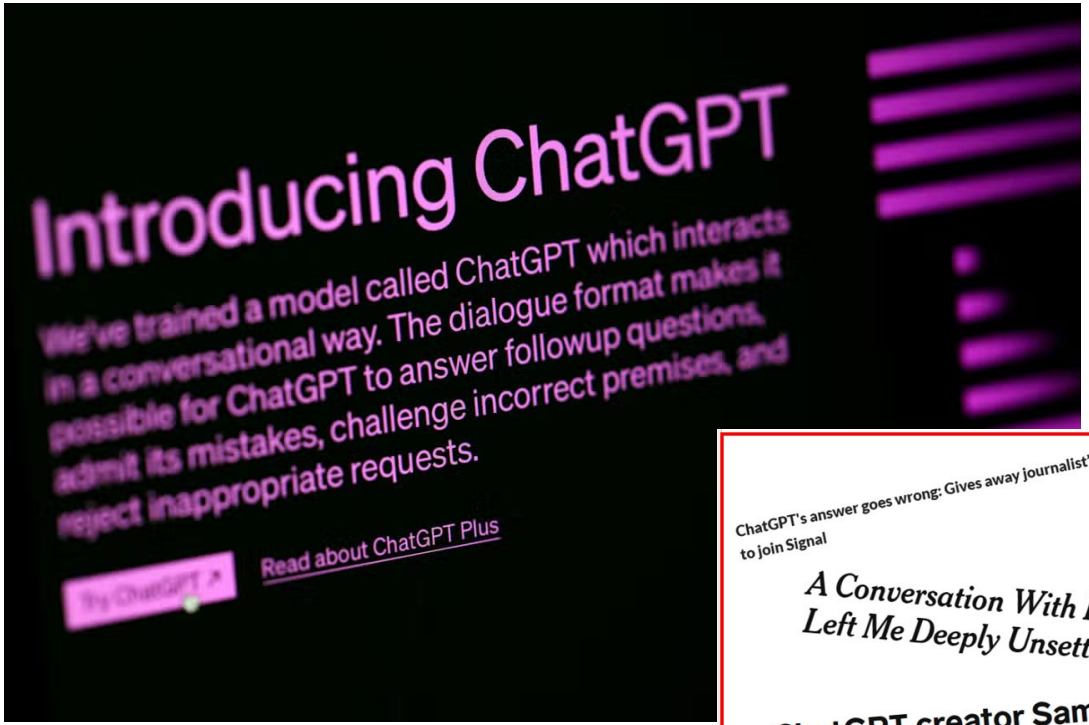
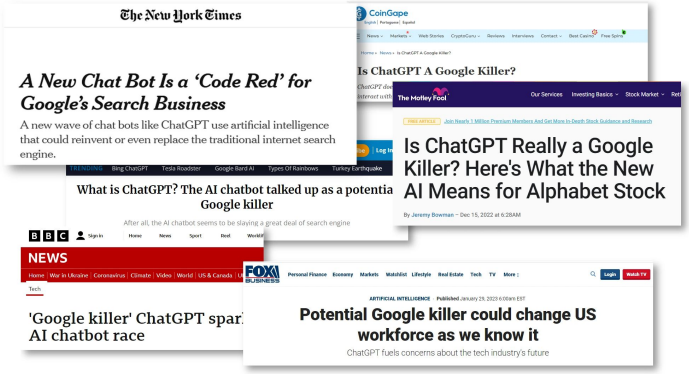
AI: Past, Present, Future

Gabriel Simmons

OLLI Lectures Fall 2024

ChatGPT - Nov 30 2022





2022-2024

The Google engineer who thinks the company's AI has come to life

AI ethicists warned Google not to impersonate humans. Now one of Google's own thinks there's a ghost in the machine.

13 min 3363

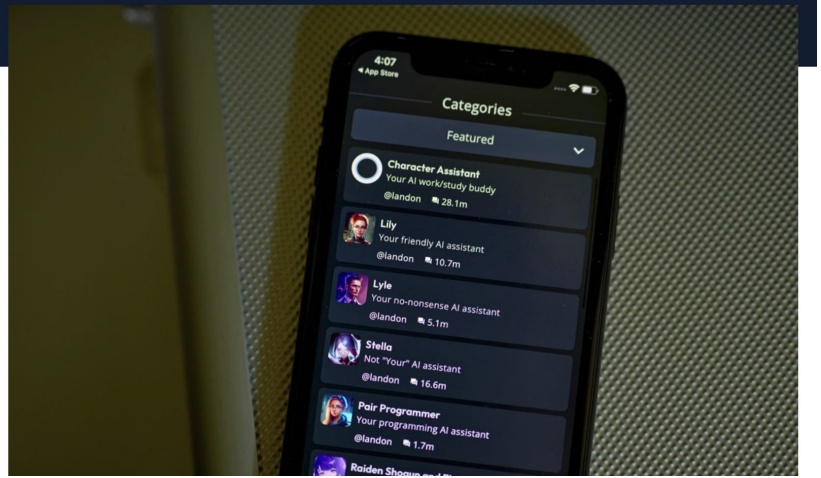
Ilya Sutskever ✓
@ilyasut

it may be that today's large neural networks are slightly conscious

3:27 PM · Feb 9, 2022

Lawsuit claims Character.AI is responsible for teen's suicide

Megan Garcia says the company's chatbots encouraged her 14-year-old son, Sewell Setzer, to take his own life, according to the lawsuit.



2024

AI for images, video, text, music, …

arms race fears

Microsoft, OpenAI plan $100 billion data-center project, media report says

By Reuters

March 29, 2024 2:14 PM PDT · Updated 7 months ago

THE SHIFT

How ChatGPT Kicked Off an A.I. Arms Race

Even inside the company, the chatbot's popularity has come as something of a shock.

OpenAI

REUTERS/Dado Ruvic/Illustration/File Photo Purchase Licensing Rights

Nuclear-Powered AI: Big Tech's Bold Solution or a Pipedream?

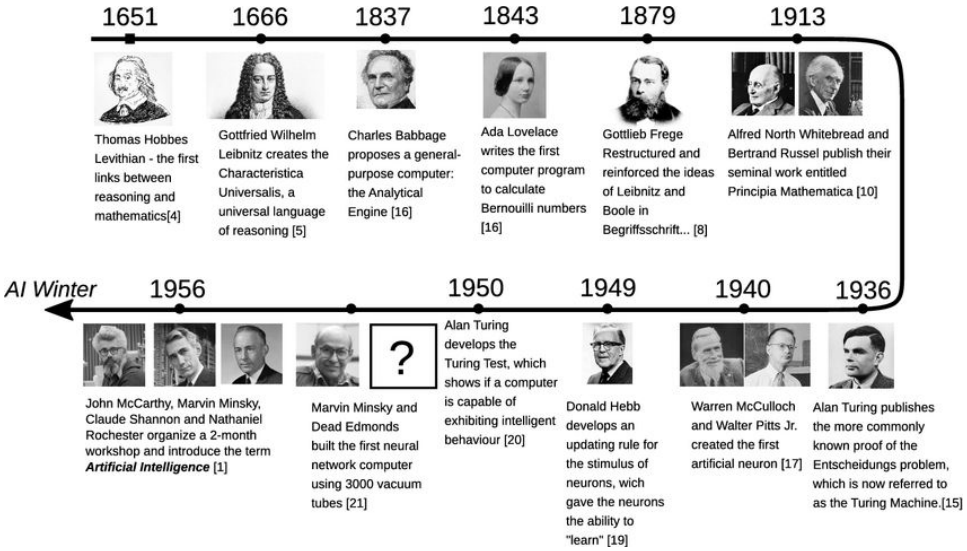
Amazon, Google and Microsoft are investing billions in nuclear power, but the projects are years away and rely on unproven technology

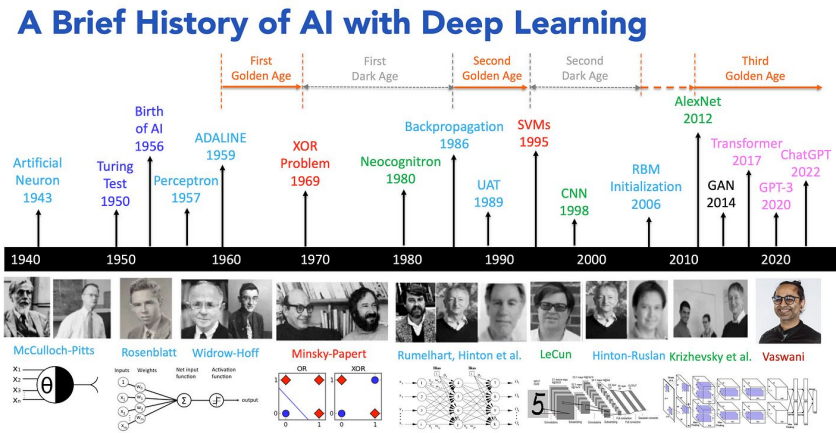
massive investment talks

AI is an exciting new frontier…

… with a long history

How did we get here? (today)





Journey of Artificial Intelligence

April 2022 · Wireless Personal Communications 123(1)

DOI: 10.1007/s11277-021-09288-0

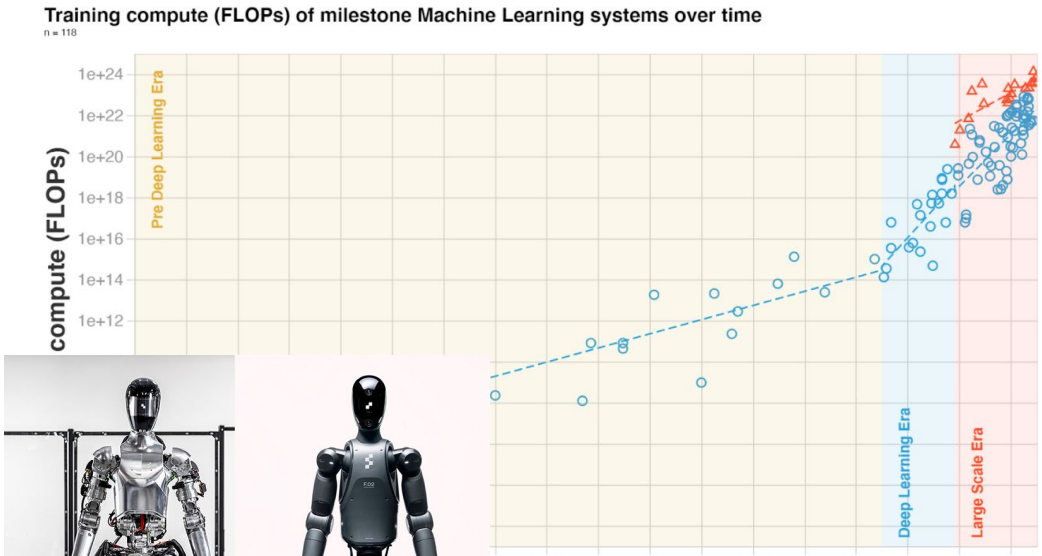
Martijn Kuipers · Ramjee Prasad

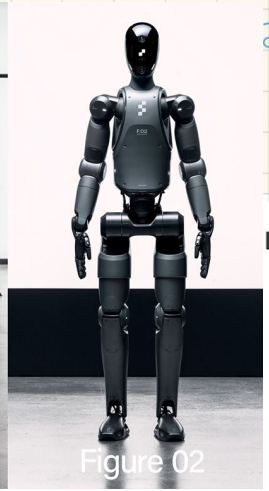
A Brief History of AI with Deep Learning

LM Po · Follow

23 min read · Aug 31, 2024

Where do we go next? (next week)





What is AI?

The problem with artificial intelligence? It's neither artificial nor intelligent

Evgeny Morozov

Let's retire this hackneyed term: while ChatGPT is good at pattern-matching, the human mind does so much more

"the field of study that gives computers the ability to learn without being explicitly programmed - Arthur Samuel, 1959"[1]

Updated: 16 August 2024
Contributors: Cole Stryker, Eda Kavlakoglu

(IBM)

What is AI?

Artificial intelligence (AI) is technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy.

The science and engineering of making intelligent machines - John McCarthy, 1955

Stop Calling Everything AI, Machine-Learning Pioneer Says ›Michael I. Jordan explains why today's artificial-intelligence systems aren't actually intelligent

BY KATHY PRETZ | 31 MAR 2021 | 6 MIN READ |

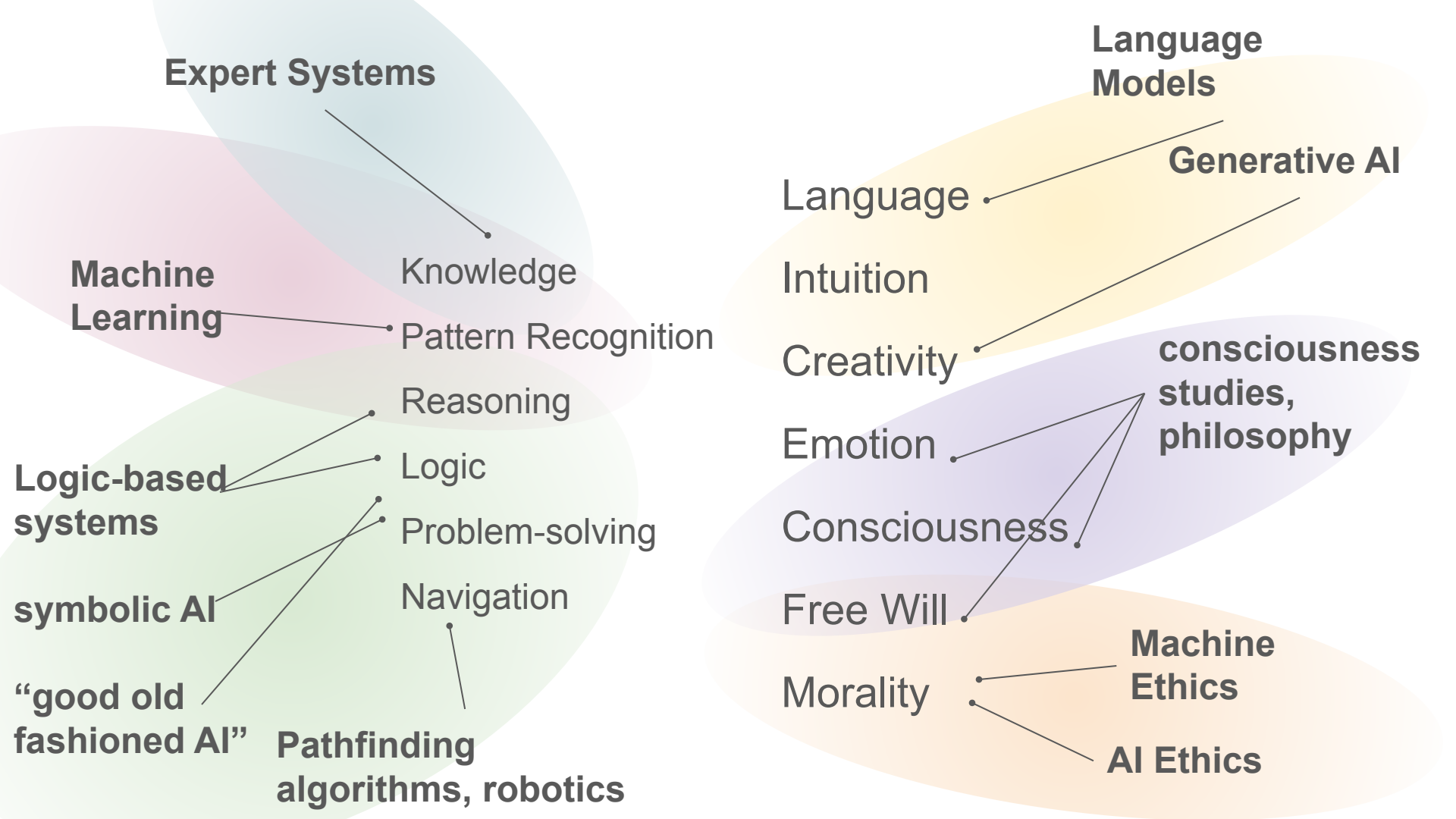
Kathy Pretz is the editor in chief of The Institute, IEEE's member publication

