

Introduction to Control and Machine Learning

Mathematical Foundations of Control and Machine Learning

Enrique Zuazua

Winter 2026

FAU Erlangen-Nürnberg

Plan of the lecture

1. **The context today.** Artificial intelligence everywhere: in science, in the economy, in what we see; representation, and the risk of hallucinations.
2. **Origins.** Aristotle; Watt's governor and Maxwell; Wiener's cybernetics and Kalman's theorem; Markov chains and PageRank; herding and parking.
3. **Machine learning objectives.** Express and represent, generalize and forecast, generate and create. Universal approximation, shallow and deep networks, supervised learning, least squares, sampling.
4. **Video material.** The 3Blue1Brown series on neural networks.

Notebook **00** of the course, *The Big Picture: Control and Machine Learning*, complements these slides. The videos of the lecture are not embedded in this PDF: click on a still frame to open the video. The original PowerPoint, with the videos embedded, is on [Google Drive](#) (300 MB).

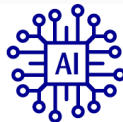
The context today

AI

The image shows the letters 'AI' in a large, bold, black, sans-serif font. The 'A' is a simple, thick letter, and the 'I' is a thick vertical bar. They are centered on a white rectangular background.

Two letters, many meanings:

- Adobe Illustrator
- Artificial Insemination
- Aortic Insufficiency
- Academic Integrity
- All Inclusive
- Automotive Industry
- Aviation Insurance
- Area of Interest
- Annual Income
- Air India



AI

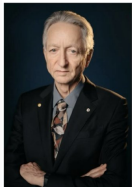


The ubiquity of AI in science

The Nobel Prize in Chemistry 2024

They cracked the code for proteins' amazing structures

The Nobel Prize in Chemistry 2024 is about proteins, life's ingenious chemical tools. David Baker has succeeded with the almost impossible feat of building entirely new kinds of proteins. Demis Hassabis and John Jumper have developed an AI model to solve a 50-year-old problem: predicting proteins' complex structures. These discoveries hold enormous potential.

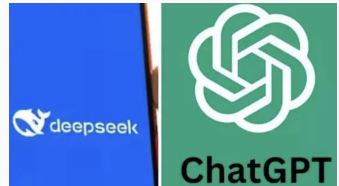


Geoffrey Hinton
Nobel Prize in Physics 2024

Born: 6 December 1947, London, United Kingdom

Affiliation at the time of the award: University of Toronto, Toronto, Canada

Prize motivation: "for foundational discoveries and inventions that enable machine learning with artificial neural networks"



The 2024 Nobel Prizes in Chemistry (Baker, Hassabis, Jumper) and Physics (Hopfield, Hinton) reward machine learning: protein structure prediction and the foundations of artificial neural networks.



This person does not exist



A face generated by a neural network: this-person-does-not-exist.com.

A computer animated hand, 1972

WIKIPEDIA

25 years of the free encyclopedia

Search Wikipedia

Search

Contents

hide

(Top)

Production

Legacy

See also

References

Bibliography

External links

A Computer Animated Hand

ArticleTalk

From Wikipedia, the free encyclopedia

A Computer Animated Hand is a 1972 American [computer-animated](#) short film produced by [Edwin Catmull](#) and [Fred Parke](#). Produced during Catmull's tenure at the [University of Utah](#), the brief was created as part of a [graduate course](#) project. After making a model of his left hand, 350 triangles and polygons were drawn in ink on its surface. The model was digitized from the data and laboriously animated in a three-dimensional animation program that Catmull wrote.

