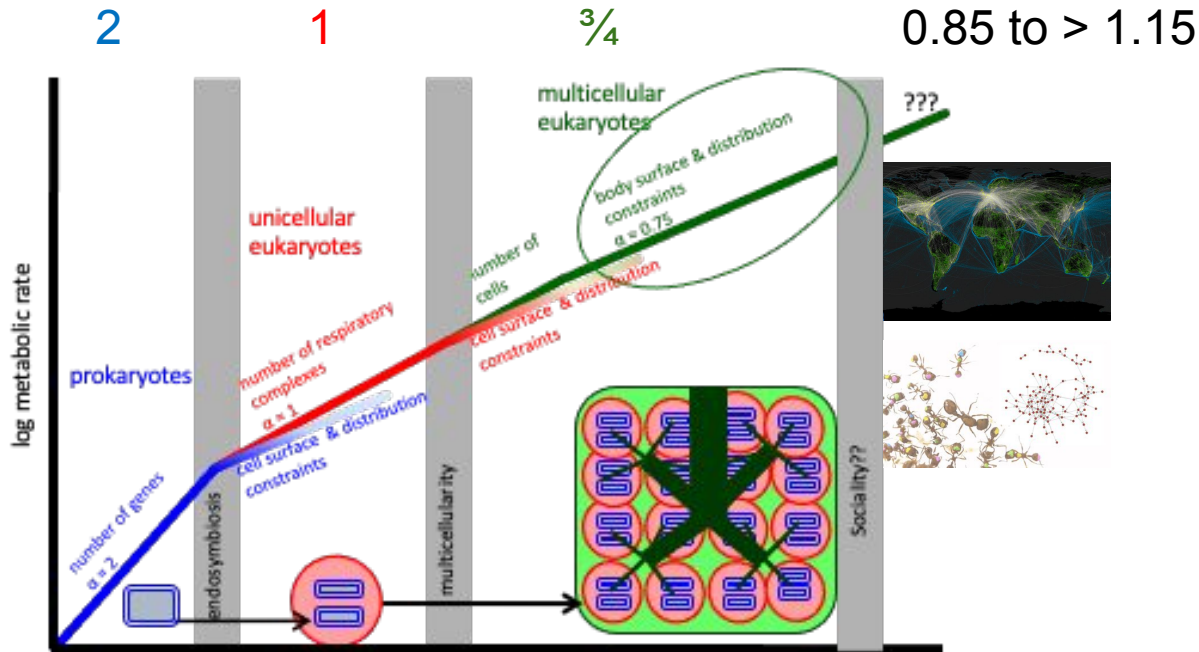


Delong et al PNAS 2010

- Innovations in chip components mimic innovation in the evolution of bacteria
- Single-core chip scaling mimics unicell scaling
- Multi-core chips echo the transition to multicellularity
- Multi-agent computation as a model for scaling in social systems
- Decentralized designs dominate in the transition to sociality

# Scaling constraints necessitate evolutionary transitions

Scaling exponents change with each transition



Delong et al PNAS 2010

What does this mean for AI? More data, more parameters, more training → more intelligence? Maybe, but...

- No exponential lasts forever
- No scaling lasts forever
- Expect Evolutionary transitions

AI will transition to collectives of agents that interact, iterate & evolve



# AI@UNM: Empowering Innovation in New Mexico

>30 UNM Faculty Actively Conducting Research in AI



## Accelerate Discovery with AI

- Supercharge sustainable solutions: smart water allocation; solar energy generation, transmission & storage; forest fire mitigation and recovery; environmental monitoring & remediation
- Secure infrastructure for trustworthy AI in medicine; law & policy; education; & national security
- Advance biomedical innovations, drug discovery & personalized health care for diverse and rural populations
- Develop novel materials, computational fabrication & advanced manufacturing
- Synthesize data to detect methane leaks, volcanic & seismic & natural hazards; improve climate models
- Develop new open source foundational architectures for accessible, applicable and approachable AI
- Develop energy-aware AI based on collective intelligence and quantum technologies

## Foster Collaboration & Catalyze Economic Growth

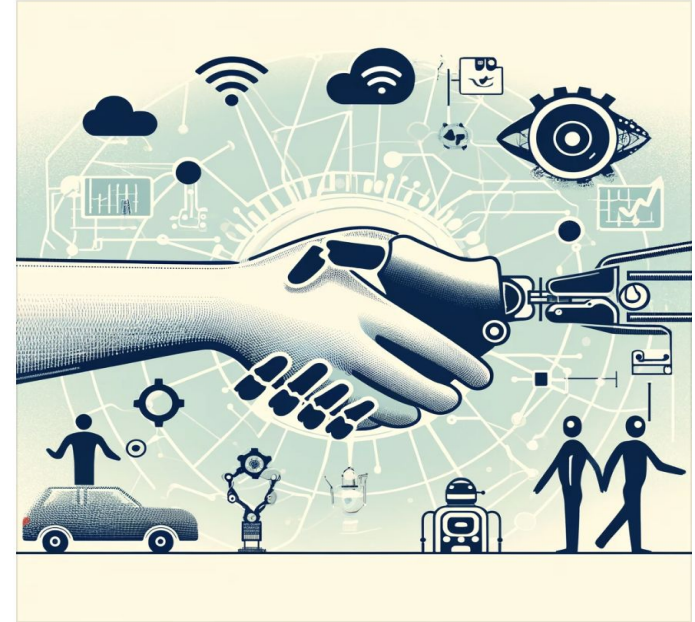
- Crystalize NM as an HPC technology hub
- Prepare NM students to thrive in the AI workforce
- Attract world class faculty
- Spur a startup ecosystem & tech transfer
- Propel NM to the forefront of 3D immersive technologies for education, workforce training, manufacturing, planning & the creative industry
- Engage diverse communities to shape AI for healthcare, law & issues affecting daily lives
- Leverage New Mexico strengths:
  - Expertise in cybersecurity and quantum computing to design next generation AI
  - Power of National Labs, abundance of landscape & natural resources & diversity and creativity of New Mexicans
  - Leadership in Hispanic Serving Institution & EPSCoR networks

