Why Should You Care About Math?

Why Should You Care About Math?

4,000 YEARS

Babylonians solved quadratics. Greeks proved theorems. Newton invented calculus to predict planets.

Why Should You Care About Math?

4,000 Years

Babylonians solved quadratics. Greeks proved theorems. Newton invented calculus to predict planets.

Hidden Power

GPS needs relativity. Spotify uses linear algebra. Your phone camera runs Fourier transforms.

Why Should You Care About Math?

4,000 Years

Babylonians solved quadratics. Greeks proved theorems. Newton invented calculus to predict planets.

Hidden Power

GPS needs relativity. Spotify uses linear algebra. Your phone camera runs Fourier transforms.

2017 $\longrightarrow$ Now

One paper — *"Attention Is All You Need"* — launched ChatGPT, Claude, Gemini, and a $3 trillion industry.

Why Should You Care About Math?

4,000 Years

Babylonians solved quadratics. Greeks proved theorems. Newton invented calculus to predict planets.

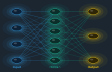
Hidden Power

GPS needs relativity. Spotify uses linear algebra. Your phone camera runs Fourier transforms.

2017 → Now

One paper — *"Attention Is All You Need"* — launched ChatGPT, Claude, Gemini, and a $3 trillion industry.

The secret? Every breakthrough in AI is built on math that already existed — most of it centuries old. Today we'll trace five mathematical ideas from ancient history to the AI running on your phone right now.



THE FIVE PILLARS

The Code of the Universe

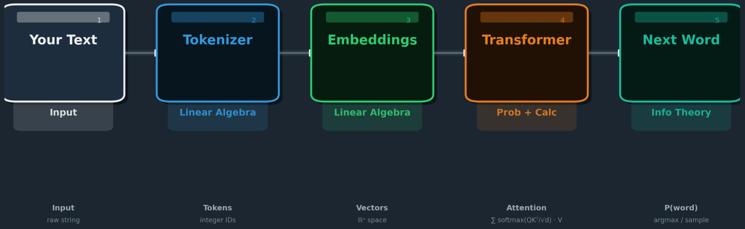
From Classical Mathematics to Large Language Models

The Five Mathematical Ideas Inside Every AI You Use

LLM Token Processing Pipeline

How an LLM Processes Text

Five mathematical stages from raw input to predicted next word



What Happens When You Ask ChatGPT a Question?

What Happens When You Ask ChatGPT a Question?

1. Your words become vectors, then giant matrices multiply — Linear Algebra

What Happens When You Ask ChatGPT a Question?

1. Your words become vectors, then giant matrices multiply — Linear Algebra
2. It computes probabilities for every possible next word — Probability

What Happens When You Ask ChatGPT a Question?

1. Your words become vectors, then giant matrices multiply — Linear Algebra
2. It computes probabilities for every possible next word — Probability
3. It learned from trillions of examples using derivatives — Calculus

What Happens When You Ask ChatGPT a Question?

1. Your words become vectors, then giant matrices multiply — Linear Algebra
2. It computes probabilities for every possible next word — Probability
3. It learned from trillions of examples using derivatives — Calculus
4. The goal: minimize surprise (cross-entropy) — Information Theory

What Happens When You Ask ChatGPT a Question?

1. Your words become vectors, then giant matrices multiply — Linear Algebra
2. It computes probabilities for every possible next word — Probability
3. It learned from trillions of examples using derivatives — Calculus
4. The goal: minimize surprise (cross-entropy) — Information Theory
5. Optimizers like Adam make trillion-parameter training possible — Optimization

How LLMs Work — Visual Intro

Live Interactive Demo

3Blue1Brown — “Large Language Models explained briefly” (∼5 min)

`https://www.youtube.com/watch?v=LPZh9BOjkQs`

Transformer Explainer — Live GPT-2

Live Interactive Demo

Type text, see tokens $\rightarrow$ embeddings $\rightarrow$ attention $\rightarrow$ prediction

`https://poloclub.github.io/transformer-explainer/`

3D LLM Visualization

Live Interactive Demo

3D walkthrough of every matrix operation in a GPT model

`https://bbycroft.net/llm`

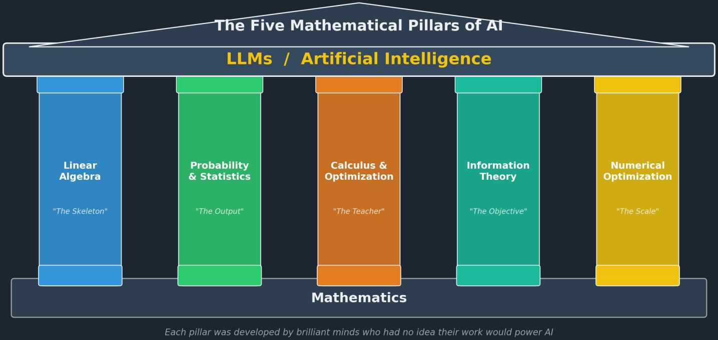
AnimatedLLM — Step by Step

Live Interactive Demo

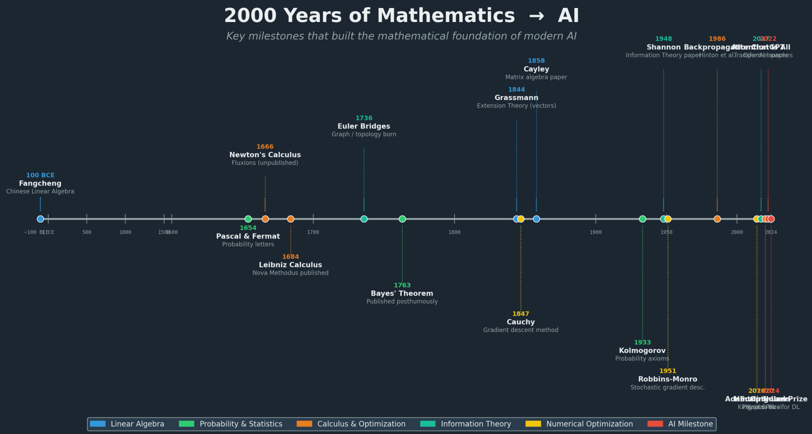
Step-by-step animation of text generation & training