The hand animation consists of three sequences, all of which rotate. The first is the data output of the hand (now called vertices) connected by lines (now called edges), but not filled with *faces*. The second is a [halftone](#) sequence that shows [flat shading](#) but lacks [smooth shading](#). The final, completed animation, featuring organic, smooth shading of the surface between the data points, depicts the hand swiveling, opening and closing, pointing at the viewer, and tilting back, allowing the camera to move to the inside of the hand. The clip also features computer-generated animations of an [artificial heart valve](#) and human faces. Snippets of the animations were used in the 1976 [Hollywood science fiction](#) film *Futureworld*.^[1]

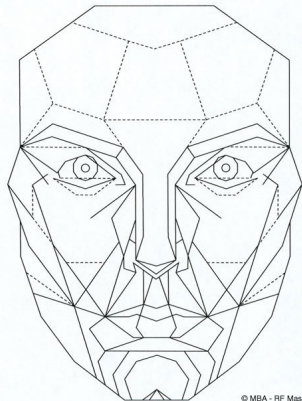
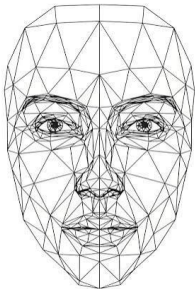
A Computer Animated Hand

The title screen for the short film: *A Computer Animated Hand*

Directed by	Edwin Catmull Fred Parke
Produced by	Edwin Catmull Fred Parke
Release date	1972
Running time	4 minutes (approx.)
Country	United States
Language	Silent (Intertitles in English)

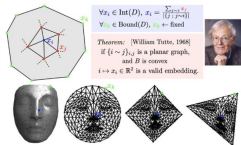
E. Catmull and F. Parke digitize a hand with 350 polygons. [Wikipedia: A Computer Animated Hand](#).

7



© MBA - RF Mask

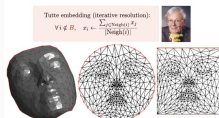
Representing a surface: Tutte embeddings



vignette

Tutte Embedding

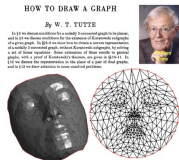
Tutte theorem defines a valid drawing of planar graphs by solving a linear system. A fundamental tool to compute surface parameterization. [link](#)



vignette

Tutte Equation

Tutte embedding is the solution of a Poisson equation. Iterative resolution is equivalent to a system of masses and springs. [link](#) [link](#)



vignette

Tutte Surfaces

Classic reference: William Thomas Tutte, How to draw a graph, 1963. Shows how to embed a planar graph in 2-D by solving a Laplacian linear system. Heavily used in mesh processing to compute surface parameterization. [link](#) [link](#)

Tutte's theorem (1963) draws a planar graph by solving a linear system; it is the basis of surface parameterization in mesh processing. Gabriel Peyré, [Mathematical Nexus](#).

Discrete differential geometry

Keenan Crane
Home News Teaching Publications Code YouTube Twitter CV FAQ Misc

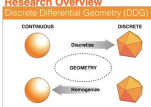


About

I am the Michael B. Donohue Associate Professor of Computer Science and Robotics at Carnegie Mellon University, and a member of the [Center for Nonlinear Analysis](#). I also mentor some incredible students in the [Geometry Collective](#). Our work in the emerging field of [Discrete Differential Geometry](#) uses insights from differential geometry to build fundamental representations and algorithms for processing, designing, and analyzing geometric data. I received my BS from UIUC, was a Google PhD Fellow at Caltech, an NSF Mathematical Sciences Postdoctoral Fellow at Columbia University. I am a Packard Fellow, recipient of an NSF CAREER Award, and a proud dad.

Research Overview

Discrete Differential Geometry (DDG)

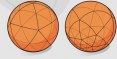


217 Smith Hall
5000 Forbes Ave
Pittsburgh, PA 15213
(412) 268 3454
kmcrcane@cs.cmu.edu



Thesis

Conformal Geometry Processing



Keenan Crane
[Caltech PhD Thesis \(2013\)](#)
[Abstract](#) [PDF](#) [BioTeX](#)

[gghthes15\(Crane 2013\)](#)
author = (Keenan Crane),
title = (Conformal Geometry Processing),
school = (Caltech),
month = (June),
year = (2013)
url = ([https://www.cs.cmu.edu/~keenan/theses/](#))

April 2026 - Graphics Top 10
I was again identified as one of the top 10 researchers in Computer Graphics over the last 5 years, according to quantitative metrics from ScholarGPS.