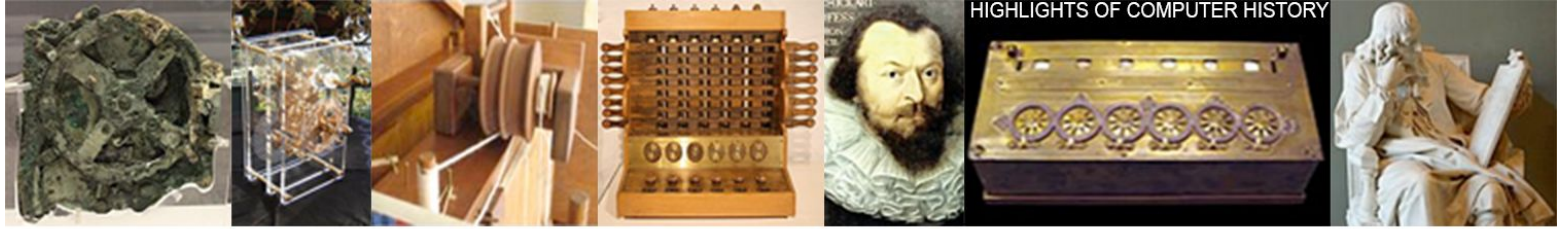
What is intelligence?

What is intelligence?

- Knowledge
- IQ
- Pattern Recognition
- Reasoning
- Logic
- Problem-solving
- Navigation
- Language

- Intuition
- Creativity
- Emotion
- Consciousness
- Free Will
- Morality


Expert Systems
Knowledge
Machine Learning
Pattern Recognition
Reasoning
Logic-based systems
Logic
Problem-solving
symbolic AI
Navigation
“good old fashioned AI”
Pathfinding algorithms, robotics
Language Models
Generative AI
Language
Intuition
Creativity
consciousness studies, philosophy
Emotion
Consciousness
Free Will
Machine Ethics
Morality
AI Ethics



1st century BC: first known gear-based calculator in Antikythera

AD 60: programmable automaton by Heron

1600s: input data! **1623:** first gear-based input-processing calculator by Schickard

1640: Pascal's superior Pascaline for simple arithmetics

1670s: Leibniz 1st computer scientist? 1st machine with memory. Principles of binary computers. Algebra of Thought.

1800: first commercial program-controlled machines (looms) by Jacquard et al. First industrial programmers; software on punchcards

1830s: Lovelace & Babbage's ideas on programs for general computers, albeit unrealized

1914: Torres y Quevedo, the pioneer of practical AI, builds a working chess end game player - chess was considered an intelligent activity

1931: Theoretical computer science founded by Gödel. First universal coding language. Exhibits the fundamental limits of math & theorem proving & AI & computing.

1935: Church extends Gödel's result to Entscheidungsproblem (decision problem). **1936:** Turing, too. Later helps to break Enigma code.

1936: Zuse's patent application. **1941:** First working programmable general-purpose computer

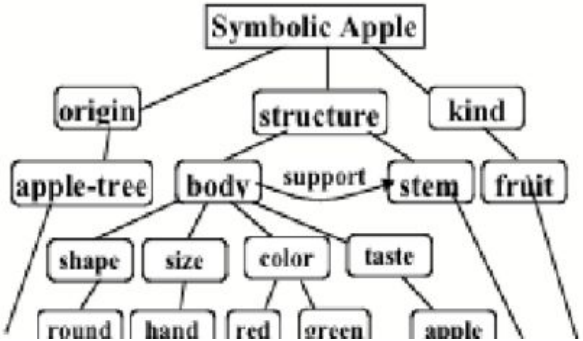
Every 5 years compute got 10 times cheaper. **2020:** 80 years ~ 10^{16}

Schmidhuber

Logic-Based Systems, Symbolic AI

Represent the world through a system of symbols

Solve tasks using rules of symbol manipulation



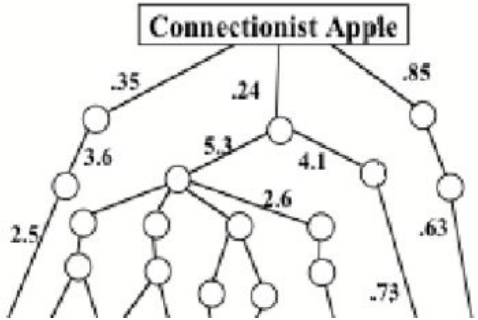
Modus Ponens

The following valid arguments show us how to apply the modus ponens (Rule of Detachment).

a) **1)** Lydia wins a ten-million-dollar lottery. p

2) If Lydia wins a ten-million-dollar lottery, then Kay will quit her job. $p \rightarrow q$

3) Therefore Kay will quit her job. $\therefore q$



Data Speaks - Data Stories that Matter

October 2017

Publisher: LUCA - Telefonica

V. Richard Benjamins

Timeline of research in logic

- 450 B.C. Stoics - ***propositional logic*** (PL), ***inference***
- 322 B.C. Aristotle - quantifiers, ***syllogism rule*****:** dominant model of correct argumentation
- 1565 Cardano - probability theory (PL + uncertainty)
- 1646 -1716 Leibniz – the founder of ***symbolic logic,*** he proposed a ***general decision procedure*** to check the validity of formulas
- 1847 George Boole – algebras, ***formalization of propositional logic***
- 1847 Augustus De Morgan: ***formal logic***
- 1879 Gottlob Frege – ***predicate or first-order logic*** (FOL)
- 1889 Peano - 9 axioms for natural numbers
- 1920 Hilbert's program
- 1921 Emil Post- ***truth tables***, 1922 Wittgenstein - ***proof by truth tables***

Timeline of research in logic (contd.)

- 1929 Gödel completeness theorem of FOL
- 1930 Herbrand - a ***proof procedure for FOL*** based on propositionalization
- 1931 Gödel incompleteness theorems for the consistency of Peano axioms
- 1936 Gentzen - a proof for the consistency of Peano axioms in set theory
- 1936 Church and Turing: ***undecidability of FOL***
- 1954 Davis - ***first machine-generated proof***
- 1955 **Beth - *Semantic Tableaux***
- 1957 Newell - First machine-generated proof in Logic Calculus
- 1958 Prawitz - First prover for FOL
- 1959 Gilmore, Wang - more provers
- 1960 **Davis - Putnam *Procedure***
- 1963 **Robinson - *Unification, resolution***
- 1968 **R.Smullyan – *Semantic tableaux proof method***

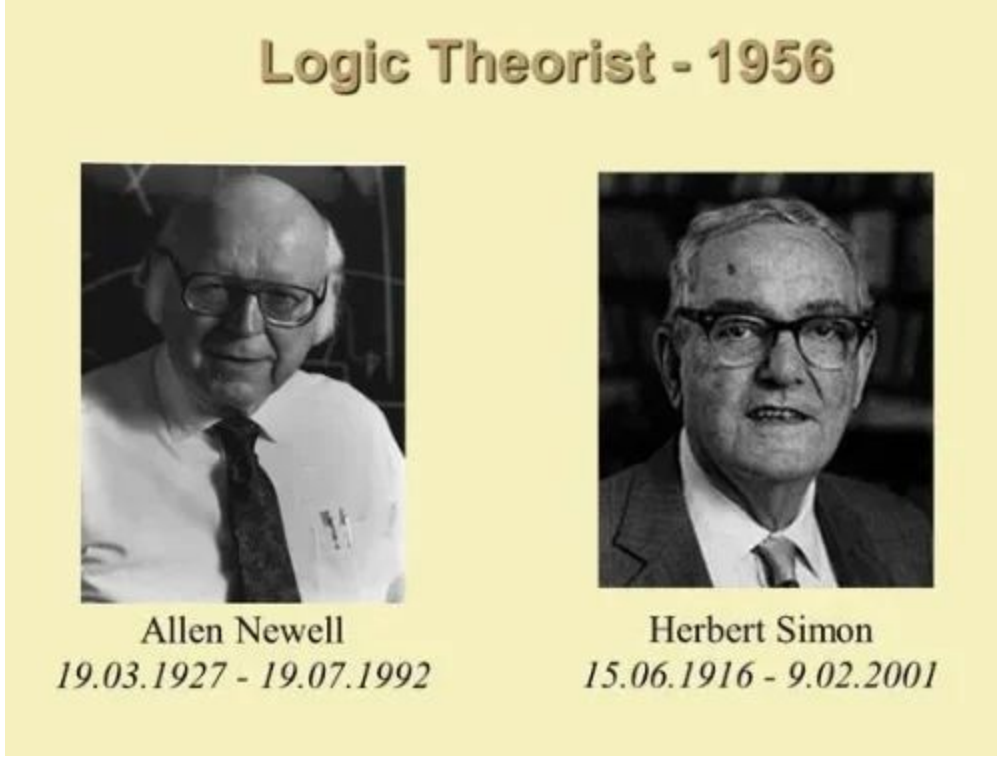
Logic Theorist - 1956

Herbert Simon, Allen Newell,
John Shaw (not pictured)

Proved 38 of the first 52 theorems in chapter 2 of the Principia Mathematica from simple axioms.

Proofs were trivial by today's standards

Superseded by General Program Solver (GPS) and many more systems for theorem proving



JUNE 8, 2024 | 12 MIN READ

AI Will Become Mathematicians' 'Co-Pilot'

Fields Medalist Terence Tao explains how proof checkers and AI programs are dramatically changing mathematics

BY CHRISTOPH DRÖSSER

AI ANTHOLOGY

Embracing change and resetting expectations

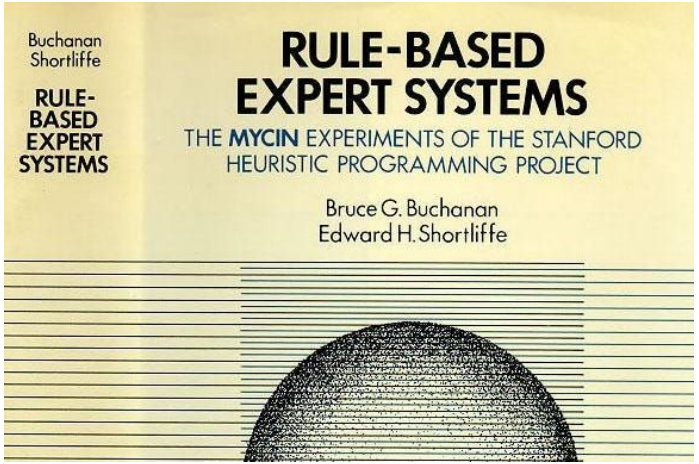
By Terence Tao
Professor of Mathematics at University of California, Los Angeles

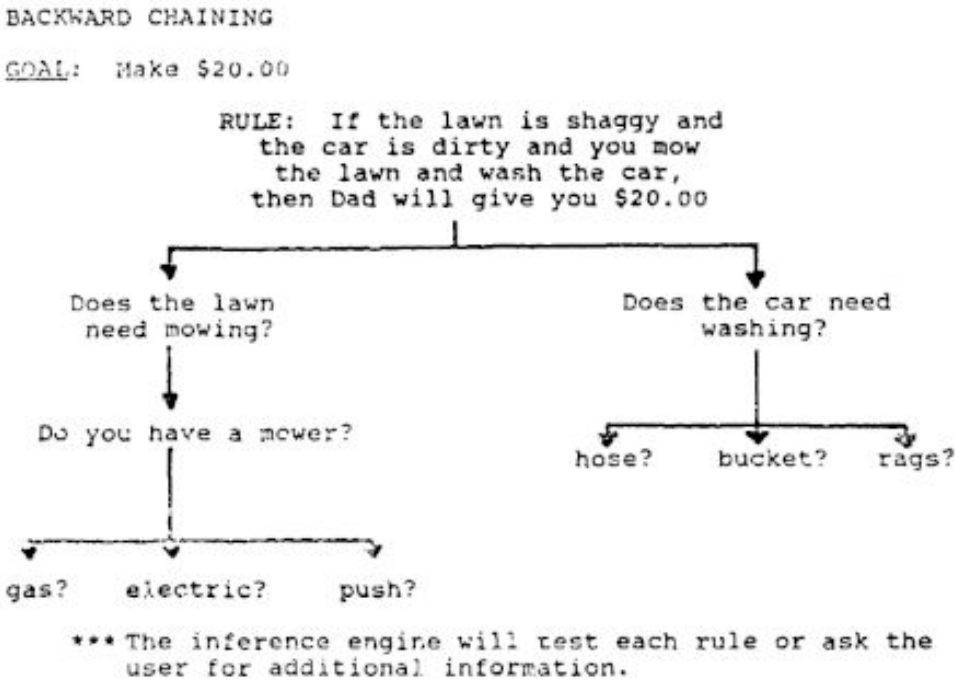
I expect, say, 2026-level AI, when used properly, will be a trustworthy co-author in mathematical research, and in many other fields as well.
- Terence Tao

Expert Systems

Use inference rules + world knowledge

to answer questions





MYCIN
Stanford, early 1970s
Edward Shortliffe

Simple inference engine +
knowledge-base of 600 rules

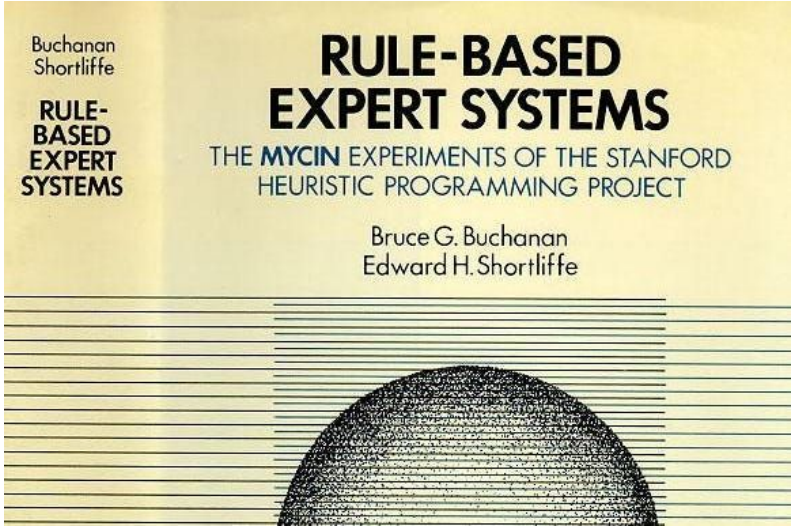
Identify likely bacterial cause
Recommend treatment

65% acceptability, on par with 42-62% acceptability of human treatment recommendations

MYCIN: A rule-based program for medical diagnosis and therapy

Why MYCIN?