`https://animatedllm.github.io/`

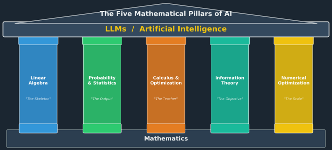
The Five Pillars of AI Mathematics



Mathematical Timeline: 100 BCE to 2024

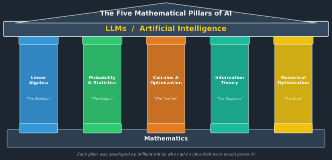


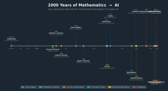
The Five Pillars



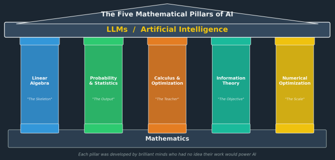
Each pillar was developed by brilliant minds who had no idea their work would power AI

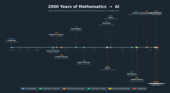
The Five Pillars





The Five Pillars





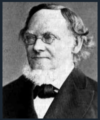
Each pillar was developed by brilliant minds who had no idea their work would power AI.

PILLAR 1

Linear Algebra

The Skeleton of AI

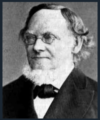
2000 Years of Linear Algebra



Hermann Grassmann

P1

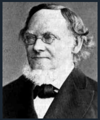
2000 Years of Linear Algebra



Hermann Grassmann

$\sim$**100 BCE** Chinese *Fangcheng* — solving systems with counting rods ORIGIN

P1 2000 Years of Linear Algebra



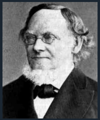
Hermann Grassmann

∼100 BCE Chinese *Fangcheng* — solving systems with counting rods ORIGIN

1844 Grassmann publishes vector spaces — universally ignored

P1

2000 Years of Linear Algebra



Hermann Grassmann

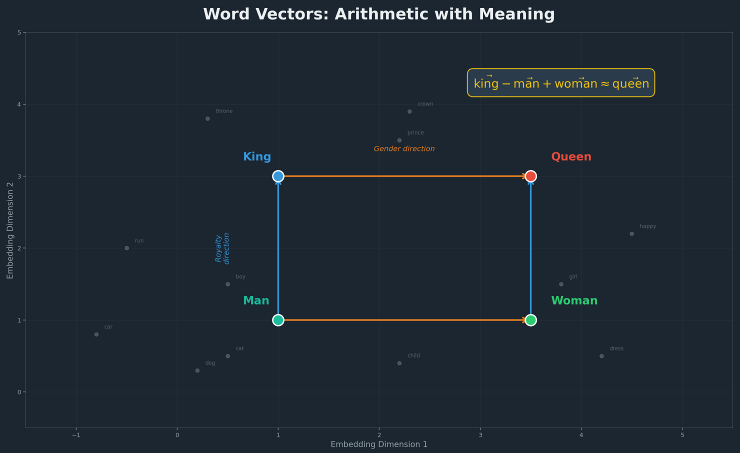
∼100 BCE Chinese *Fangcheng* — solving systems with counting rods ORIGIN

1844 Grassmann publishes vector spaces — universally ignored

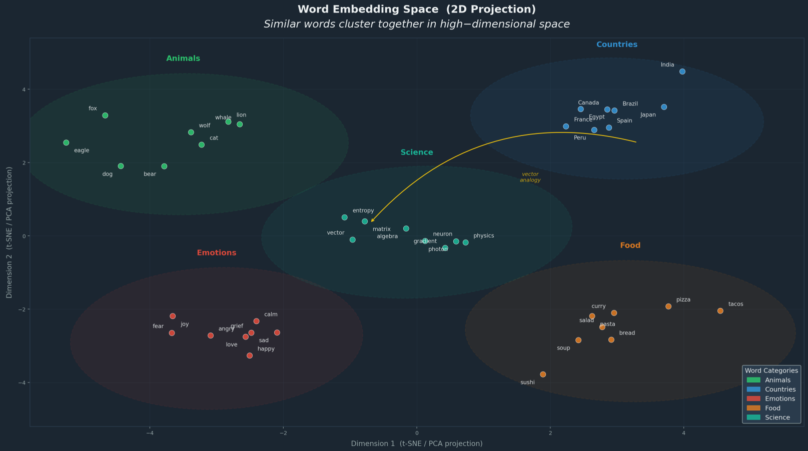
1858 Cayley invents matrix theory — while working as a lawyer

Arthur Cayley

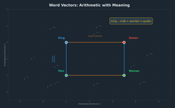
Word Vectors: King − Man + Woman = Queen



2D Word Embedding Space

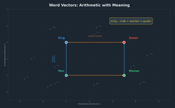


P1 Words as Vectors AI Connection

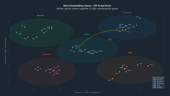


$$\vec{\text{king}} - \vec{\text{man}} + \vec{\text{woman}} \approx \vec{\text{queen}}$$