# TrustworthyAI @ UNM

## For Trusted Autonomy



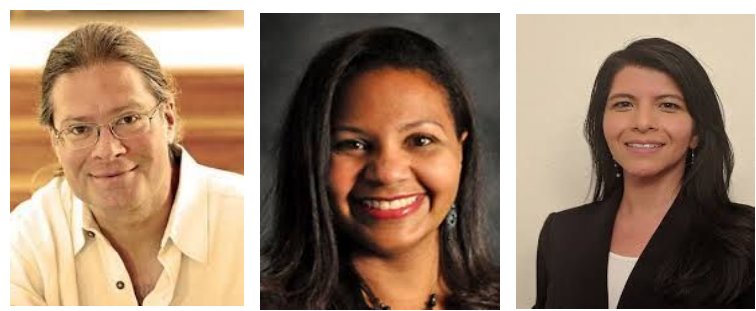
- Develop tools to **detect unsafe behavior** in neural networks
- **Build safe data sets** to train neural networks to produce results within pre-specified safe bounds
- **Develop mitigation strategies** to eliminate dangerous or unexpected predictions & behaviors
- **Iteratively improve neural networks** by removing causes of unsafe behavior
- Build upon our toolbox for neural net verification



Mathematical Foundations



# TrustworthyAI In Law & Medicine @ UNM



- **Analyze accuracy and fairness in risk assessment algorithms used in criminal justice**
- Build on close collaborations with UNM's Institute for Social Research, the Administrative Office of the Courts, the New Mexico Corrections Department, and City and County governments.
- **Identify sources of implicit bias, anomalies and uncertainties in medical records**
- **Provide trustworthy interpretations of medical reports** in close collaboration with stakeholders ensuring that AI is transparent, fair, and inclusive.



Apply trustworthy AI to  
protect rights & improve health



# Interdisciplinary Working Group on Algorithmic Justice



Melanie Moses  
UNM / SFI



CS, Cris Moore  
SFI



Kathy Powers  
Poli Sci, UNM/ SFI



Alfred Mathewson  
Law, UNM



Sonia Rankin Law,  
UNM



Mirta Galesic  
SFI



Josh Garland  
Arizona State



Matthew Fricke  
CS, UNM



Gabe Sanchez  
Poli Sci, UNM



Tina Eliassi-Rad  
CS, NEU /SFI



Mahzarin Banaji  
Psych, Harvard/SFI



Trilce Estrada  
CS, UNM



Nadiyah Humber  
Law, UConn

[nature](#) > [nature reviews electrical engineering](#) > [comment](#) > article

Comment | Published: 12 April 2024

# Medical artificial intelligence should do no harm

[Melanie E. Moses](#)  & [Sonia M. Gipson Rankin](#)

[Nature Reviews Electrical Engineering](#) (2024) | [Cite this article](#)

40 Accesses | 4 Altmetric | [Metrics](#)

**Bias and distrust in medicine have been perpetuated by the misuse of medical equations, algorithms and devices. Artificial intelligence (AI) can exacerbate these problems. However, AI also has potential to detect, mitigate and remedy the harmful effects of bias to build trust and improve healthcare for everyone.**

# AI @UNM

---

## **Presenters**

### **Kent Kiehl**

Professor

Department of Psychology

Presentation Title: "Machine learning of brain science in legal contexts"

### **Sarah Dreier**

Assistant Professor

Department of Political Science

Presentation Title: "AI and text quality: Identifying sources of training biases and imperfect classification outcomes"

### **Jessica Feezell**

Associate Professor

Department of Political Science

Presentation Title: "Algorithm-driven News and Political Polarization in the U.S."

### **Cristyn Elder**

Associate Professor

Department of English Language & Literature

### **Iain Thompson**

Professor

Department of Philosophy

Presentation Title: "What Is Called Thinking in the Age of AI?"