- Diagnose likely infecting organisms in blood and meningitis infections
- Use test results and information about patient supplied by doctor
- Prescribe an effective antibiotic treatment
- Do this early in the course of disease, before all possible information is available
- To counteract:
  - overuse of antibiotics
  - irrational use of antibiotics
  - maldistribution of expertise

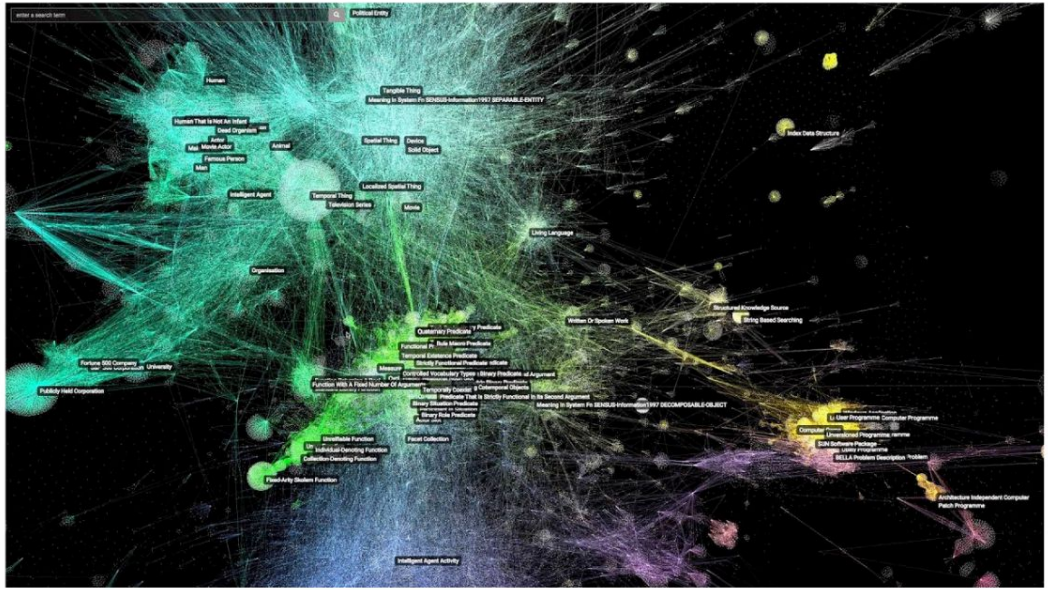


Cyc Knowledge Base

Started in 1980s by Douglas Lennat

Over 1 million assertions (facts)

*"**Cyc** (pronounced /ˈsaɪk/ SYKE) is a long-term artificial intelligence project that aims to assemble a comprehensive ontology and knowledge base that spans the basic concepts and rules about how the world works. **Hoping to capture common sense knowledge, Cyc focuses on implicit knowledge that other AI platforms may take for granted."** - Wikipedia*



The World's Broadest and Deepest Common Sense Knowledge Base

Practical, Retargetable, and Reusable Real-World Knowledge

Evolutionary Computation

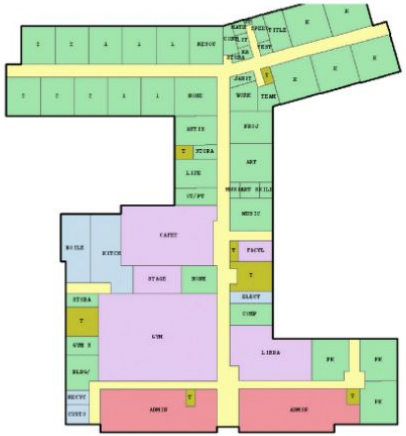
Algorithms inspired by the process of natural selection.

Maintain a pool of solutions, mutate and combine, select the best according to a fitness function.

Famous examples:

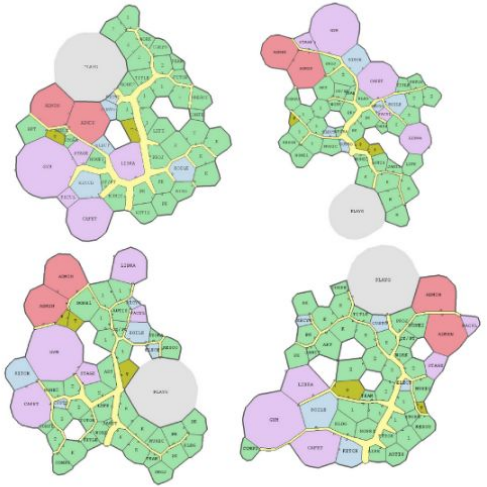
- Genetic Algorithms (1950s)
- Genetic Programming (1990s)
- Neuroevolution (2000s)

Before

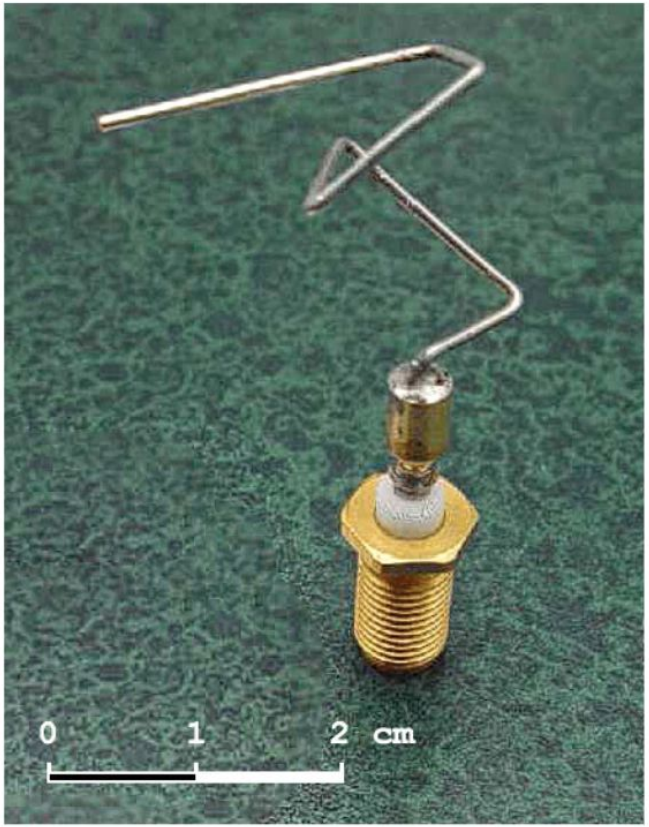


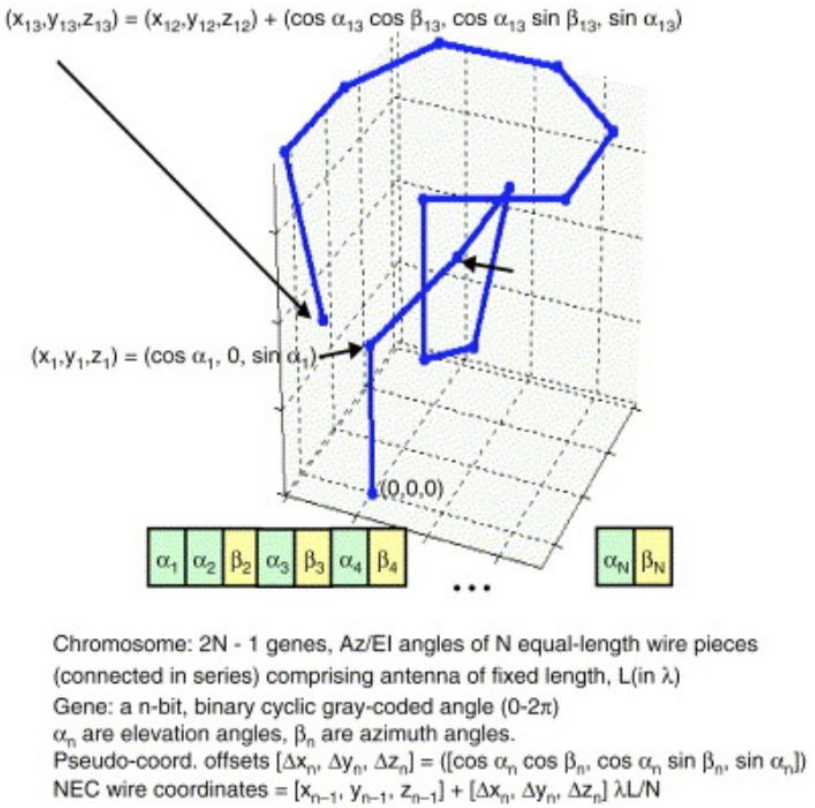
The original elementary school. Found somewhere in Maine.

After



https://www.joelsimon.net/evo_floorplans.html

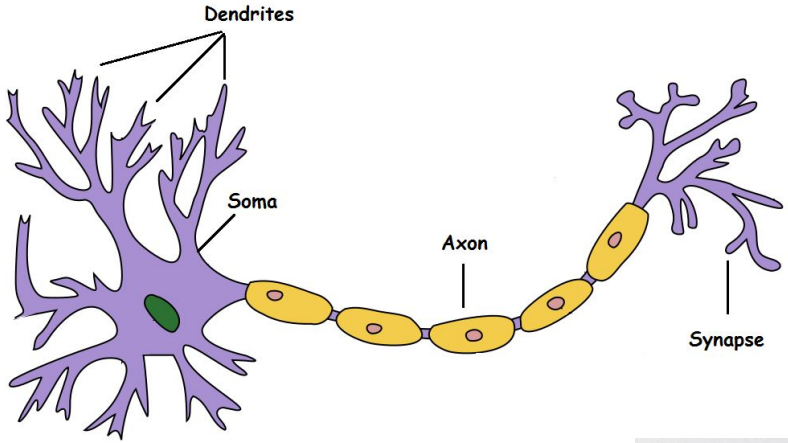


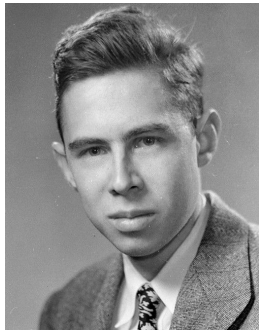


Chromosome: 2N - 1 genes, Az/El angles of N equal-length wire pieces (connected in series) comprising antenna of fixed length, L(in λ)
Gene: a n-bit, binary cyclic gray-coded angle (0-2π)
α_n are elevation angles, β_n are azimuth angles.
Pseudo-coord. offsets $[\Delta x_n, \Delta y_n, \Delta z_n] = ([\cos\alpha_n \cos\beta_n, \cos\alpha_n \sin\beta_n, \sin\alpha_n])$
NEC wire coordinates = $[x_{n-1}, y_{n-1}, z_{n-1}] + [\Delta x_n, \Delta y_n, \Delta z_n]\ \lambda L/N$

https://www.sciencedirect.com/science/article/pii/S

Connectionism





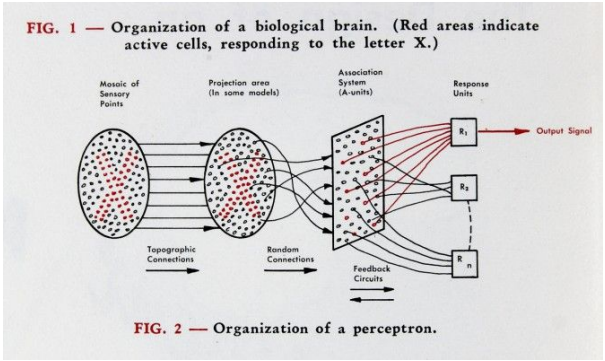
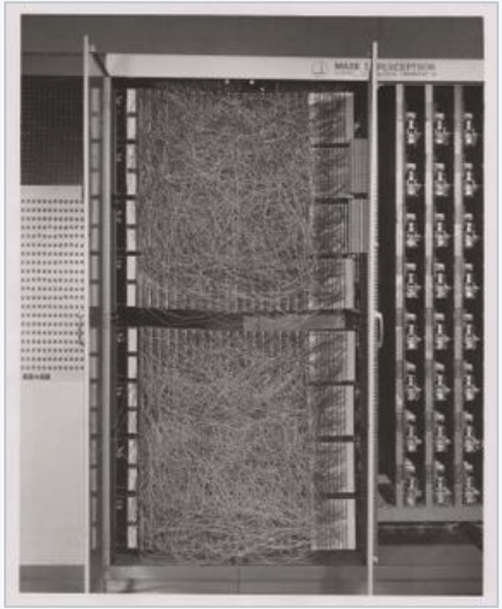


FIG. 1 — Organization of a biological brain. (Red areas indicate active cells, responding to the letter X.)

FIG. 2 — Organization of a perceptron.



Mark I Perceptron machine, the first implementation of the perceptron algorithm. It was connected to a camera with 20×20 cadmium sulfide photocells to make a 400-pixel image. The main visible feature is the sensory-to-association plugboard, which sets different combinations of input features. To the right are arrays of potentiometers that implemented the adaptive weights.[2]:213

"The Design of an Intelligent Automaton," in Research Trends, a Cornell Aeronautical Laboratory publication, Summer 1958.