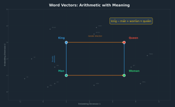
P1 Words as Vectors AI CONNECTION



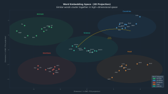
$$\vec{\text{king}} - \vec{\text{man}} + \vec{\text{woman}} \approx \vec{\text{queen}}$$



P1 Words as Vectors AI CONNECTION

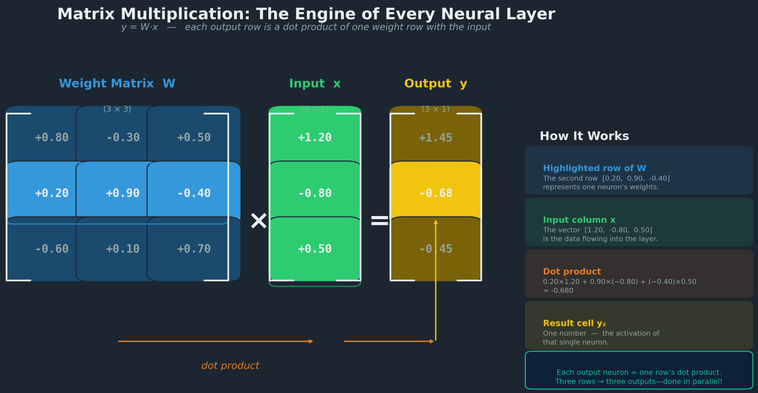


$$\vec{\text{king}} - \vec{\text{man}} + \vec{\text{woman}} \approx \vec{\text{queen}}$$



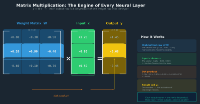
Mikolov et al., 2013 — Word2Vec: meaning encoded as geometry.

Matrix Multiplication: The Core Operation



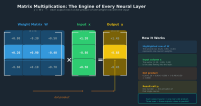
The Engine: Matrix Multiplication

$$\text{output} = W \cdot \vec{x} + \vec{b}$$



The Engine: Matrix Multiplication

$$\text{output} = W \cdot \vec{x} + \vec{b}$$

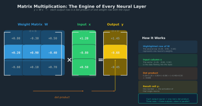


Every layer: multiply input vector by weight matrix

P1

The Engine: Matrix Multiplication

$$\text{output} = W \cdot \vec{x} + \vec{b}$$



Every layer: multiply input vector by weight matrix

GPT-4: ∼1.8 trillion such multiplications per token

Attention Weight Heatmap

Attention: Which Words Look at Which?

Each word decides which other words are relevant ("Attention Is All You Need", 2017)

How to Read This

"The cat sat on the mat"

"cat" → The, sat

"cat" checks the article before it and the verb it performs

"sat" → cat, on

"sat" needs to know who is sitting and the location preposition

"mat" → on, the

"mat" ties back to "on" and the article "the" before it

Diagonal ≈ 0.18–0.25

Every word attends to itself with moderate weight

Bright cells = strong attention

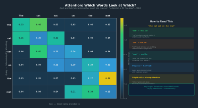
Yellow / green cells show the most important connections

Key insight: Transformers learn WHICH words to pay attention to — automatically!

P1 Attention = Three Matrix Multiplies BREAKTHROUGH

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V$$

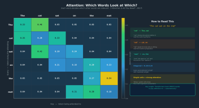
$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$



P1 Attention = Three Matrix Multiplies BREAKTHROUGH

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$



Three matrices. That's all attention is. Cayley's invention, applied to language.

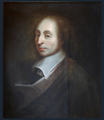
PILLAR 2

Probability & Statistics

The Language of Uncertainty

P2

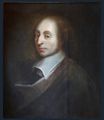
Born from Gambling



Blaise Pascal

Born from Gambling

1654 Pascal & Fermat exchange letters about a gambling problem ORIGIN



Blaise Pascal

Fermat

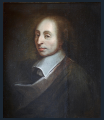
Bayes

Kolmogorov

Born from Gambling

1654 Pascal & Fermat exchange letters about a gambling problem ORIGIN

1763 Bayes' theorem published posthumously



Blaise Pascal

Fermat

Bayes

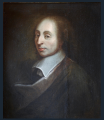
Kolmogorov

P2 Born from Gambling

1654 Pascal & Fermat exchange letters about a gambling problem ORIGIN

1763 Bayes' theorem published posthumously

1933 Kolmogorov writes the axioms — probability becomes rigorous



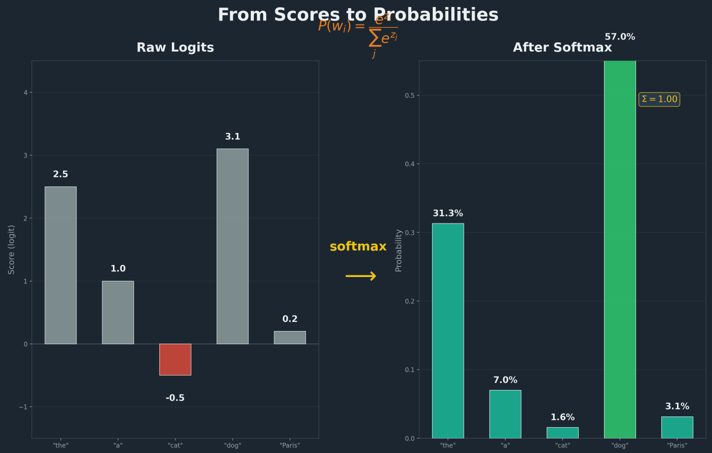
Blaise Pascal

Fermat

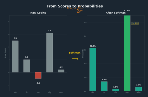
Bayes

Kolmogorov

Softmax: From Logits to Probabilities

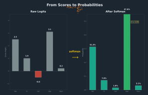


Turning Scores into Probabilities



$$P(w_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

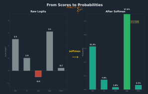
Turning Scores into Probabilities



$$P(w_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

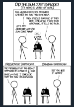
50,000+ words. One probability each. Kolmogorov's axioms in action.