Teaching

Computer Graphics (15-462/662) **Discrete Differential Geometry (15-458/766)** **Monte Carlo Methods (15-327/627/860, 21-387)**
[Videos](#) [Code](#) [Webpage](#) [Videos](#) [Code \[C++/JS\]](#) [Webpage](#) [Notes](#) [Webpage](#)

Publications

Rectangular Surface Parameterization



Corman, Crane
[ACM Trans. on Graph. \(2025\)](#)
[Abstract](#) [PDF](#) [Supplement](#) [BibTeX](#)
[Best Paper \(Honorable Mention\)](#)

Solving Partial Differential Equations in Participating Media



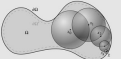
Miller, Sawhney, Crane, Gkioulekas
[ACM Trans. on Graph. \(2025\)](#)
[Abstract](#) [PDF](#) [Project](#) [BibTeX](#)

Putting Rigid Bodies to Rest



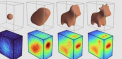
Baktash, Sharp, Zhou, Crane, Jacobson
[ACM Trans. on Graph. \(2025\)](#)
[Abstract](#) [PDF](#) [Data](#) [BibTeX](#)

State of the Art in Grid-Free Monte Carlo Methods for Partial Differential Equations



Sawhney, Miller, Gkioulekas, Crane
[SIGGRAPH Courses \(2025\)](#)
[Abstract](#) [Project](#) [HTML](#) [BibTeX](#)

Differential Walk on Spheres



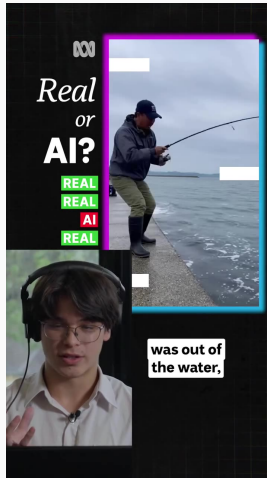
Miller, Sawhney, Crane, Gkioulekas
[ACM Trans. on Graph. \(2024\)](#)
[Abstract](#) [Project](#) [PDF](#) [BibTeX](#)

Keenan Crane, Carnegie Mellon University: geometry processing, Monte Carlo methods for PDEs.

The risk of hallucinations

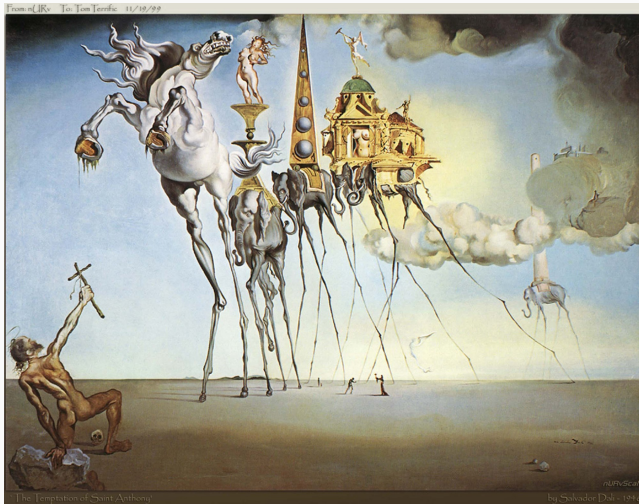


Real or AI?