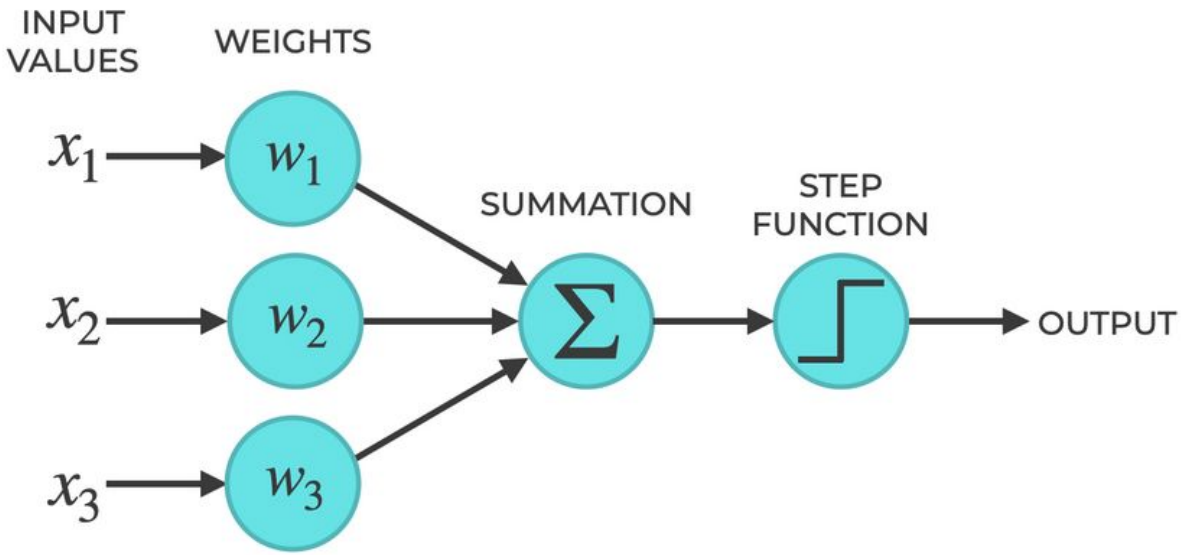
Connectionism

Invented 1943 McCulloch & Pitts

Implemented by Rosenblatt 1957

Learns to classify inputs into two categories by adjusting weights

Basis for modern neural networks

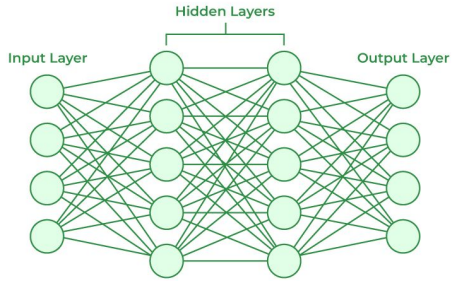


Machine Learning

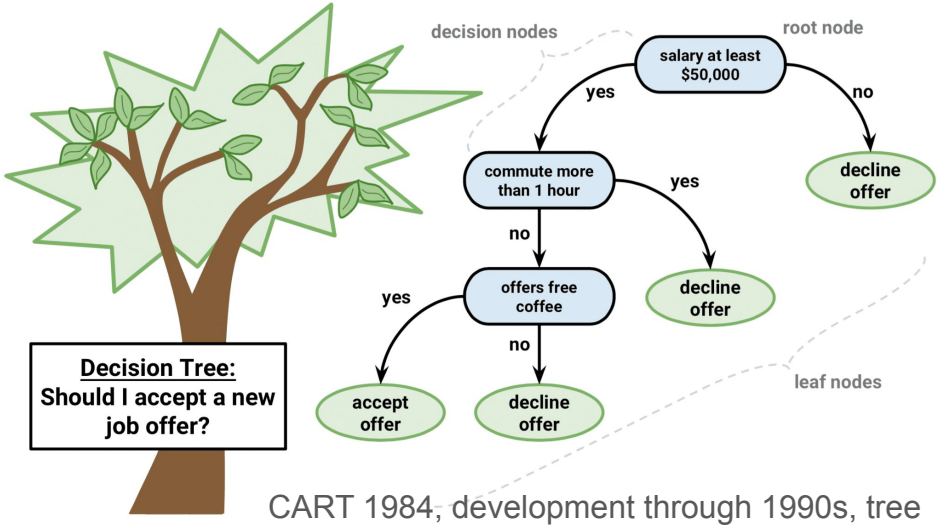
Learning from data

Modern machine learning inherits from the connectionist tradition as well as statistics and mathematics

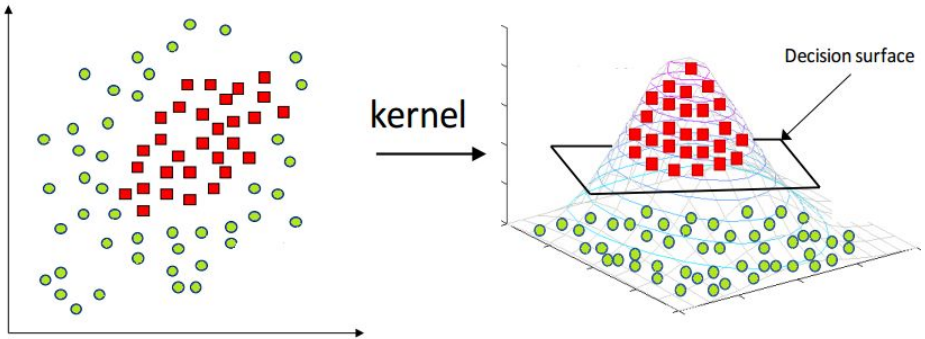
View intelligence as **function approximation** - modeling a data generating process



Neural Network
Perceptron theory 1943, implementation 1957
Multi-layer perceptron proposed 1958



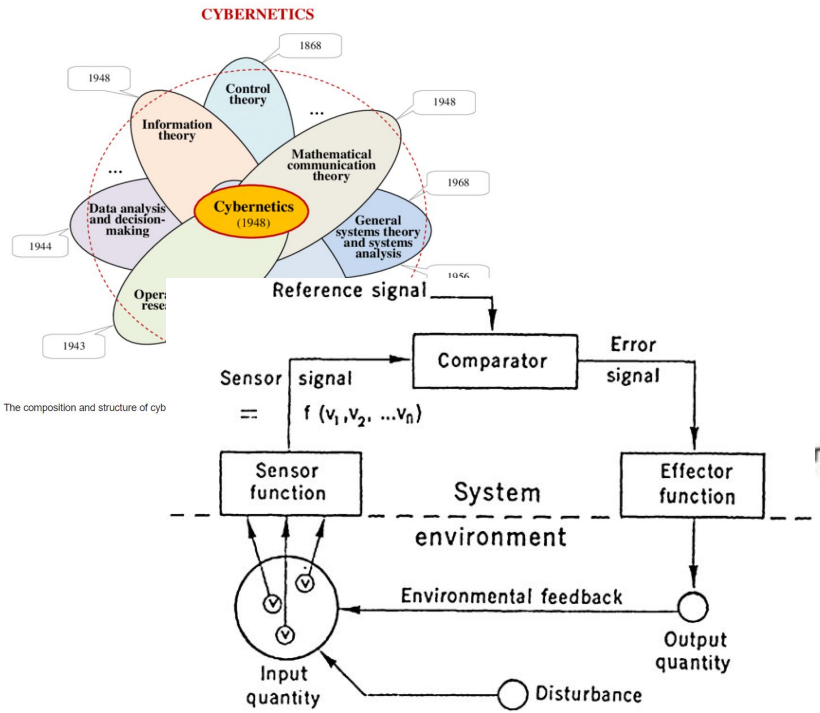
CART 1984, development through 1990s, tree ensembles still developing today

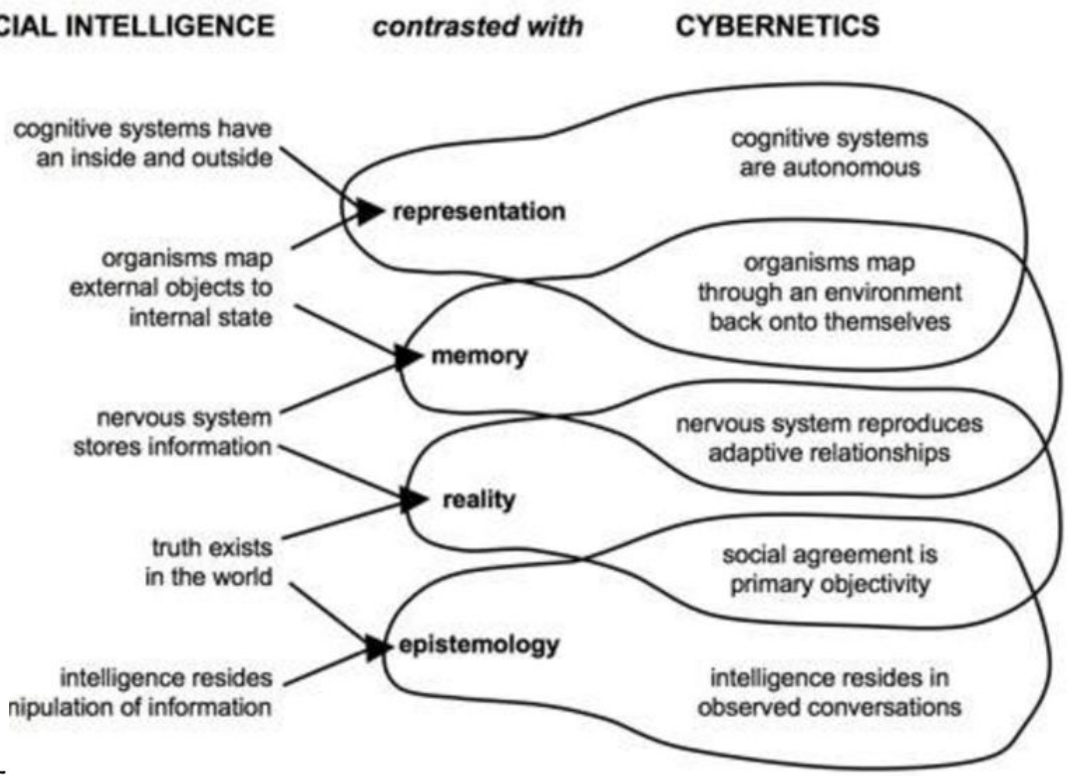


Support Vector machines, proposed 1964 Vapnik & Chervonenkis, Kernel SVM 1992

Cybernetics

The alternative history of AI





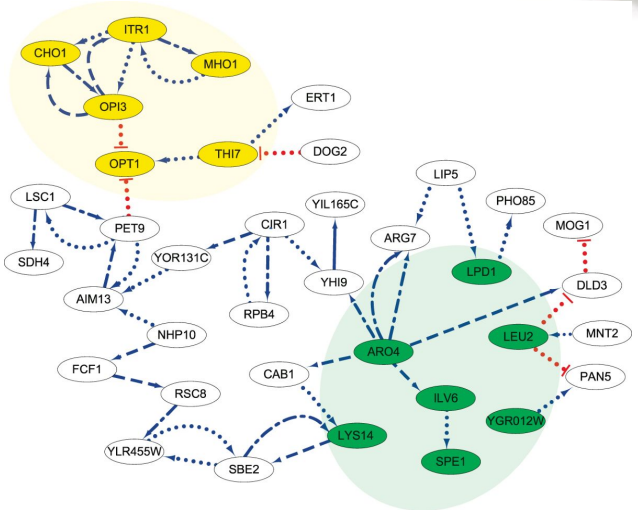
Cybernetics vs. AI, 1990, Paul Pangaro

Cybernetics

Intelligence as an emergent phenomenon in situated systems

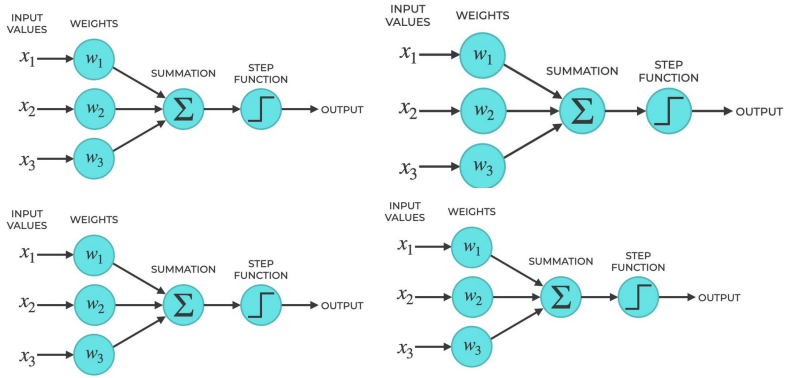
Everything from thermostats to gene regulatory networks

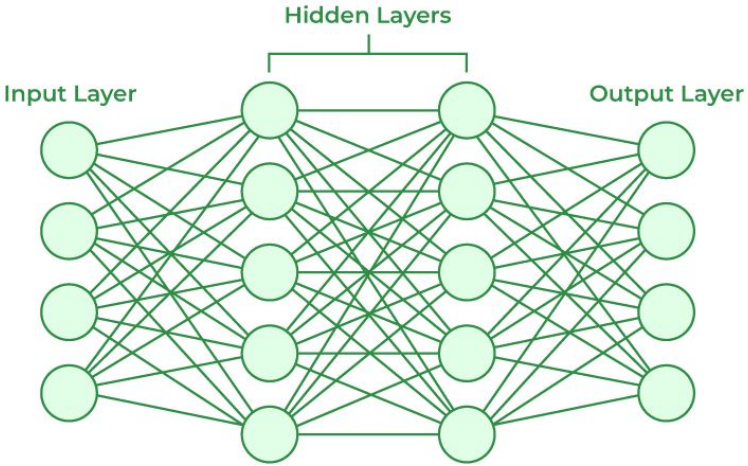
Spillover to “mainstream” AI via reinforcement learning



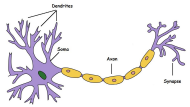
Deep Learning

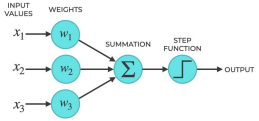
Compose many perceptron-like units into an artificial neural network with multiple layers