Turning Scores into Probabilities



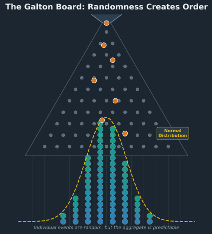
$$P(w_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

50,000+ words. One probability each. Kolmogorov's axioms in action.



xkcd.com/1132 (CC BY-NC 2.5)

Randomness Creates Order

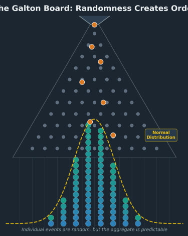


Wikimedia Commons (CC BY-SA 4.0)

Individual events are random.

But the aggregate forms a pattern.

Randomness Creates Order



Wikimedia Commons (CC BY-SA 4.0)

Individual events are random.

But the aggregate forms a pattern.

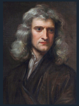
LLMs work the same way: each token is sampled randomly, but the sequence is coherent.

PILLAR 3

Calculus & Optimization

The Teacher

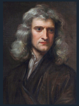
P3 The Calculus Wars ORIGIN



Newton (1666)

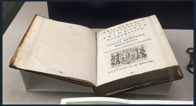
Leibniz (1684)

P3 **The Calculus Wars** ORIGIN



Newton (1666)

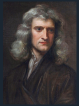
Leibniz (1684)



Principia Mathematica, 1713 ed.

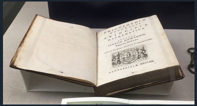
The Calculus Wars

ORIGIN



Newton (1666)

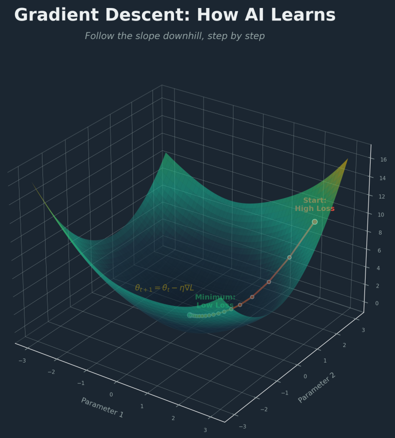
Leibniz (1684)



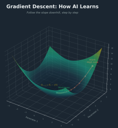
Principia Mathematica, 1713 ed.

Both invented calculus independently. We use Leibniz's notation: $\dfrac{dy}{dx}$

Gradient Descent on a Loss Surface

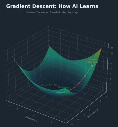


Gradient Descent: How AI Learns



Cauchy (1847)

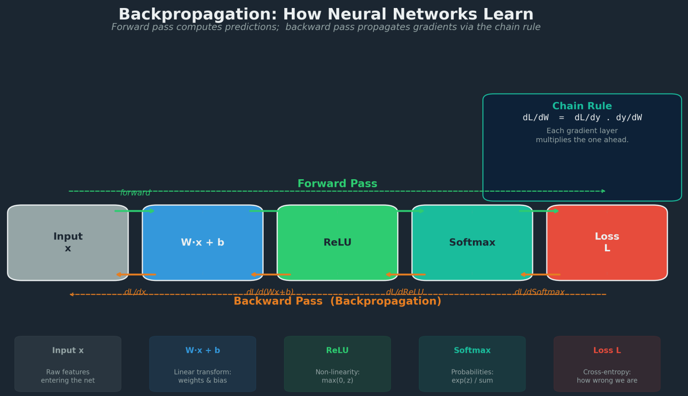
Gradient Descent: How AI Learns



Cauchy (1847)

Cauchy invented this in 1847 — for tracking planetary orbits.

Backpropagation: Forward and Backward Pass

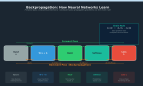


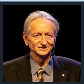
P3

Backpropagation = The Chain Rule

BREAKTHROUGH

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial w}$$

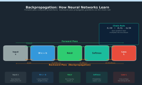


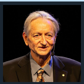


Hinton (Nobel 2024)

Backpropagation = The Chain Rule BREAKTHROUGH

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial w}$$



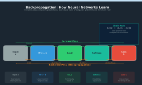


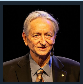
Hinton (Nobel 2024)

The chain rule: derivatives flow backward through every layer

P3 Backpropagation = The Chain Rule BREAKTHROUGH

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial w}$$





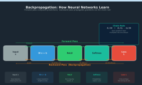
Hinton (Nobel 2024)

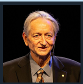
The chain rule: derivatives flow backward through every layer

1986: Rumelhart, Hinton & Williams publish in *Nature*

P3 Backpropagation = The Chain Rule BREAKTHROUGH

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial w}$$





Hinton (Nobel 2024)

The chain rule: derivatives flow backward through every layer

1986: Rumelhart, Hinton & Williams publish in *Nature*

2024: Hinton wins the Nobel Prize in Physics

PILLAR 4

Information Theory

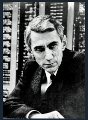
The Objective Function

P4

Shannon: Father of Information Theory

ORIGIN

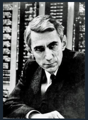
> *"Information is the resolution of uncertainty."*
>
> — Claude Shannon