### **Brandon Schmandt**

Professor

Department of Earth & Planetary Sciences

Presentation Title: "Classification of seismic signals with machine learning: applications to environmental change, earthquakes in New Mexico, and national security"

### **Joseph Galewsky**

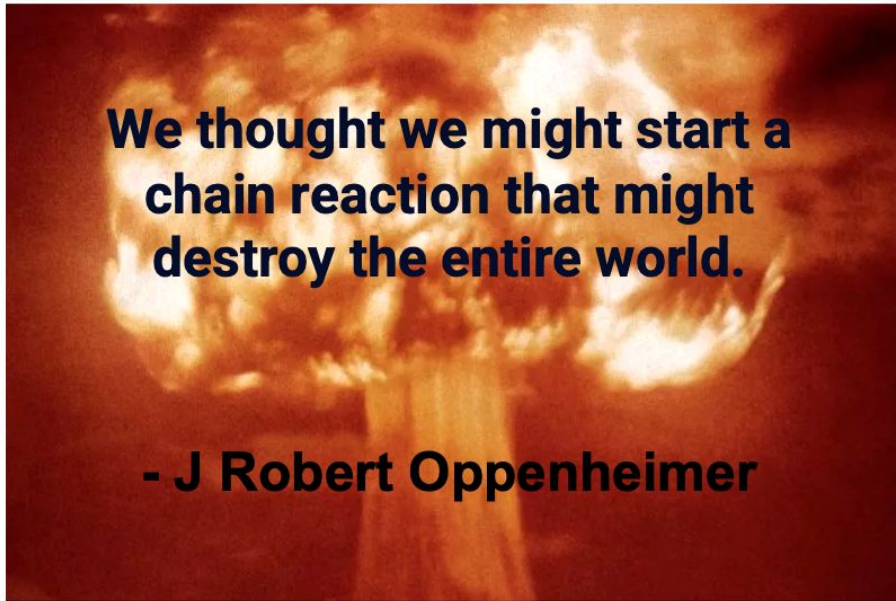
Professor

Department of Earth & Planetary Sciences

Presentation Title: "Cloudy with a Chance of Bits: Using AI to improve our understanding of water vapor's role in the climate system"

# New Mexico AI Consortium

- Los Alamos, Sandia, UNM, NMSU, NMC and others are forming the NMAIC
- Goals:
  - Establish long-term collaboration in AI & ML & HPC
  - Position NM to be a national/international leader in AI for science, natl security & public good
  - Develop a thriving NM research and innovation ecosystem and economy
  - Advance fundamental and applied research in AI
  - Engage & support the diverse communities of New Mexico in shaping an AI-enhanced future
- Please contact me [melaniem@unm.edu](mailto:melaniem@unm.edu) about ongoing or potential collaborations within the labs and NM Universities and colleges.
- Look for an announcement later this summer



Atomic bomb Trinity test.

UNIVERSAL HISTORY ARCHIVE/UIG VIA GETTY IMAGES



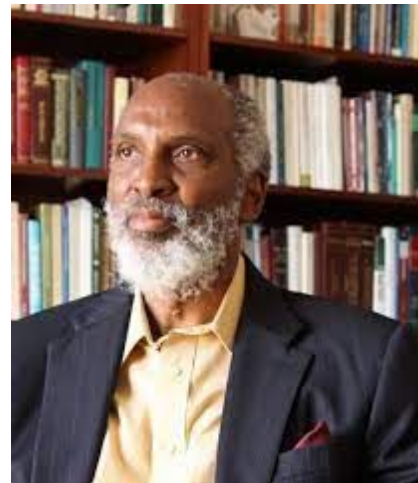


John Powell  
UC Berkeley Law Professor

Kennedy told us that we ALL had to contribute for America to reach the moon

**AI is trained on data from  
(almost) all of humanity**

**All of humanity needs to be engaged  
to shape the future of AI**



# Resources

## How AI works

<https://arstechnica.com/science/2023/07/a-jargon-free-explanation-of-how-ai-large-language-models-work/>

## Politico: legislation for transparency and audits

<https://www.politico.com/news/magazine/2024/04/30/ai-legislation-states-mistake-00155006>

## Complexity & AI

<https://www.newamerica.org/planetary-politics/policy-papers/how-complexity-thinking-can-help-the-world-navigate-ai/>

## Moses & Gipson Rankin article on Trustworthy AI in Medicine

<https://www.nature.com/articles/s44287-024-00049-2>

## AI and Algorithmic Justice Resources @UNM & SFI

<https://airesources.unm.edu/students.html>

<https://algorithmicjustice.cs.unm.edu/index.html>

<https://www.santafe.edu/research/projects/algorithmic-justice>

[melaniem@unm.edu](mailto:melaniem@unm.edu) <https://moseslab.cs.unm.edu>



## Questions?



# Even deep fakes have positive uses

## An AI Salvador Dalí will answer any question when called on his famous 'lobster phone'

UPDATED APRIL 21, 2024 · 1:08 PM ET

 Chloe Veltman



Ask Dalí at the Dalí Museum in St. Petersburg, Fla., allows visitors talk to the famous surrealist artist via an AI-generated version of his voice.

Martin Pagh Ludvigsen/Goodby Silverstein & Partners

<https://www.npr.org/2024/04/21/1246117954/salvador-dali-artificial-intelligence-lobster-telephone-museum>