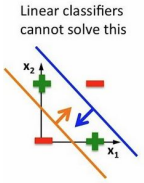


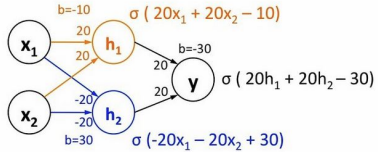
Deep Learning History





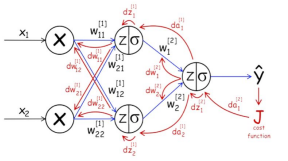
Solving XOR with a Neural Net



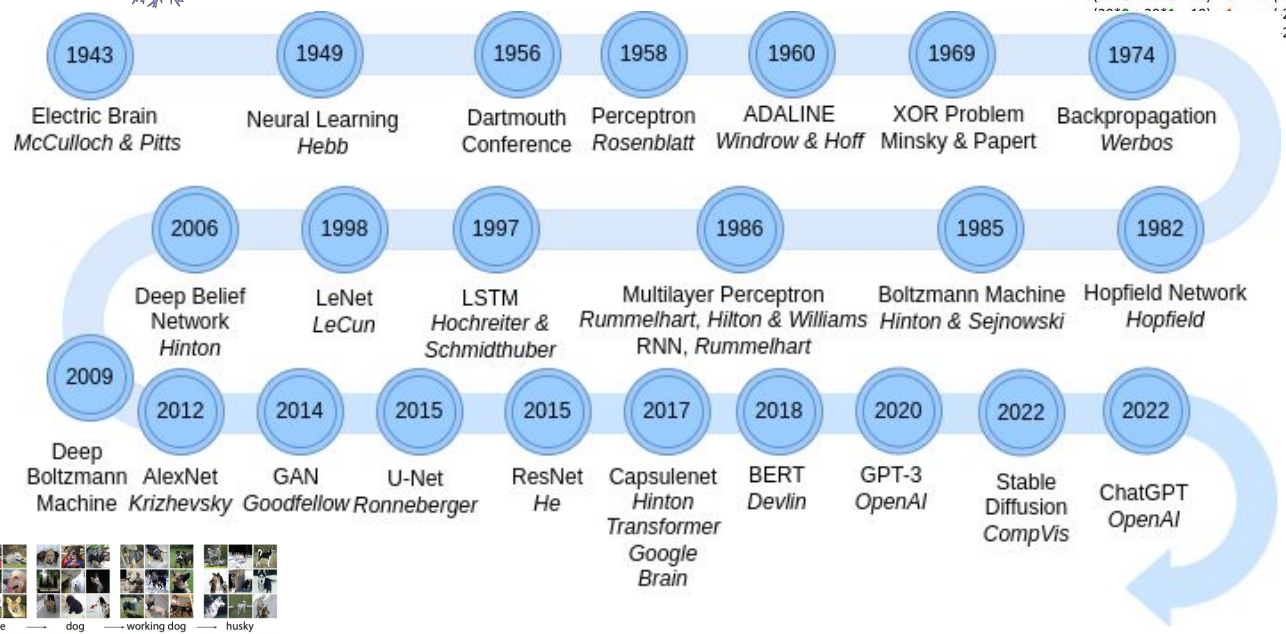


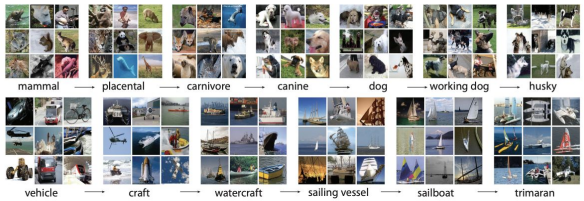
σ(20*0 + 20*0 – 10) ≈ 0
σ(20*1 + 20*1 – 10) ≈ 1
σ (-20*0 – 20*0 + 30) ≈ 1
σ (-20*1 – 20*1 + 30) ≈ 0
σ (20*0 + 20*1 – 30) ≈ 0
σ (20*1 + 20*0 – 30) ≈ 0
20*0 – 20*1 + 30) ≈ 1
20*1 – 20*0 + 30) ≈ 1
σ (20*0 + 20*1 – 30) ≈ 1
σ (20*1 + 20*1 – 30) ≈ 1

Victor Lavrenko



Backpropagation

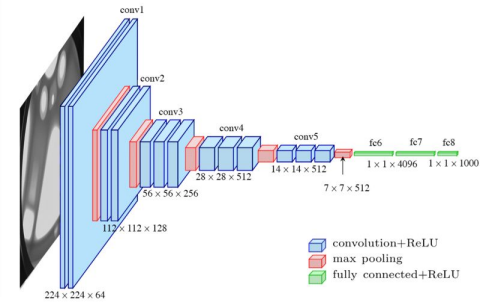




https://towardsai.net/p/l/a-brief-history-of-neural-nets

ImageNet, AlexNet

“The paradigm shift of the ImageNet thinking is that while a lot of people are paying attention to models, let’s pay attention to data. Data will redefine how we think about models.” – Fei-Fei Li



At the time, the compute resource to train such a network was scarce. But the introduction of optimized GPUs made the possibility of training deep conventional neural network achievable. The particular GPU used to train the AlexNet CNN architecture was the NVIDIA GTX 580 3GB GPU.

Nvidia GTX 580

Key Dates

2009

IMAGENET

ImageNet is presented for the first time as a poster at the Conference on Computer Vision and Pattern Recognition (CVPR) in Florida.

2012

ALEXNET

The deep convolutional neural network architecture AlexNet beats the field in the ImageNet Challenge by a whopping 10.8% — arguably kickstarting the current boom in computer vision.

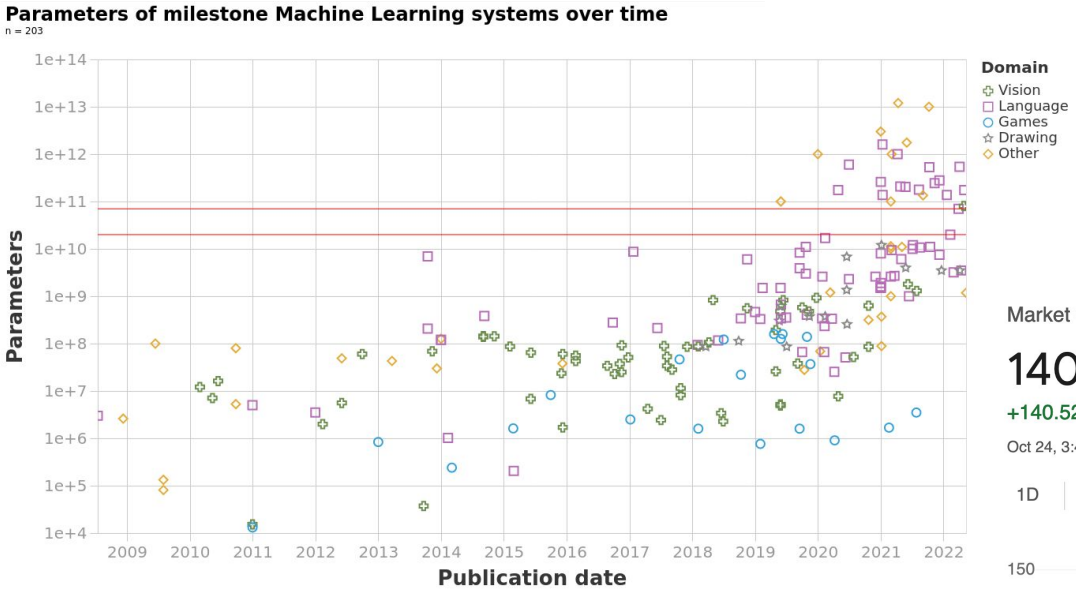
2017

95% ACCURACY IN COMPUTER VISION

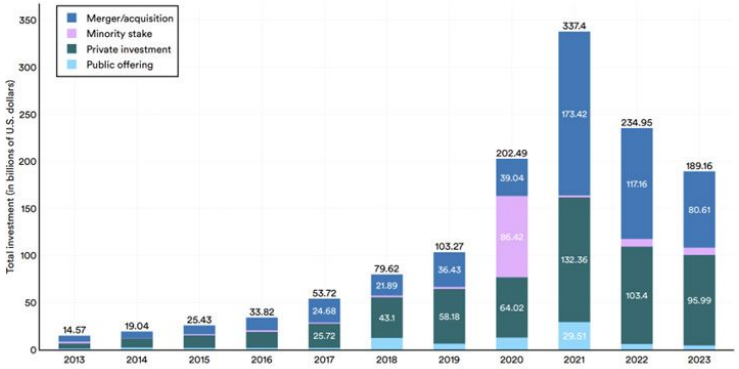
29 of 38 the teams competing in the ImageNet Challenge achieve greater than 95% accuracy. Image recognition has been taken to unprecedented levels.

History of Data Science

Confluence of data, computing hardware improvement, algorithmic improvement, investment, excitement



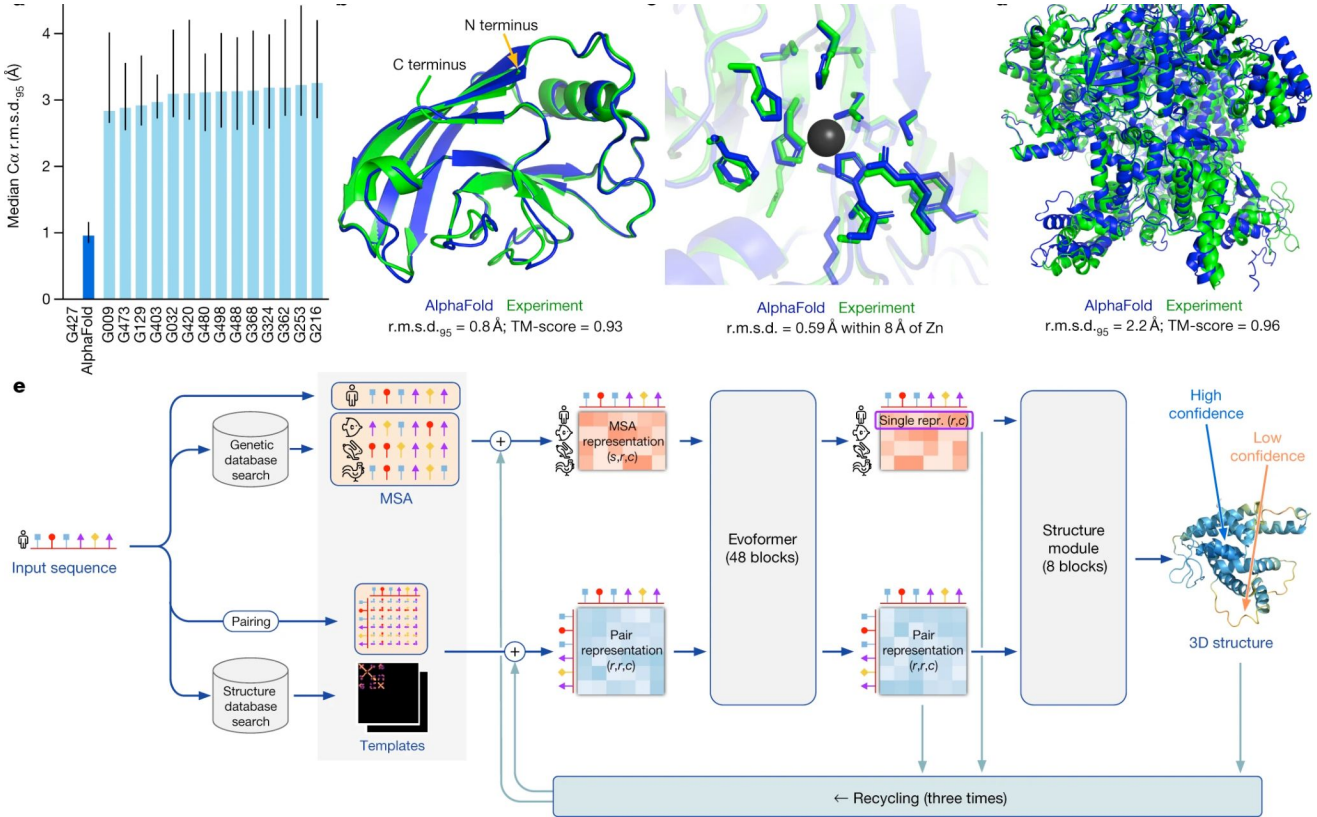
Epoch AI

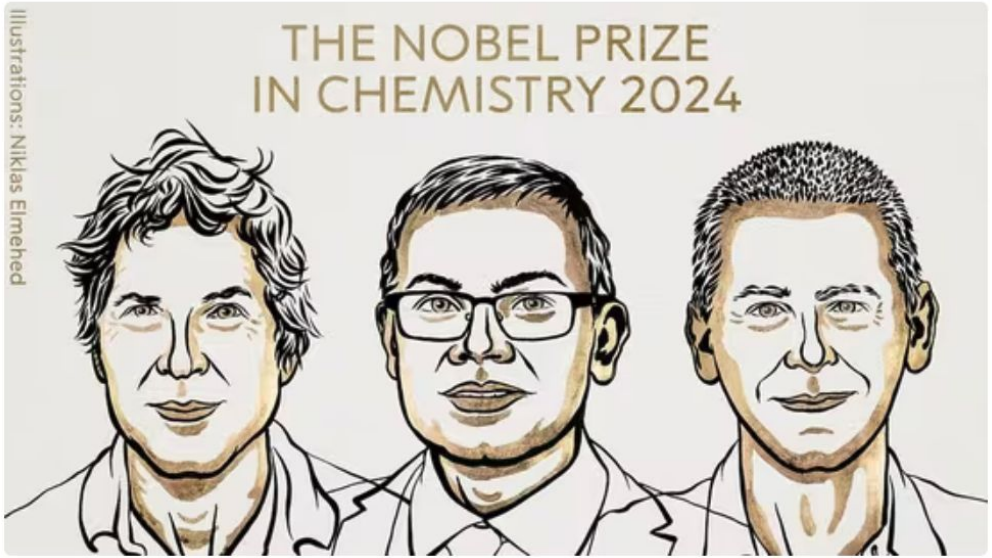


Source: Nestor Maslej, Loredana Fattorini, Raymond Perrault, Vanessa Parli, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald and Jack Clark, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2024. ("The AI Index 2024 Annual Report" hereafter)

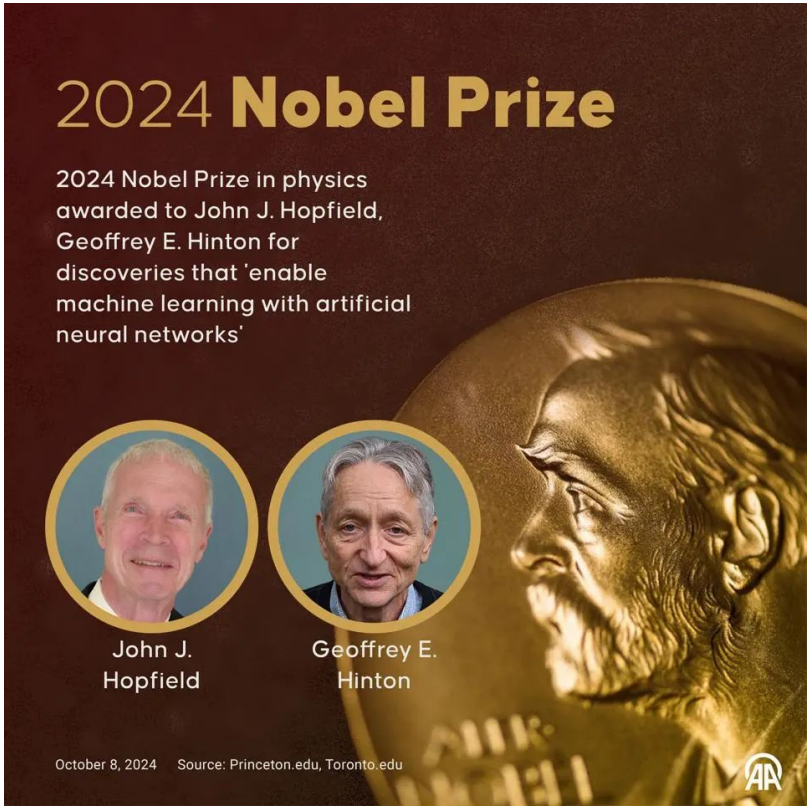
Fig. 1: AlphaFold produces highly accurate structures.