Claude Shannon

P4 Shannon: Father of Information Theory ORIGIN

> *"Information is the resolution of uncertainty."*
>
> — Claude Shannon

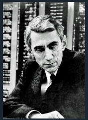


Claude Shannon

1948: "A Mathematical Theory of Communication" — invented the **bit**

P4 Shannon: Father of Information Theory ORIGIN

> *"Information is the resolution of uncertainty."*
>
> — Claude Shannon

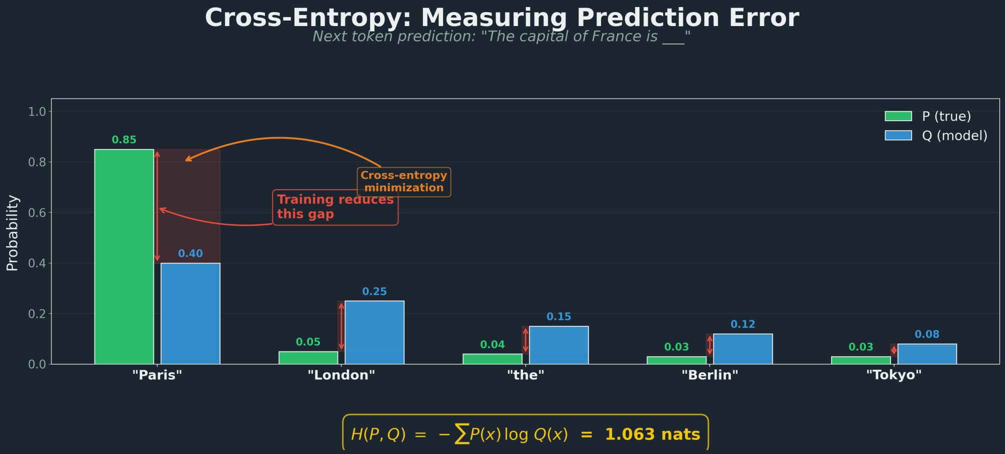


Claude Shannon

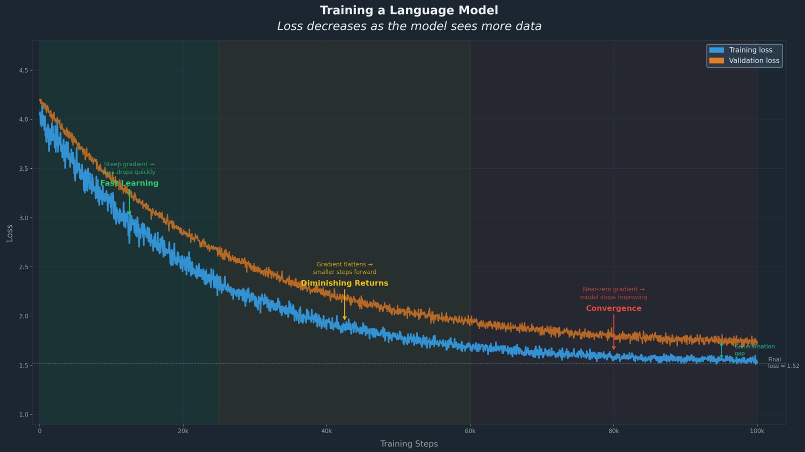
1948: "A Mathematical Theory of Communication" — invented the **bit**

Fun fact: Shannon juggled while riding a unicycle through Bell Labs

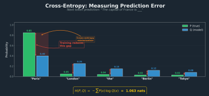
Cross-Entropy: Predicted vs True Distribution



Training Loss Curve Over Time

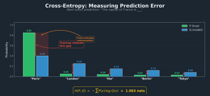


P4 Cross-Entropy: The LLM Loss Function AI CONNECTION



$$H(P,Q) = -\sum_x P(x) \log Q(x)$$

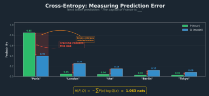
P4 Cross-Entropy: The LLM Loss Function AI CONNECTION



$$H(P, Q) \quad = \quad -\sum_x P(x) \log Q(x)$$

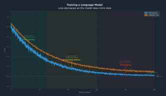
Shannon's 1948 formula IS the training objective of every LLM.

P4 Cross-Entropy: The LLM Loss Function AI CONNECTION



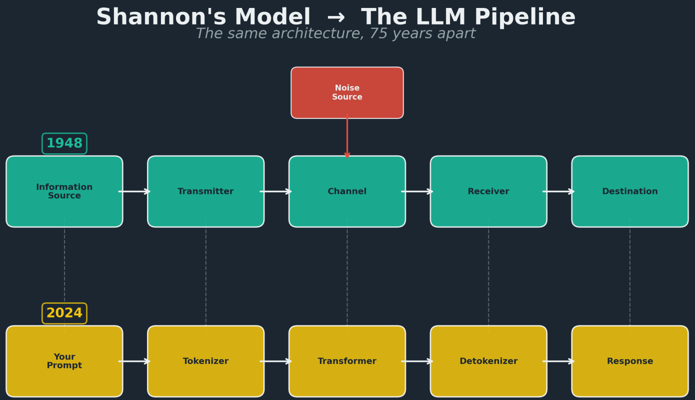
$$H(P, Q) = -\sum_x P(x) \log Q(x)$$

Shannon's 1948 formula IS the training objective of every LLM.