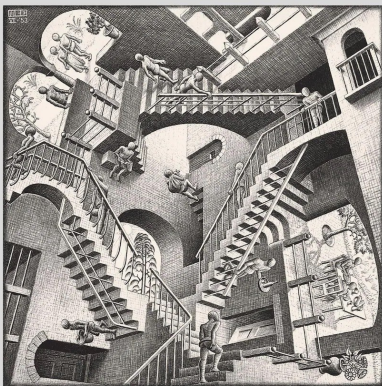


- ABC News (Australia): *Is it real or AI? It's getting harder than ever to tell the difference* (video, 3 min)

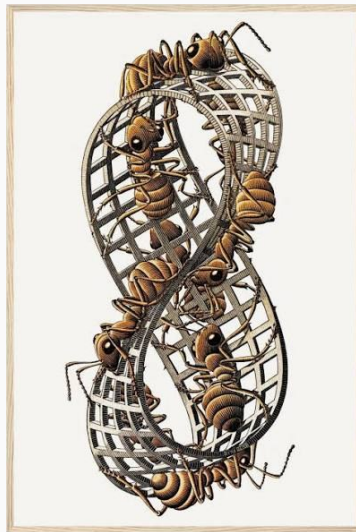
The risk of hallucinations



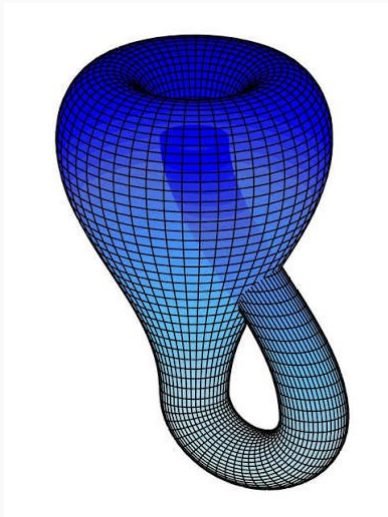
Salvador Dalí, *The Temptation of Saint Anthony* (1946).



M. C. Escher, *Relativity* (1953).



M. C. Escher, *Möbius Strip II* (1963).



The Klein bottle. Felix Klein was professor in Erlangen (1872–1875), where he formulated the Erlangen program.

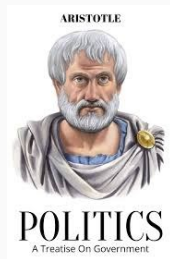
Origins

When did artificial intelligence emerge, or when did the race toward what we now call artificial intelligence begin?

Aristotle

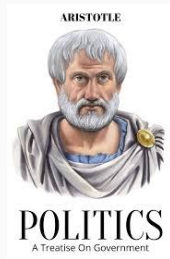
"If every instrument could accomplish its own work, obeying or anticipating the will of others, like the statues of Daedalus or the tripods of Hephaestus. . . then master-craftsmen would have no need of assistants, and masters no need of slaves."

— Aristotle, *Politics*, Book I, Ch. 4 (335–323 BC)



"If every instrument could accomplish its own work, obeying or anticipating the will of others, like the statues of Daedalus or the tripods of Hephaestus. . . then master-craftsmen would have no need of assistants, and masters no need of slaves."

— Aristotle, *Politics*, Book I, Ch. 4 (335–323 BC)



- **Automation.** He imagined a world where tools (instruments) could perform tasks autonomously, anticipating or obeying commands.
- **Liberation from labor.** This automation would eliminate the need for human labor, freeing people from mundane work.
- **Relevance today.** Aristotle's thought experiment foreshadows modern discussions on automation, robotics and artificial intelligence.



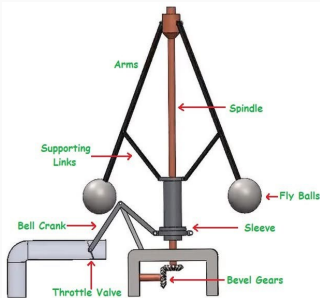
Control in an Information Rich World, SIAM, R. Murray (ed.), 2003.

Maxwell, 1868: *On Governors*

The following communications were read :—

I. “ On Governors.” By J. CLERK MAXWELL, M.A., F.R.SS.L. & E.
Received Feb. 20, 1868.

A Governor is a part of a machine by means of which the velocity of the machine is kept nearly uniform, notwithstanding variations in the driving-power or the resistance.



Mr. J. C. Maxwell on Governors.

March 5, 1868.