From: [Highly accurate protein structure prediction with AlphaFold]





David Baker, Demis Hassabis, and John Jumper.(X / @NobelPrize)



Sam Altman warns AI could kill us all. But he still wants the world to use it

By Samantha Kelly, CNN

7 minute read · Published 6:00 AM EDT, Tue October 31, 2023

PauseAI Proposal

Implement a pause on the training of AI systems more powerful than GPT-4, until we know how to build them safely and keep them under democratic control.

Version: Oct 9th, 2024 Individual countries can and should implement this measure *right now*. Especially the US (or California, specifically) should implement a Pause, since it is home to virtually all leading AI companies. Many scientists and industry leaders agree that a Pause is necessary, and the (US) public also strongly supports a pause (64% - 69%).

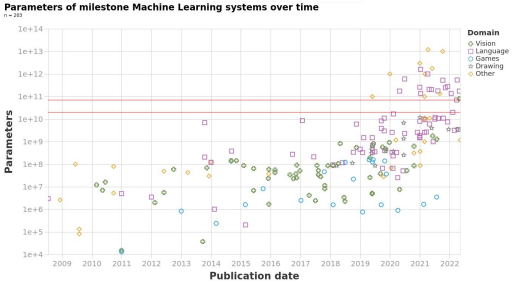
However, we cannot expect countries or companies to risk their competitive advantage by pausing AI training runs for a long time if other countries or companies do not do the same. This is why we need a **global Pause**.

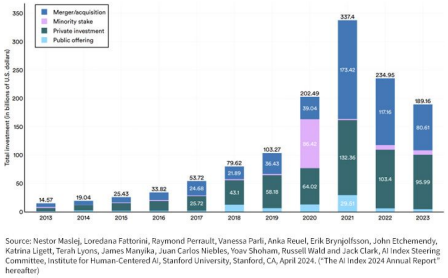
There's a long tail of things of varying degrees of badness that could happen. I think at the extreme end is the Nick Bostrom style of fear that an AGI could destroy humanity. I can't see any reason in principle why that couldn't happen. - Dario Amodei, Anthropic CEO

The upshot is simply a question of time, but that the time will come when the machines will hold the real supremacy over the world and its inhabitants is what no person of a truly philosophic mind can for a moment question.

- Samuel Butler, 1863 Darwin Among the Machines

AI Present

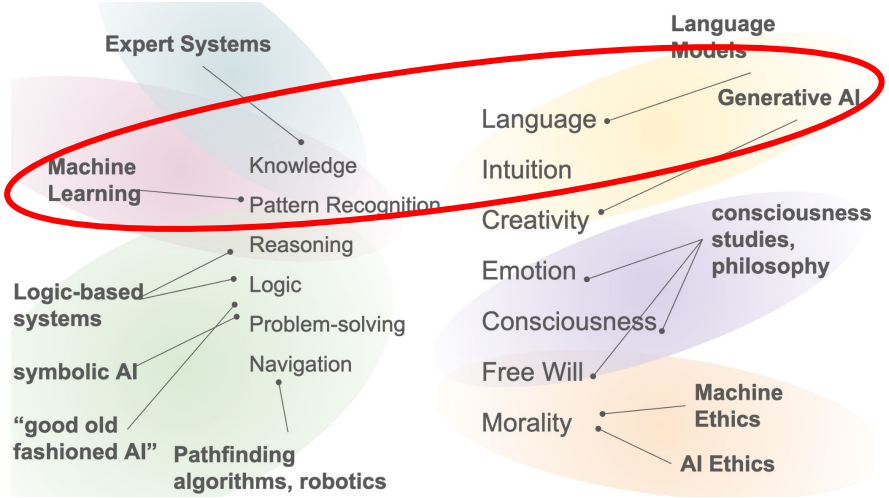


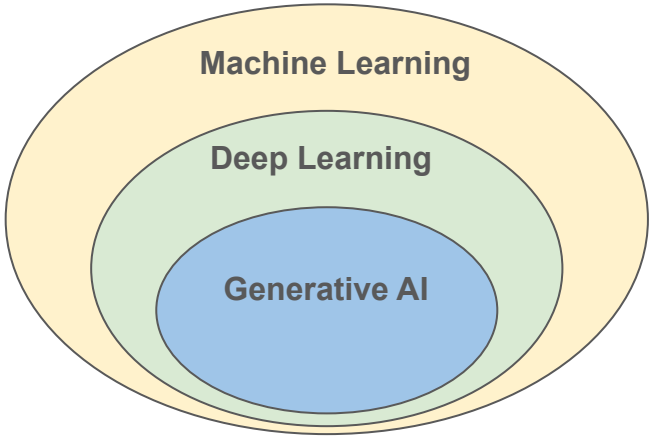


Source: Nestor Maslej, Loredana Fattorini, Raymond Perrault, Vanessa Parli, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald and Jack Clark, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2024. ("The AI Index 2024 Annual Report" hereafter)

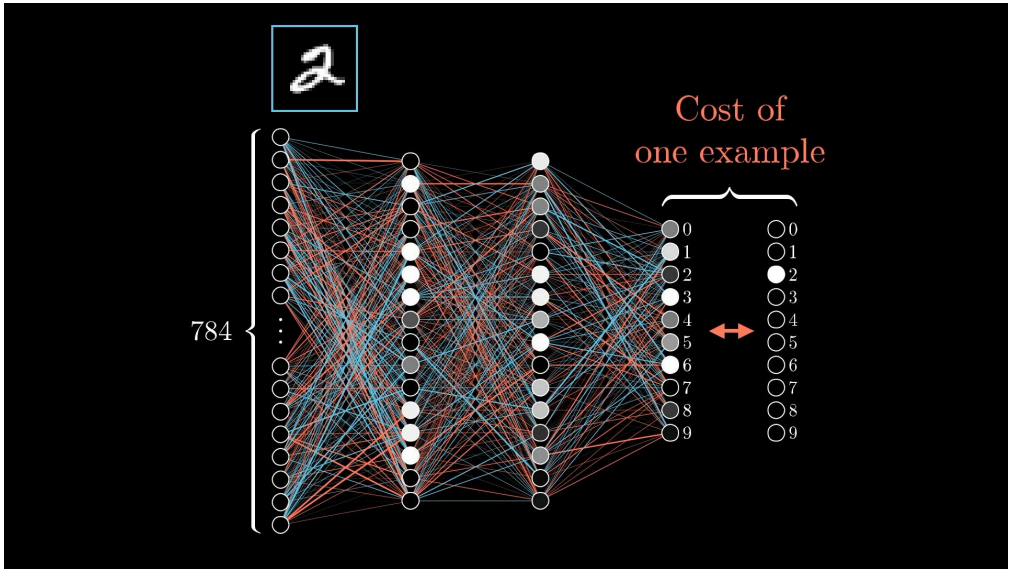
Many paradigms, but **Machine Learning**, **Deep Learning**, and **Generative AI** have seen most commercial interest.

Positive feedback cycle between data, compute, model architectures, and investment.





How does Deep Learning Work?



1. Start with a **random network**
2. Collect (input, output) examples
3. Put the inputs into the network
4. **Compare the network output with the expected output**
5. **Update the network weights** until the network output matches the expected output
6. **Repeat**

Grant Sanderson, Josh Pullen
(3Blue1Brown)

Self-Supervision

1. Start with a **random network**
2. Collect (input, output) examples
3. Put the inputs into the network
4. **Compare the network output with the expected output**
5. **Update the network weights** until the network output matches the expected output
6. **Repeat**

$$$ Costly!

How can we skip this step?

Self-Supervision uses data that already exists

With a clever trick to get “labels” for free

Fill-in-the-blank:

prompt	prediction	probability
You are likely to find a **snake** in a ___	field	0.066
One effect of **exercising** is feeling ___	better	0.296
You could be **sick** because you are ___	pregnant	0.209
If you want to **learn** then you need a ___	teacher	0.122

Self-Supervised Language Modeling:

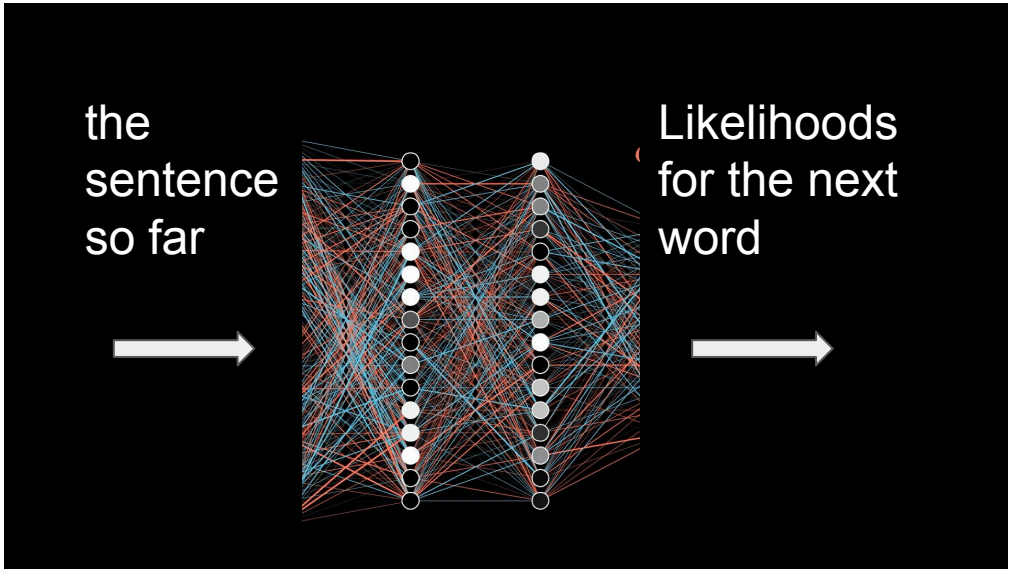
1. Find a source of lots of sentences (the Internet)

2. For each sentence:

   first 9 words: input
   10th word: output

3. Train a model to predict output from input

Generative Language Modeling



Language Modeling is flexible

x "input"	w "**text** output"
An author	A document written by that author
A topic label	An article about that topic
{SPAM, NOT_SPAM}	An email
A sentence in French	Its English translation
A sentence in English	Its French translation
A sentence in English	Its Chinese translation
An image	A text description of the image
A document	Its summary
A document	Its translation
Meterological measurements	A weather report
Acoustic signal	Transcription of speech
Conversational history + database	Dialogue system response
A question + a document	Its answer
A question + an image	Its answer

Chris Dyer, DeepMind, CMU

How much data?

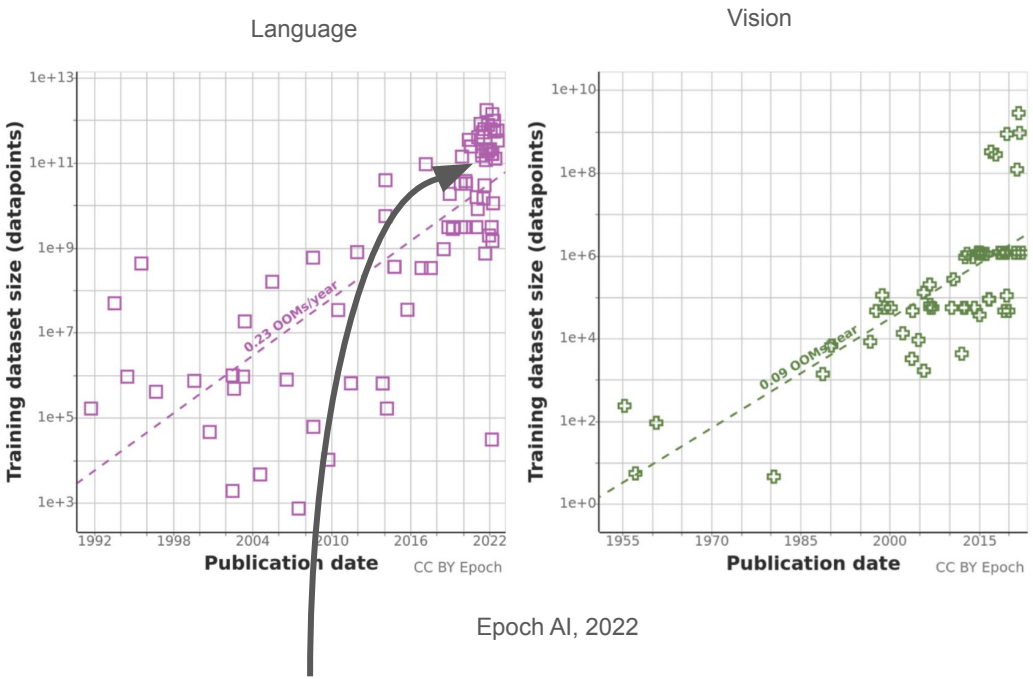
Self-Supervised Language Modeling:

1. Find a source of lots of sentences (the Internet)

2. For each sentence:

   first 9 words: input
   10th word: output

3. Train a model to predict output from input



Epoch AI, 2022

~750 billion words to train frontier AI

vs.

~30-40 million in a child's first 4 years

[Submitted on 7 Aug 2024 (v1), last revised 8 Oct 2024 (this version, v2)]

Is Child-Directed Speech Effective Training Data for Language Models?

Steven Y. Feng, Noah D. Goodman, Michael C. Frank

What is Deep Learning good at?

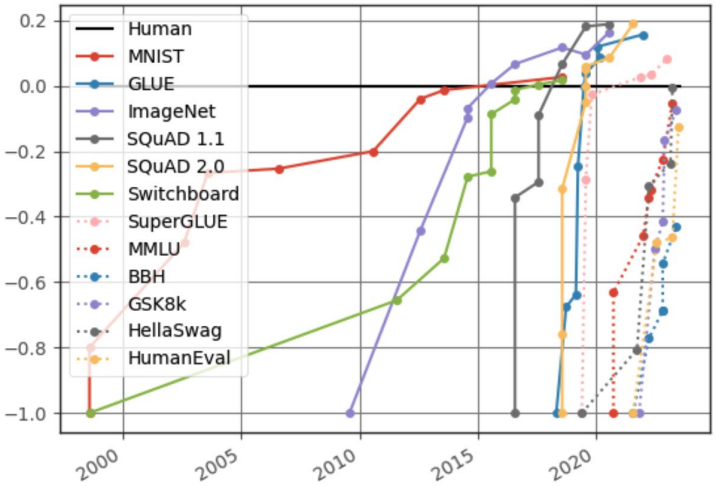
Problem-Solving ✅

Intelligence is the **ability to solve tasks, period.**

Common (implicit or explicit) view among ML practitioners in industry.

Behaviorism: If it walks like a duck, quacks like a duck, it's probably a duck.