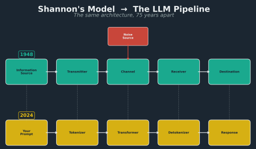


Training loss decreasing over time

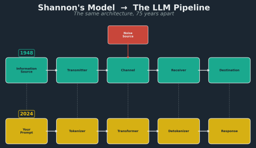
Shannon's Communication Model as LLM Pipeline



P4 Shannon’s Model → The LLM Pipeline



P4 Shannon’s Model → The LLM Pipeline



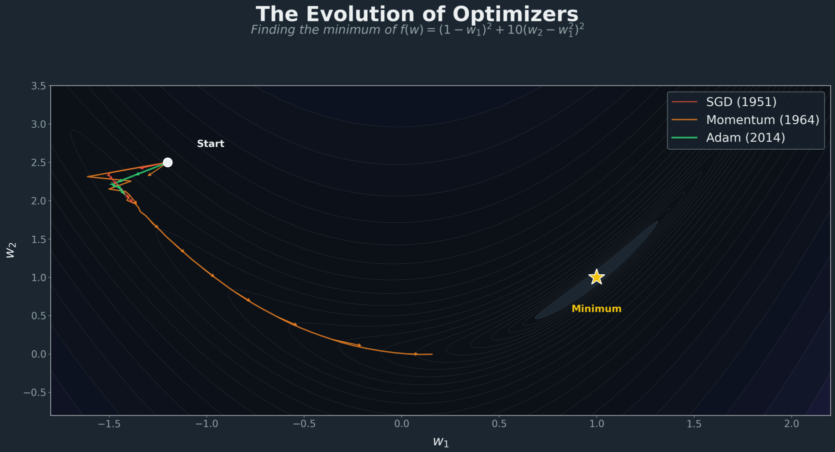
Shannon designed this for telephone lines. 75 years later, it describes exactly how ChatGPT works.

PILLAR 5

Numerical Optimization

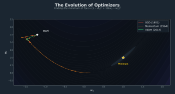
Training at Scale

SGD to Momentum to Adam

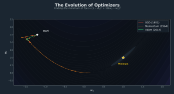


The Evolution of Optimizers

DISCOVERY



P5 The Evolution of Optimizers DISCOVERY

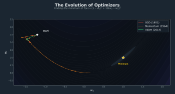


1951: Robbins & Monro invent SGD

P5

The Evolution of Optimizers

DISCOVERY



1951: Robbins & Monro invent SGD

1964: Polyak adds momentum

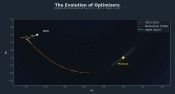
P5 The Evolution of Optimizers DISCOVERY

1951: Robbins & Monro invent SGD

1964: Polyak adds momentum

2014: Kingma & Ba create Adam — **200,000+ citations**

P5 The Evolution of Optimizers DISCOVERY

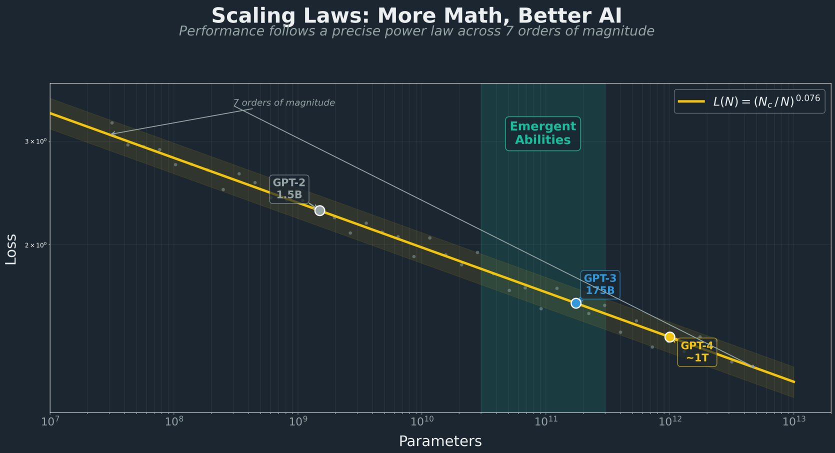


1951: Robbins & Monro invent SGD

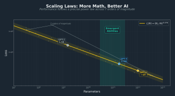
1964: Polyak adds momentum

2014: Kingma & Ba create Adam — **200,000+ citations**

Neural Scaling Laws (Kaplan et al., 2020)

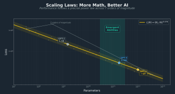


P5 More Math, Better AI AI CONNECTION



$$L(N) = \left(\frac{N_c}{N}\right)^{0.076}$$

P5 More Math, Better AI AI Connection



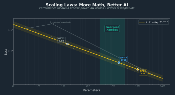
$$L(N) = \left(\frac{N_c}{N}\right)^{0.076}$$

Kaplan et al., 2020 — why companies spend billions on bigger models.

P5

More Math, Better AI

AI Connection

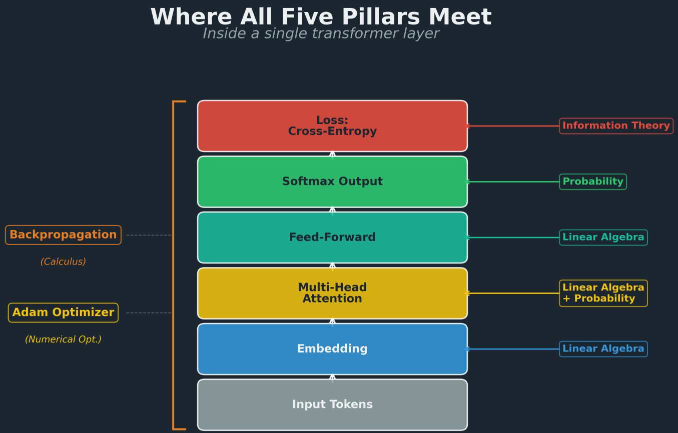


$$L(N) = \left(\frac{N_c}{N}\right)^{0.076}$$

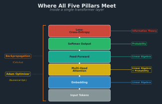
Kaplan et al., 2020 — why companies spend billions on bigger models.

xkcd.com/2048 (CC BY-NC 2.5)

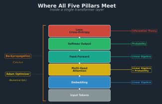
All Five Pillars in One Forward-Backward Pass



Where All Five Pillars Meet

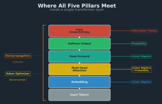


Where All Five Pillars Meet



Linear Algebra	Probability	Calculus	Info Theory	Optimization
Skeleton	Language	Teacher	Objective	Scale

Where All Five Pillars Meet



Linear Algebra Skeleton	**Probability** Language	**Calculus** Teacher	**Info Theory** Objective	**Optimization** Scale

Five branches of pure mathematics, developed over 2000 years, all running simultaneously in a single forward-backward pass.