Watt's centrifugal governor (1788) keeps the speed of a steam engine nearly uniform.

Maxwell asks when such a feedback device settles and when it oscillates: the first stability analysis.

J. C. Maxwell, *On Governors*, Proc. Royal Society of London 16 (1868), 270–283.





The linear finite d -dimensional system

$$x'(t) = Ax(t) + Bu(t), \quad t \in (0, T); \quad x(0) = x^0, \quad (1)$$

with $m \ll d$ controls. $A \in M_{d \times d}$, $B \in M_{d \times m}$ and $x^0 \in \mathbb{R}^d$; $x : [0, T] \rightarrow \mathbb{R}^d$ represents the **state** and $u : [0, T] \rightarrow \mathbb{R}^m$ the **control**.

Can we control d states with only m controls, even if $d \gg m$?

Theorem (R. E. Kálmán, 1960)

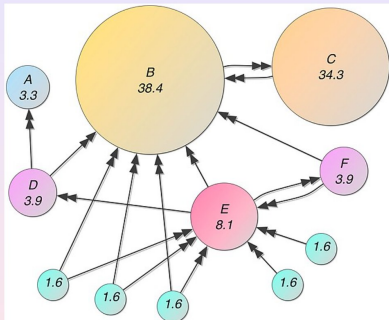
System (1) is controllable if and only if

$$\text{rank} [B, AB, \dots, A^{d-1}B] = d.$$



A new paradigm: mathematics + artificial intelligence

PageRank: Larry Page and Sergey Brin (Google, 1998). The importance of a web page is the stationary distribution of a random walk on the graph of links.



Inspired on the **Perron–Frobenius theorem**: Oskar Perron (1880–1975), Georg Frobenius (1849–1917).

Andrey Markov (1856–1922)

- Born 14 June 1856 in Ryazan; died 20 July 1922 in Petrograd. St. Petersburg University.
- Known for **Markov chains** and **Markov processes**: random sequences in which the next state depends only on the present one.
- Markov chains are the mathematics behind PageRank and the simplest example of a stochastic dynamical system (lecture 01).

[Wikipedia: Andrey Markov.](#)

Andrey Markov

Андрей Марков



A. A. Markov (1886).

Markov in 1886

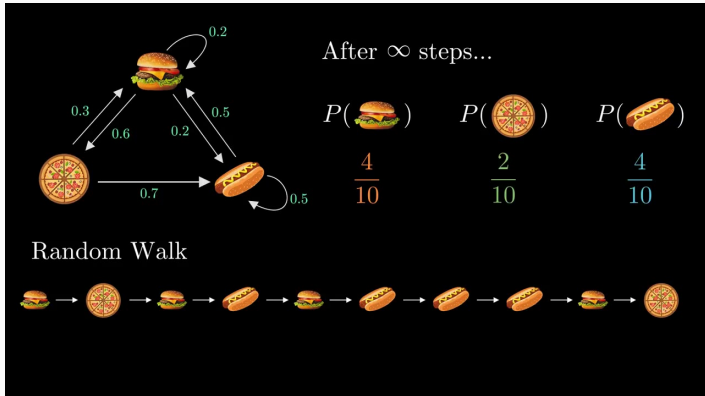
Born 14 June 1856
[Ryazan, Russia](#)

Died 20 July 1922 (aged 66)
[Petrograd, Russia](#)

Alma mater [St. Petersburg University](#)

Known for [Markov chains](#)
[Markov processes](#)

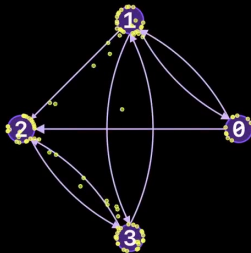
Video: introduction to Markov chains



► Normalized Nerd: *Markov Chains Clearly Explained! Part 1* (video, 9 min)

Video: Google's algorithm through Markov chains

How to Measure Relative Importance?