"saturation" on benchmarks for handwriting recognition, image recognition, question answering, grade-school math, programming



Douwe Kiela et. al. Dynabench, 2021

Shortcut Learning and Spurious Correlations ✖

Shortcuts *are decision rules that* ***perform well on standard benchmarks*** *but* ***fail to transfer to more challenging testing conditions****…*

Shortcut Learning in Deep Neural Networks, Geirhos et. al., 2023

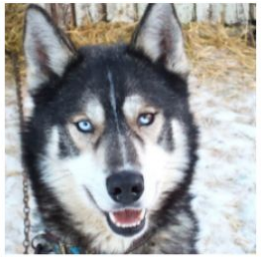
Ribiero, et. al. 2016

"Why Should I Trust You?"
Explaining the Predictions of Any Classifier

Marco Tulio Ribeiro
University of Washington
Seattle, WA 98105, USA
marcotcr@cs.uw.edu

Sameer Singh
University of Washington
Seattle, WA 98105, USA
sameer@cs.uw.edu

Carlos Guestrin
University of Washington
Seattle, WA 98105, USA
guestrin@cs.uw.edu



(a) Husky classified as wolf

(b) Explanation

Adversarial Examples ✖

Ubiquitous

Shared between models

Sensible?

Figure 1: A demonstration of fast adversarial example generation applied to GoogLeNet (Szegedy et al., 2014a) on ImageNet. By adding an imperceptibly small vector whose elements are equal to the sign of the elements of the gradient of the cost function with respect to the input, we can change GoogLeNet's classification of the image. Here our ϵ of .007 corresponds to the magnitude of the smallest bit of an 8 bit image encoding after GoogLeNet's conversion to real numbers.

Our findings prompt us to ***view adversarial examples as a fundamentally human phenomenon.***

… we should not be surprised that classifiers exploit highly predictive features … given such features exist in real-world datasets

Adversarial Examples Are Not Bugs, They Are Features

Andrew Ilyas* MIT ailyas@mit.edu

Shibani Santurkar* MIT shibani@mit.edu

Dimitris Tsipras* MIT tsipras@mit.edu

Logan Engstrom* MIT engstrom@mit.edu

Brandon Tran MIT btran115@mit.edu

Aleksander Mądry MIT madry@mit.edu

Sample Efficiency ✖

Intelligence is the **conversion of information into problem-solving behavior**.

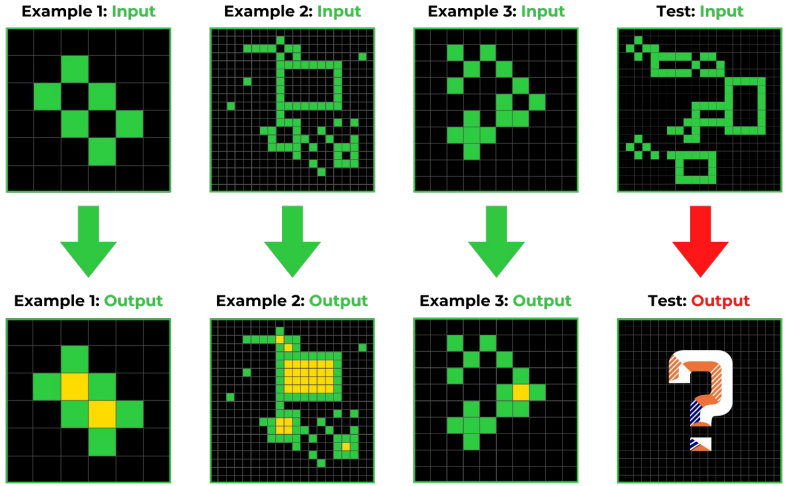
The most intelligent system is one that can solve hard tasks from little data.

Human performance: 90%+
Current AI Performance: ~40%

as of 2024

https://lab42.global/arc/

ARC (Abstraction and Reasoning Corpus) example



AI Future

Will AI progress continue? How far will it go?

Under the current paradigm of self-supervised deep learning?
Under a new or hybrid paradigm?
In theory or in practice?

Will AI exceed human intelligence?

To what extent will humans and AI be integrated?

How should we govern and control AI?

What's at stake?

Will AI systems become moral patients?

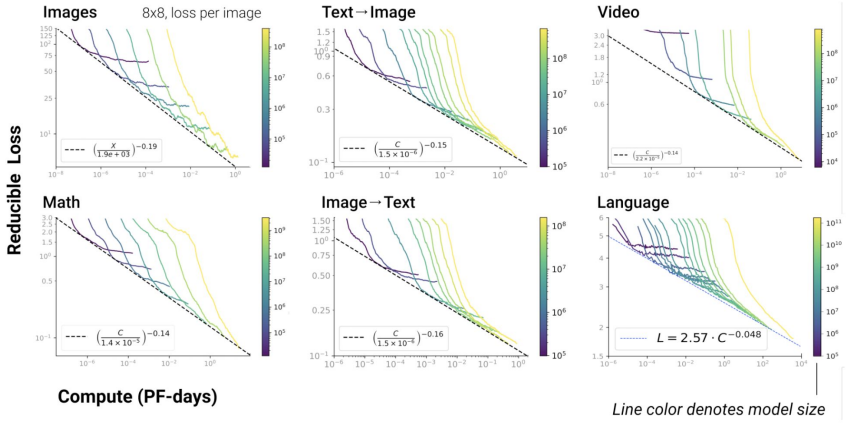
Will AI Progress Continue?

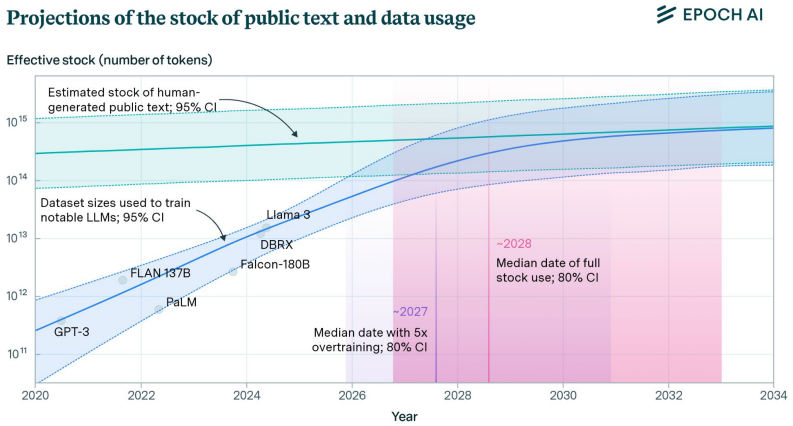
“Scaling Laws” suggest that 10x increases in data, model, and compute scale will continue to yield progress

Will there always be 10x more data?
10x more compute?

Various projections indicate that we **may exhaust text and image data sources in the ~1-2 decades**

Caveat: what about other paradigms? other sources of data?





AI timelines: What do experts in artificial intelligence expect for the future?

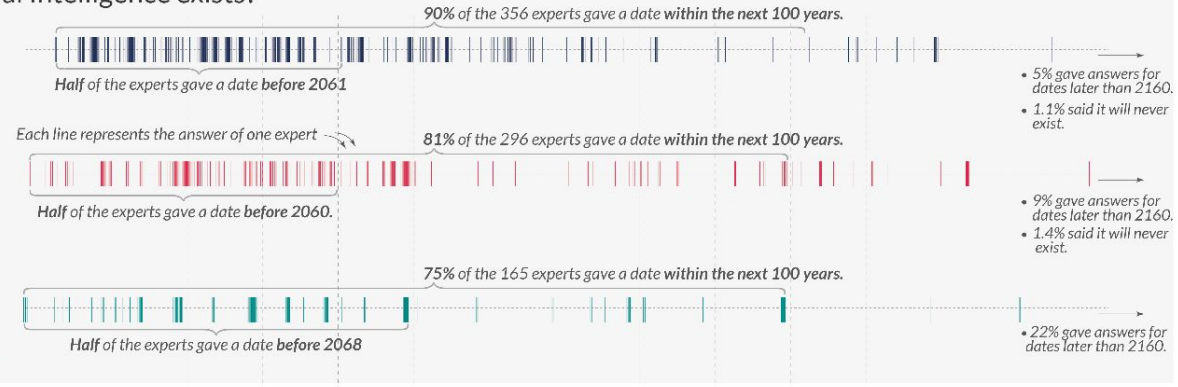
Our World in Data

When will there be a 50% chance that Human-level Artificial Intelligence exists?

1) Timelines of **356 AI experts**, surveyed **in 2022** by **Katja Grace et al.**:
The experts were asked when unaided machines will be able to accomplish every task better and more cheaply than human workers.

2) Timelines of **296 AI experts**, surveyed **in 2019** by **Baobao Zhang et al.**:
The experts were asked when machines will collectively be able to perform more than 90% of all tasks that are economically relevant better than the median human paid to do that task.

3) Timelines of **165 AI experts**, surveyed **in 2018** by **Gruetzemacher et al.**:
The experts were asked when AI systems will collectively be able to accomplish 99% of tasks that humans are paid to do at or above the level of a typical human.



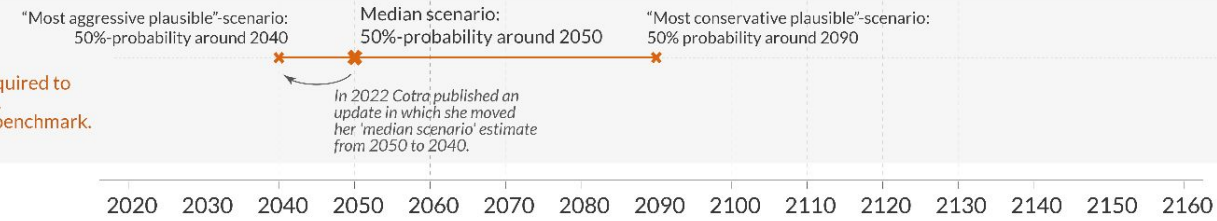
When will the first 'Artificial General Intelligence'-system be devised, tested, and publicly announced?

Community prediction of 315 forecasters on **Metaculus.com**, as of Oct. 27, 2022:

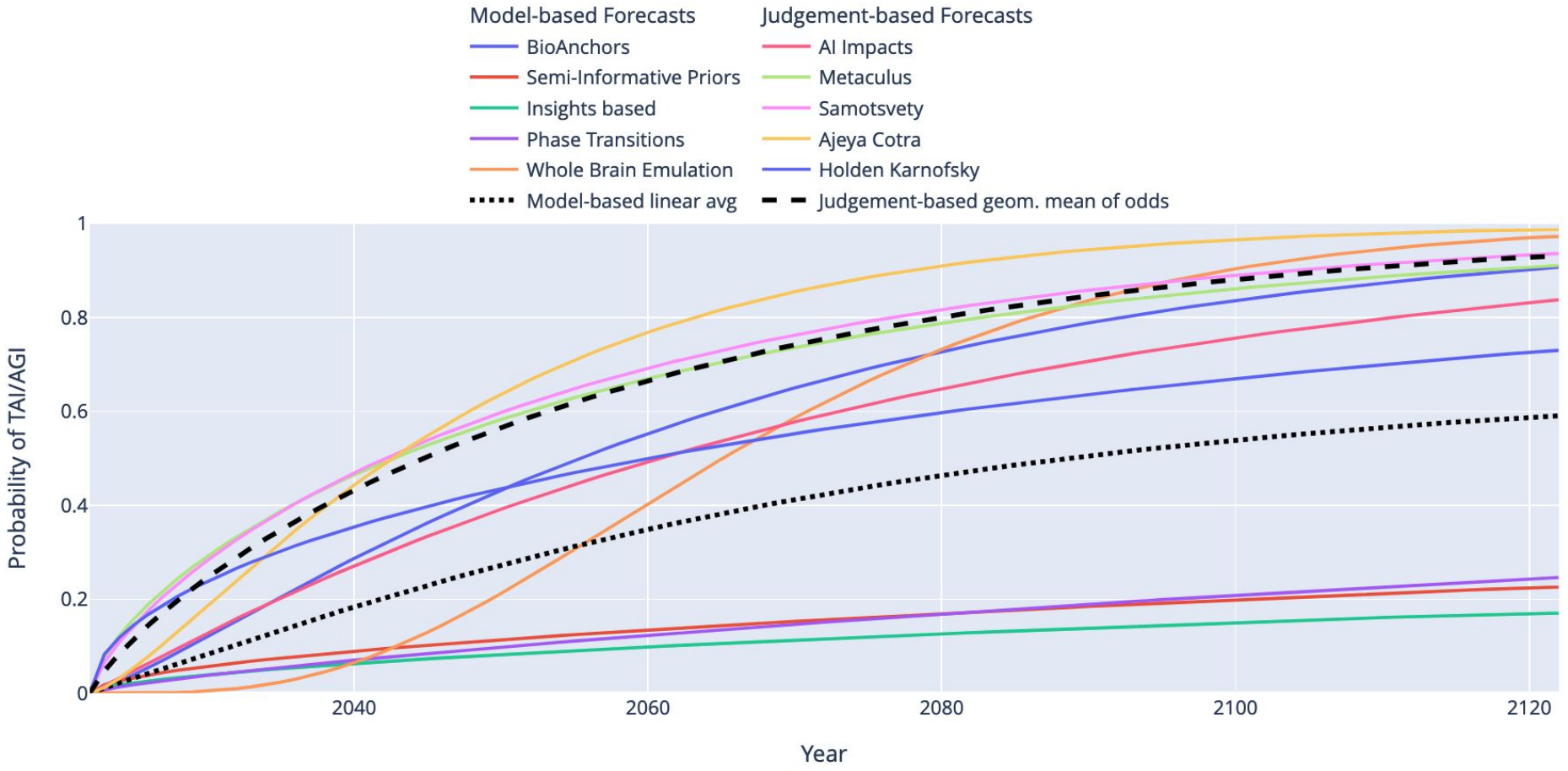
The forecasters on this open forecasting platform are asked when the first general AI system — a single, unified software system — will be "devised, tested, and publicly announced".
The system must possess certain capabilities, such as "general robotic capabilities" and "high competency at diverse fields of expertise."

Ajeya Cotra's timeline for 'Transformative AI'

In her research Ajeya Cotra estimated the computation that would be required to train a human-level AI-system using existing architecture and algorithms.
The estimates of the required computation rely on the human brain as a benchmark.



Full details on all studies and the questions that the AI experts were asked can be found in the text at OurWorldInData.org/AI-timelines.