What LLMs Can Actually Do BREAKTHROUGH

> *"An AI won a Nobel Prize and a Math Olympiad gold medal. In back-to-back years."*

What LLMs Can Actually Do BREAKTHROUGH

> *"An AI won a Nobel Prize and a Math Olympiad gold medal. In back-to-back years."*

35/42

IMO GOLD MEDAL 2025

Gemini Deep Think — only 67 of 630 humans earned gold

What LLMs Can Actually Do BREAKTHROUGH

> *"An AI won a Nobel Prize and a Math Olympiad gold medal. In back-to-back years."*

35/42

IMO GOLD MEDAL 2025

Gemini Deep Think — only 67 of 630 humans earned gold

Nobel

CHEMISTRY 2024

AlphaFold solved 50-year protein folding problem

What LLMs Can Actually Do BREAKTHROUGH

> *"An AI won a Nobel Prize and a Math Olympiad gold medal. In back-to-back years."*

35/42

IMO Gold Medal 2025

Gemini Deep Think — only 67 of 630 humans earned gold

Nobel

Chemistry 2024

AlphaFold solved 50-year protein folding problem

92%

HumanEval Coding

Claude on standard benchmark

What LLMs Can Actually Do BREAKTHROUGH

> *"An AI won a Nobel Prize and a Math Olympiad gold medal. In back-to-back years."*

35/42

IMO GOLD MEDAL 2025

Gemini Deep Think — only 67 of 630 humans earned gold

Nobel

CHEMISTRY 2024

AlphaFold solved 50-year protein folding problem

92%

HUMANEVAL CODING

Claude on standard benchmark

These are not predictions. These already happened.

The Numbers Are Stupid Big

The Numbers Are Stupid Big

1.7T

Parameters in GPT-4

The Numbers Are Stupid Big

1.7T

Parameters in GPT-4

$100M+

Training cost

25,000 GPUs for 90 days

The Numbers Are Stupid Big

1.7T

Parameters in GPT-4

$100M+

Training cost

25,000 GPUs for 90 days

15T

Training tokens

= 2,750 Wikipedias = 84,000 years of reading

The Numbers Are Stupid Big

1.7T

Parameters in GPT-4

$100M+

Training cost

25,000 GPUs for 90 days

15T

Training tokens

= 2,750 Wikipedias = 84,000 years of reading

800M

Weekly ChatGPT users

1 in 10 humans on Earth (Oct 2025)

The Numbers Are Stupid Big

1.7T

Parameters in GPT-4

15T

Training tokens

= 2,750 Wikipedias = 84,000 years of reading

$100M+

Training cost

25,000 GPUs for 90 days

800M

Weekly ChatGPT users

1 in 10 humans on Earth (Oct 2025)

Plot twist: DeepSeek R1 matched GPT-4 performance for $6 million. Open source.

Brilliant and Broken

> *“It solved an IMO problem but can’t count the letters in “strawberry.” Both true.”*

Brilliant and Broken

> *"It solved an IMO problem but can't count the letters in "strawberry." Both true."*

Strawberry: Says there are 2 R's in "strawberry"

9.11 vs 9.9: Many LLMs claim 9.11 $>$ 9.9

Brilliant and Broken

> *"It solved an IMO problem but can't count the letters in "strawberry." Both true."*

Strawberry: Says there are 2 R's in "strawberry"

9.11 vs 9.9: Many LLMs claim 9.11 > 9.9

Mata v. Avianca: Lawyer fined $5K for fake citations

Air Canada: Ordered to honor a hallucinated refund policy

Brilliant and Broken

> *"It solved an IMO problem but can't count the letters in "strawberry." Both true."*

Strawberry: Says there are 2 R's in "strawberry"

9.11 vs 9.9: Many LLMs claim 9.11 > 9.9

Mata v. Avianca: Lawyer fined $5K for fake citations

Air Canada: Ordered to honor a hallucinated refund policy

Why? LLMs are statistical pattern completers, not fact databases. There is no internal fact-checker.