► Reducible: *PageRank: A Trillion Dollar Algorithm* (video, 25 min)

E. Hairer
G. Wanner

UNDERGRADUATE TEXTS IN MATHEMATICS

Analysis by Its History

 Springer

Analysis by its history

... I should like to bring up again for emphasis ... points, in which my exposition differs especially from the customary presentation in the textbooks: 1. Illustration of abstract considerations by means of figures. 2. Emphasis upon its relation to neighboring fields, such as calculus of differences and interpolation. . . 3. Emphasis upon historical growth. It seems to me extremely important that precisely the prospective teacher should take account of all of these.
(F. Klein 1908, Engl. ed. p. 236)

Traditionally, a rigorous first course in Analysis progresses (more or less) in the following order:

sets, mappings $\Rightarrow$ limits, continuous functions $\Rightarrow$ derivatives $\Rightarrow$ integration.

On the other hand, the historical development of these subjects occurred in reverse order:

Cantor 1875	$\Leftarrow$	Cauchy 1821	$\Leftarrow$	Newton 1665	$\Leftarrow$	Archimedes
Dedekind		Weierstrass		Leibniz 1675		Kepler 1615
						Fermat 1638

E. Hairer and G. Wanner, *Analysis by Its History*, Undergraduate Texts in Mathematics, Springer, 1996 (preface).

True herding



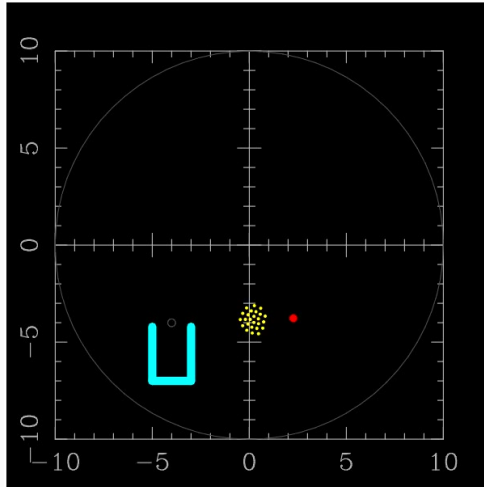
- A sheepdog trial: the shepherd steers the flock through the dog (video, 1:51)

Robotic herding



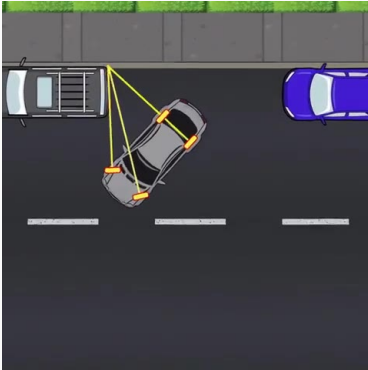
- University of the Sunshine Coast: *Robotic dog herding sheep in regional Queensland* (video, 15 s)

Virtual herding



► Simulation: one controlled agent (red) drives the flock (yellow) into the pen (video, 20 s)

An example: Nelson's car



► Parallel parking (simulation, 40 s)

The state is four-dimensional, position (x_1, x_2) , orientation θ and steering angle φ , and there are two controls, the speed and the rotation of the wheel. **Two controls suffice to control a four-dimensional dynamical system.**

E. Sontag, *Mathematical Control Theory*, 2nd ed., Springer-Verlag, New York, 1998, Figure 4.1.

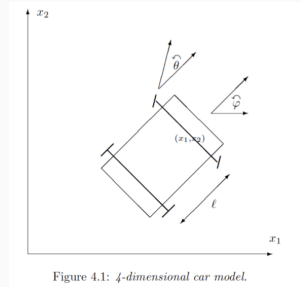


Figure 4.1: 4-dimensional car model.