Model-based Forecasts
BioAnchors
Semi-Informative Priors
Insights based
Phase Transitions
Whole Brain Emulation
Model-based linear avg
Judgement-based Forecasts
AI Impacts
Metaculus
Samotsvety
Ajeya Cotra
Holden Karnofsky
Judgement-based geom. mean of odds
Probability of TAI/AGI
1
0.8
0.6
0.4
0.2
0
2040
2060
2080
2100
2120
Year

Summary

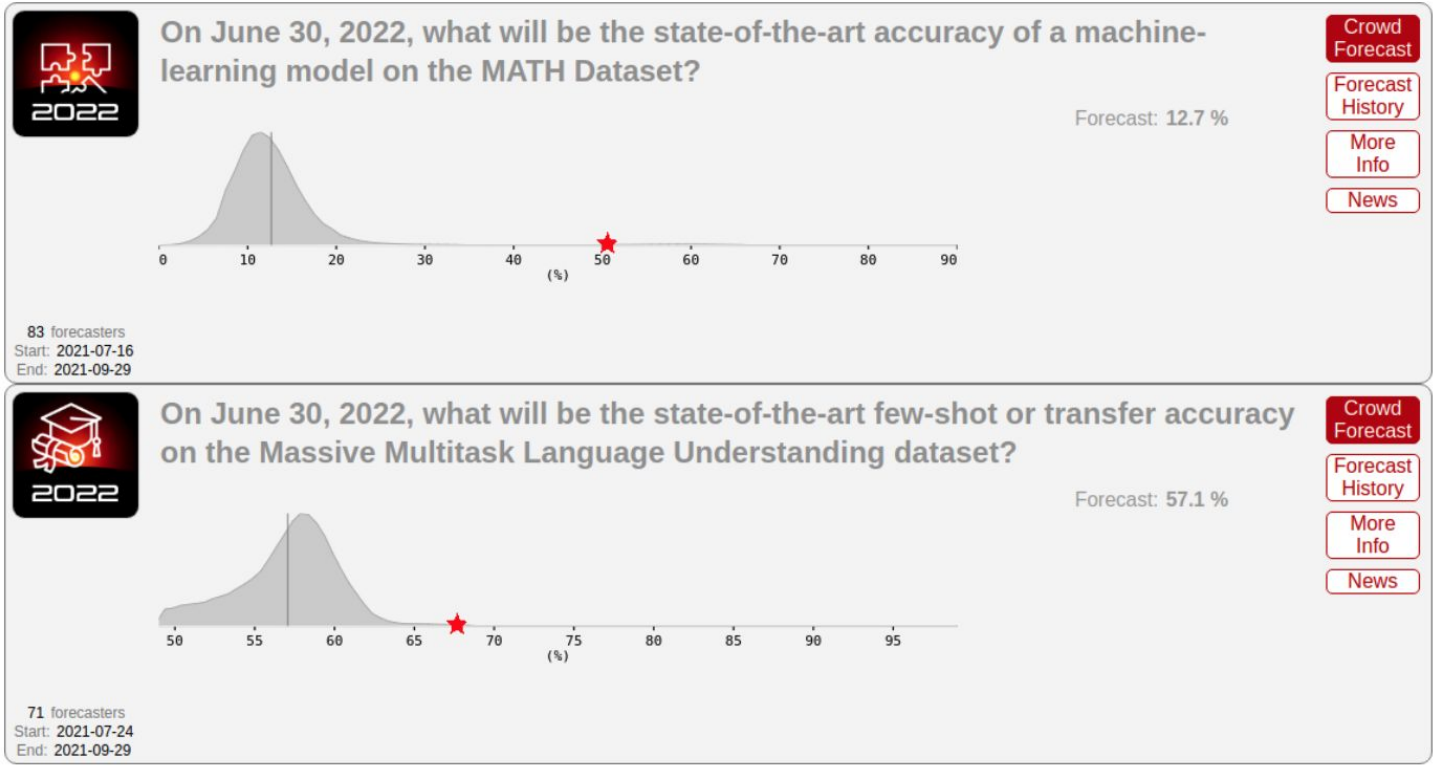
Experts **vary widely on timelines**

Many estimate a **high (>50%) chance of “Transformative AI”/“AGI” in the next 100 years**, with some timelines as short as a handful of years.

Can Experts be trusted with forecasting? (Steinhardt, UCB)

1. Forecasters' predictions were not very good in general: two out of four forecasts were outside the 90% credible intervals.
2. However, they were better than my personal predictions, and I suspect better than the median prediction of ML researchers (if the latter had been preregistered).
3. Specifically, progress on ML benchmarks happened significantly **faster** than forecasters expected. But forecasters predicted faster progress than I did personally, and my sense is that I expect somewhat faster progress than the median ML researcher does.
4. Progress on a *robustness* benchmark was slower than expected, and was the only benchmark to fall short of forecaster predictions. This is somewhat worrying, as it suggests that machine learning capabilities are progressing quickly, while safety properties are progressing slowly.

Can Experts be trusted with forecasting? (Steinhardt, UCB)



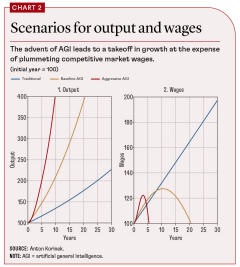
What's at stake?

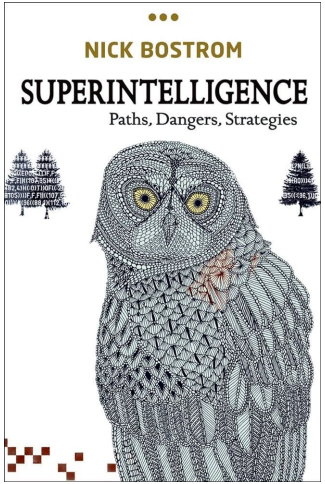
Individuals?
Communities?
Global economy?
The human species?

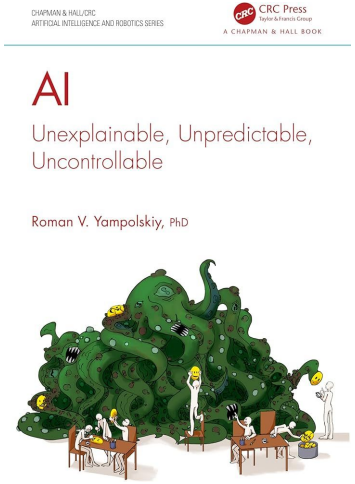
On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜

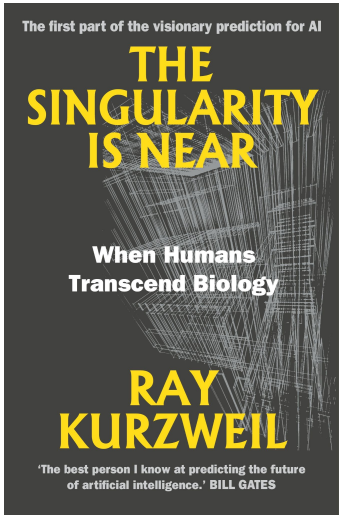
Authors: Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, Shmargaret Shmitchell | Authors Info & Claims

FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency • Pages 610 - 623 • https://doi.org/10.1145/3442188.3445922





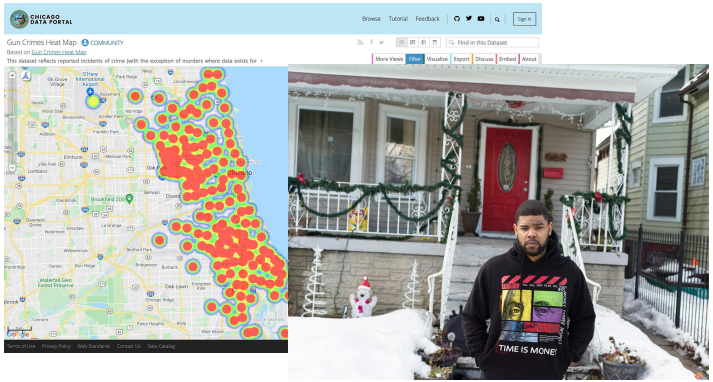




Present-day impacts

One narrative says that…

- speculation about the future of AI receives too much attention
- distracts from present-day impacts
- inequitable costs and benefits from AI



IP theft

misinformation/disinformation

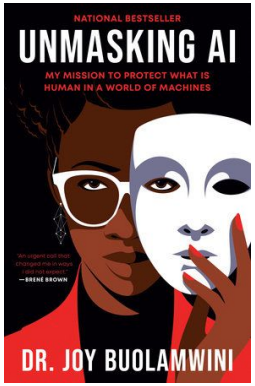
economic impacts

environmental impacts

psychological toll of data labeling

misuse

…



Look for articles by…

Deb Raji
Joy Buolamwini
Timni Gebru
among others

'It's destroyed me completely': Kenyan moderators decry toll of training of AI models

Employees describe the psychological trauma of reading and viewing graphic content, low pay and abrupt dismissals

Global Economic Impacts

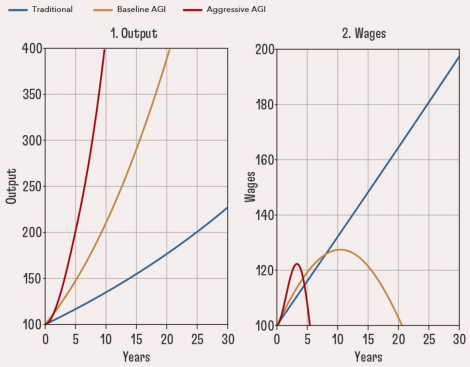
Anton Korinek
UVA, Brookings Institute

Daron Acemoglu
MIT

CHART 2

Scenarios for output and wages

The advent of AGI leads to a takeoff in growth at the expense of plummeting competitive market wages.
(initial year = 100)



SOURCE: Anton Korinek.
NOTE: AGI = artificial general intelligence.

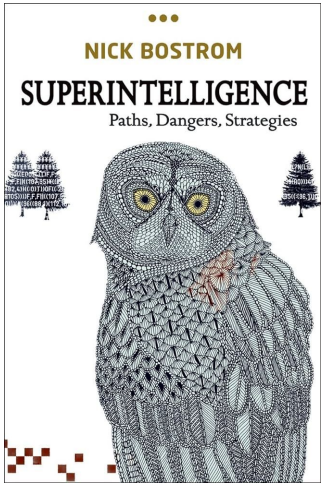
Don't Believe the AI Hype

May 21, 2024 | **DARON ACEMOGLU**

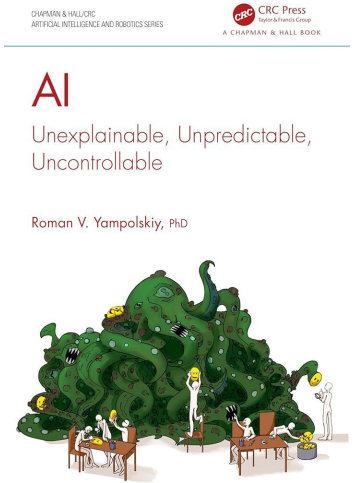
If you listen to tech industry leaders, business-sector forecasters, and much of the media, you may believe that recent advances in generative AI will soon bring extraordinary productivity benefits, revolutionizing life as we know it. Yet neither economic theory nor the data support such exuberant forecasts.

FEATU
1

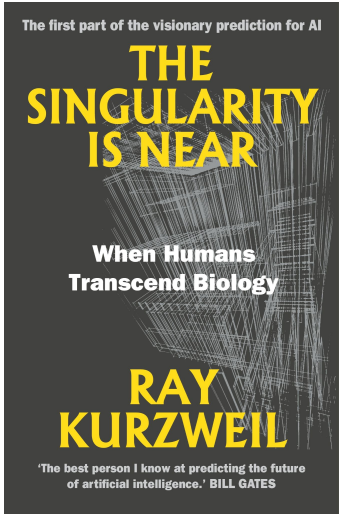
Species-Level Impacts



Nick Bostrom



Roman Yampolskiy



Ray Kurzweil

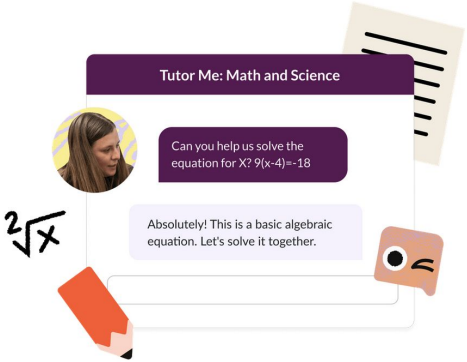
In the meantime…

OCTOBER 1, 2023 | 14 MIN READ

Artificial Intelligence Could Finally Let Us Talk with Animals

AI is poised to revolutionize our understanding of animal communication

BY LOIS PARSHLEY



Tablet

Obverse

1. nin-ukken-ne2
2. u3-na-a-du11
3. 1(barig) sze ur-{d}szul-pa-e3-ra

reverse

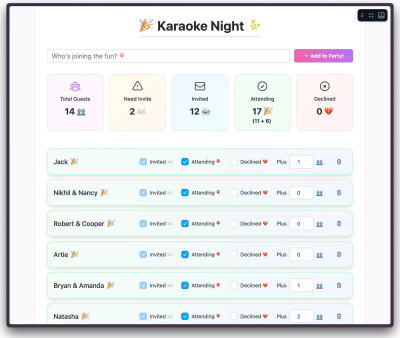
1. he2-na-ab-[szum2]-mu

   blank space
   seal impression

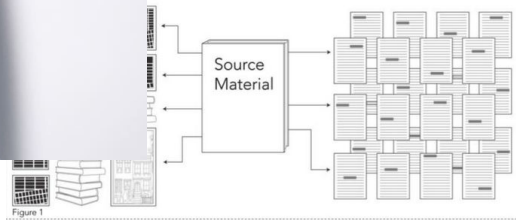
2. [...]

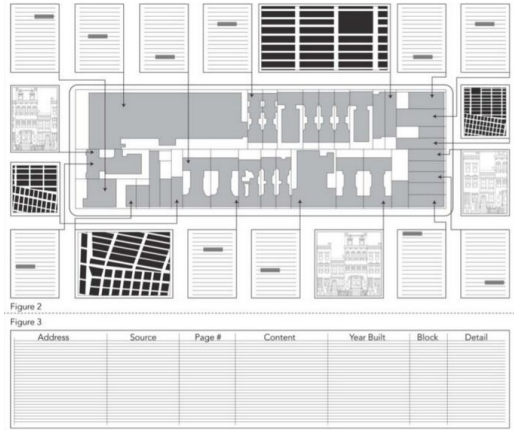
seal 1

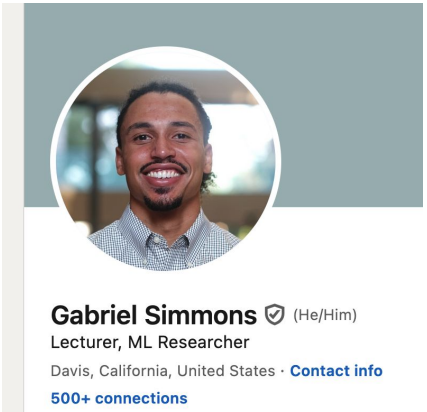
1. lu2-du10-ga
2. dub-sar
3. dumu ur-{d}nin-tu



Thank you!







https://www.linkedin.com/in/gabriel-simmons/

gsimmons@ucdavis.edu