The Race — Zero to Gold in 8 Years

The Race — Zero to Gold in 8 Years

2017 **“Attention Is All You Need” paper** ORIGIN

The Race — Zero to Gold in 8 Years

2017 “Attention Is All You Need” paper ORIGIN

2020 GPT-3 launches — 175 billion parameters

The Race — Zero to Gold in 8 Years

2017 "Attention Is All You Need" paper ORIGIN

2020 GPT-3 launches — 175 billion parameters

Nov 2022 ChatGPT: 100M users in 2 months (Instagram took 2.5 years)

The Race — Zero to Gold in 8 Years

2017 "Attention Is All You Need" paper ORIGIN

2020 GPT-3 launches — 175 billion parameters

Nov 2022 ChatGPT: 100M users in 2 months (Instagram took 2.5 years)

Mar 2023 GPT-4 passes the bar exam BREAKTHROUGH

The Race — Zero to Gold in 8 Years

2017 "Attention Is All You Need" paper ORIGIN

2020 GPT-3 launches — 175 billion parameters

Nov 2022 ChatGPT: 100M users in 2 months (Instagram took 2.5 years)

Mar 2023 GPT-4 passes the bar exam BREAKTHROUGH

2024 Two Nobel Prizes go to AI work (Physics + Chemistry) DISCOVERY

The Race — Zero to Gold in 8 Years

2017 “Attention Is All You Need” paper ORIGIN

2020 GPT-3 launches — 175 billion parameters

Nov 2022 ChatGPT: 100M users in 2 months (Instagram took 2.5 years)

Mar 2023 GPT-4 passes the bar exam BREAKTHROUGH

2024 Two Nobel Prizes go to AI work (Physics + Chemistry) DISCOVERY

Jan 2025 DeepSeek R1 drops — Nvidia loses $600B in one day

The Race — Zero to Gold in 8 Years

2017 "Attention Is All You Need" paper ORIGIN

2020 GPT-3 launches — 175 billion parameters

Nov 2022 ChatGPT: 100M users in 2 months (Instagram took 2.5 years)

Mar 2023 GPT-4 passes the bar exam BREAKTHROUGH

2024 Two Nobel Prizes go to AI work (Physics + Chemistry) DISCOVERY

Jan 2025 DeepSeek R1 drops — Nvidia loses $600B in one day

Jul 2025 AI scores IMO gold medal (35/42) AI CONNECTION

What YOU Can Do Right Now

1. **Get the GitHub Student Developer Pack**
   Free Copilot, free cloud credits, free everything
2. **Take the Kaggle Intro to ML course**
   Free, hands-on, takes one weekend
3. **Open Google Colab and run a notebook**
   Free GPU, no setup, works in your browser
4. **Try a HuggingFace model**
   Thousands of pre-trained models, one line of code
5. **Enter a Kaggle competition**
   Real data, real problems, real community

What YOU Can Do Right Now

1. **Get the GitHub Student Developer Pack**
   Free Copilot, free cloud credits, free everything
2. **Take the Kaggle Intro to ML course**
   Free, hands-on, takes one weekend
3. **Open Google Colab and run a notebook**
   Free GPU, no setup, works in your browser
4. **Try a HuggingFace model**
   Thousands of pre-trained models, one line of code
5. **Enter a Kaggle competition**
   Real data, real problems, real community

$186K

ML ENGINEER MEDIAN SALARY

What YOU Can Do Right Now

1. **Get the GitHub Student Developer Pack**
   Free Copilot, free cloud credits, free everything
2. **Take the Kaggle Intro to ML course**
   Free, hands-on, takes one weekend
3. **Open Google Colab and run a notebook**
   Free GPU, no setup, works in your browser
4. **Try a HuggingFace model**
   Thousands of pre-trained models, one line of code
5. **Enter a Kaggle competition**
   Real data, real problems, real community

$186K

ML ENGINEER MEDIAN SALARY

Free tools: ChatGPT, Claude, GitHub Copilot (free for students), Google Colab, Kaggle

What YOU Can Do Right Now

1. **Get the GitHub Student Developer Pack**
   Free Copilot, free cloud credits, free everything
2. **Take the Kaggle Intro to ML course**
   Free, hands-on, takes one weekend
3. **Open Google Colab and run a notebook**
   Free GPU, no setup, works in your browser
4. **Try a HuggingFace model**
   Thousands of pre-trained models, one line of code
5. **Enter a Kaggle competition**
   Real data, real problems, real community

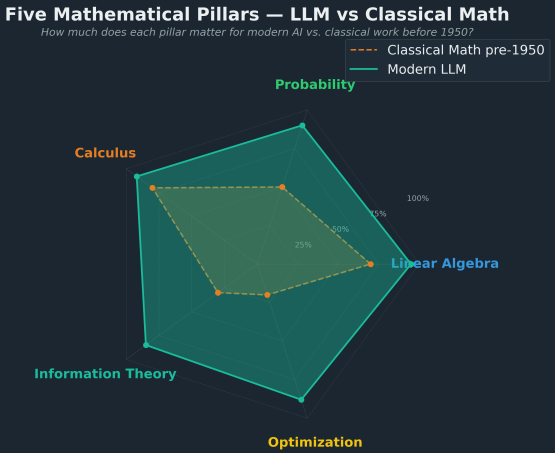
$186K

ML ENGINEER MEDIAN SALARY

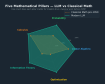
Free tools: ChatGPT, Claude, GitHub Copilot (free for students), Google Colab, Kaggle

The tools are free. The courses are free. What are you doing this weekend?

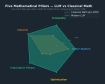
Five Pillars: Convergence Radar



The Code Is Still Being Written

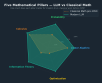


The Code Is Still Being Written



- AI / ML Engineer
- Data Scientist
- Research Mathematician
- Quantitative Analyst
- AI Safety Researcher

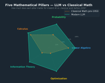
The Code Is Still Being Written



- AI / ML Engineer
- Data Scientist
- Research Mathematician
- Quantitative Analyst
- AI Safety Researcher

> *"The mathematicians who built these tools never imagined AI. The AI researchers who use them stand on 2000 years of shoulders."*

The Code Is Still Being Written



- AI / ML Engineer
- Data Scientist
- Research Mathematician
- Quantitative Analyst
- AI Safety Researcher

> *"The mathematicians who built these tools never imagined AI. The AI researchers who use them stand on 2000 years of shoulders."*

Thank you. Questions?