► Automatic parking (video, 30 s)

Machine learning objectives

- Express / Represent
- Generalize / Forecast
- Generate / Create

Express / Represent

Why does it work? Universal approximation

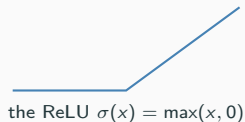
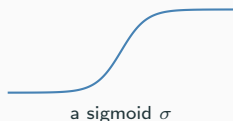
N. Wiener, Tauberian theorems, *Annals of Mathematics* 33 (1) (1932), 1–100.

$$f(x) \sim \sum_{j=1}^K w_j G(x + b_j)$$

G. Cybenko, Approximation by superpositions of a sigmoidal function, *Mathematics of Control, Signals and Systems* 2 (1989), 303–314.

$$\lim_{x \rightarrow -\infty} \sigma(x) = 0, \quad \lim_{x \rightarrow +\infty} \sigma(x) = 1$$

$$f(x) \sim \sum_{j=1}^K w_j \sigma(a_j \cdot x + b_j)$$



Neural networks: shallow versus deep

$$f(x) \sim \sum_{j=1}^K w_j \sigma(a_j \cdot x + b_j)$$

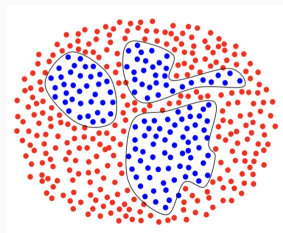
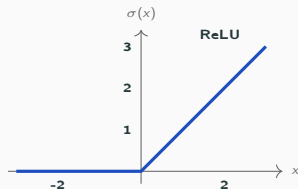


$$x^{k+1} = x^k + h w^k \sigma(a^k \cdot x^k + b^k)$$



$$\frac{dx(t)}{dt} = w(t) \sigma(a(t) \cdot x(t) + b(t))$$

A shallow network is a sum of K neurons; a deep (residual) network composes layers, $k = 0, \dots, L$; as the step $h \rightarrow 0$ the layers become the time of a **neural ODE**.

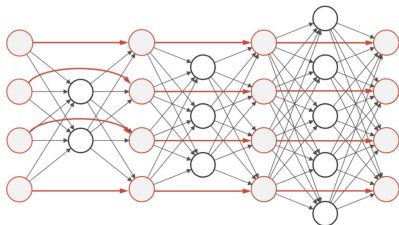


Classification: separate the blue from the red points.

Supervised learning: computational practice

$$\underbrace{\frac{1}{N} \sum_{i=1}^N \text{loss}(x^i, \ell^i)}_{\text{empirical risk} := E(x(\cdot))} + \alpha \sum_{j=1}^K \|(\mathbf{a}_j, \mathbf{w}_j, \mathbf{b}_j)\|^2$$

$$\underbrace{\frac{1}{N} \sum_{i=1}^N \text{loss}(x^i, \ell^i)}_{\text{empirical risk} := E(x(\cdot))} + \alpha \sum_{j=1}^K \|(\mathbf{a}_j, \mathbf{w}_j, \mathbf{b}_j)\|^2$$



Given N samples x^i with labels ℓ^i , choose the parameters $\theta = (a, b, w)$ that minimize the empirical risk, plus a penalty on their size.

GPT-3 (2020)

175 billion parameters

$\theta = (a, b, w)$

$N = 300$ billion tokens in the training corpus

dimension $d = 8,192$

Least squares

The objective is to adjust the parameters β of a model function $f(x, \beta)$ to best fit a data set (x_i, y_i) , $i = 1, \dots, n$. The fit at a data point is measured by its **residual**

$$r_i = y_i - f(x_i, \beta).$$

The least-squares method finds the optimal parameters by minimizing the **sum of squared residuals**

$$S(\beta) = \sum_{i=1}^n r_i^2.$$

- In the simplest case $f(x, \beta) = \beta$ and the minimizer is the arithmetic mean of the y_i .
- For a straight line, $f(x, \beta) = \beta_0 + \beta_1 x$: *linear regression*.

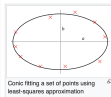
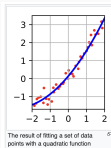
Legendre (1805) and Gauss (1809). [Wikipedia: Least squares](#).



A.-M. Legendre



C. F. Gauss



Sampling (statistics)

🌐 41 languages ▾

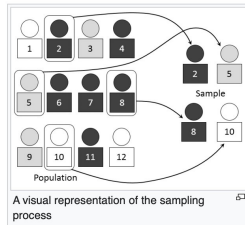
[Article](#) [Talk](#)

[Read](#) [Edit](#) [View history](#) [Tools](#) ▾

From Wikipedia, the free encyclopedia

For other uses, see [Sampling \(disambiguation\)](#).

In this [statistics](#), [quality assurance](#), and [survey methodology](#), **sampling** is the selection of a subset or a **statistical sample** (termed **sample** for short) of individuals from within a [statistical population](#) to estimate characteristics of the whole population. The subset is meant to reflect the whole population, and statisticians attempt to collect samples that are representative of the population. Sampling has lower costs and faster data collection compared to recording data from the entire population (in many cases, collecting the whole population is impossible, like getting sizes of all stars in the universe), and thus, it can provide insights in cases where it is infeasible to measure an entire population.

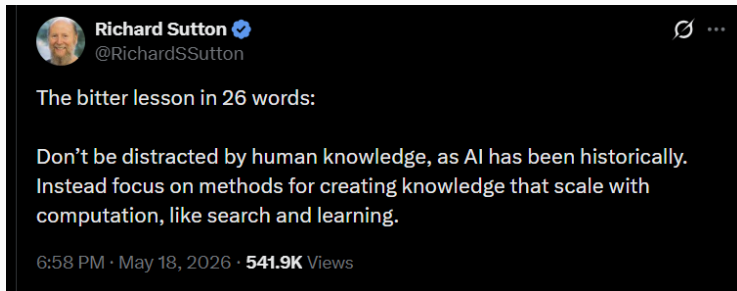


Each [observation](#) measures one or more properties (such as weight, location, colour or mass) of independent objects or individuals. In [survey sampling](#), weights can be applied to the data to adjust for the sample design, particularly in [stratified sampling](#).^[1] Results from [probability theory](#) and [statistical theory](#) are employed to guide the practice. In business and medical research, sampling is widely used for gathering information about a population.^[2] [Acceptance sampling](#) is used to determine if a production lot of material meets the governing [specifications](#).

A model is trained on a **sample** of the population it should describe. [Wikipedia: Sampling \(statistics\)](#).

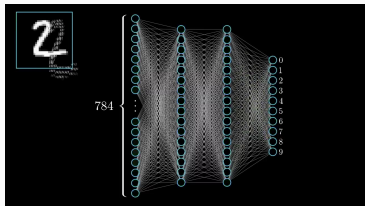
The bitter lesson (Rich Sutton, 2019)

The biggest progress in AI has come from **general methods that leverage computation**, rather than methods that try to incorporate human knowledge.

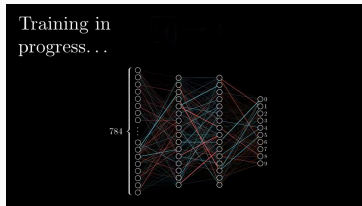


R. Sutton, *The Bitter Lesson* (2019), and the summary posted by its author in 2026.

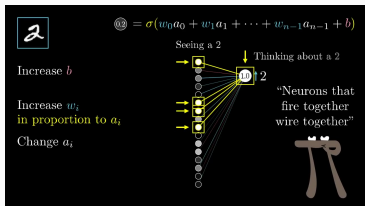
Video material: 3Blue1Brown, *Neural networks*



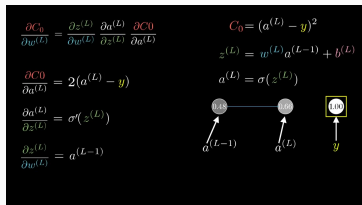
- 1. But what is a neural network? (19 min)



- 2. Gradient descent, how neural networks learn (21 min)



- 3. What is backpropagation really doing? (13 min)



- 4. Backpropagation calculus (10 min)