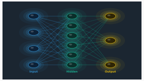


THE FIVE PILLARS

The Code of the Universe

From Classical Mathematics to Large Language Models

UAE Mathematics Conference 2026 · Prof. Jörg Osterrieder

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.

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 → Now

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

Local Impact

Dubai's autonomous metro runs on optimization. Abu Dhabi's Falcon LLM uses every pillar we will cover today.

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.

Local Impact

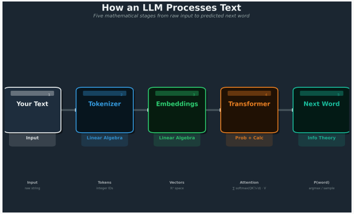
Dubai's autonomous metro runs on optimization. Abu Dhabi's Falcon LLM uses every pillar we will cover today.

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

What Happens When You Ask ChatGPT a Question?

What Happens When You Ask ChatGPT a Question?

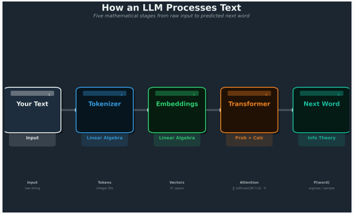
1. Your words become vectors, then giant matrices multiply — Linear Algebra



Token processing pipeline

What Happens When You Ask ChatGPT a Question?

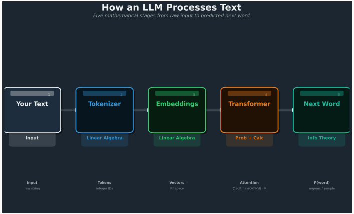
1. Your words become vectors, then giant matrices multiply — Linear Algebra
2. It computes probabilities for every possible next word — Probability



Token processing pipeline

What Happens When You Ask ChatGPT a Question?

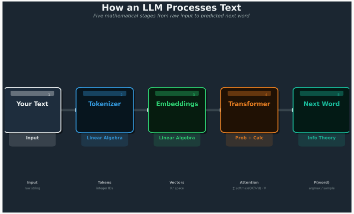
1. Your words become vectors, then giant matrices multiply — Linear Algebra
2. It computes probabilities for every possible next word — Probability
3. It learned from trillions of examples using derivatives — Calculus



Token processing pipeline

What Happens When You Ask ChatGPT a Question?

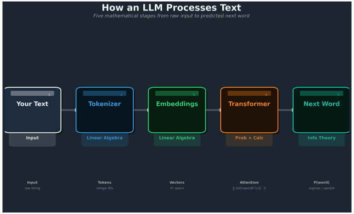
1. Your words become vectors, then giant matrices multiply — Linear Algebra
2. It computes probabilities for every possible next word — Probability
3. It learned from trillions of examples using derivatives — Calculus
4. The goal: minimize surprise (cross-entropy) — Information Theory



Token processing pipeline

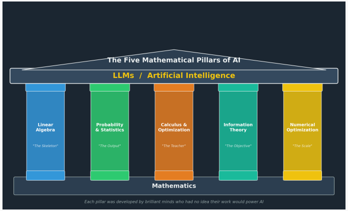
What Happens When You Ask ChatGPT a Question?

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 practical — Optimization

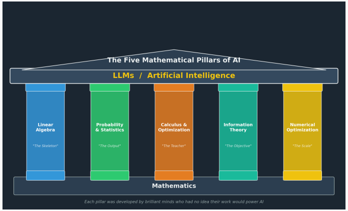


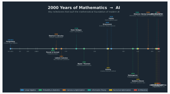
Token processing pipeline

The Five Pillars of AI Mathematics

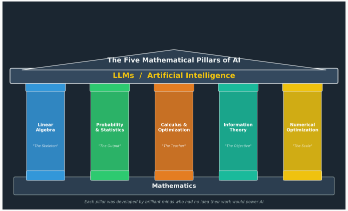


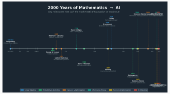
The Five Pillars of AI Mathematics





The Five Pillars of AI Mathematics





Each pillar was developed by brilliant minds who had no idea their work would power AI. We will visit each one, meet the mathematicians who built it, and show exactly where it appears inside a modern LLM.

Interactive Demos: How LLMs Work

Demo 1

3Blue1Brown — "Large Language Models explained briefly" (∼5 min). Visual walkthrough: tokenization, embeddings, attention, next-token prediction.

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

Demo 2

Transformer Explainer — type text, see tokens → embeddings → attention → prediction in a live GPT-2 model.

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

PILLAR 1

Linear Algebra

The Skeleton of AI

P1 2000 Years of Linear Algebra

Hermann Grassmann

P1

2000 Years of Linear Algebra

Hermann Grassmann

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

P1

2000 Years of Linear Algebra

Hermann Grassmann

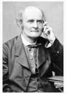
∼100 BCE Chinese *Fangcheng* — solving systems with counting rods ORIGIN

1844 Grassmann publishes vector spaces — almost universally ignored

P1

2000 Years of Linear Algebra

Hermann Grassmann



Arthur Cayley

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

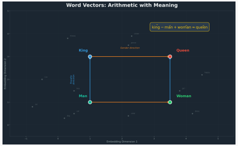
1844 Grassmann publishes vector spaces — almost universally ignored

1858 Cayley invents matrix theory — while working as a lawyer

P1

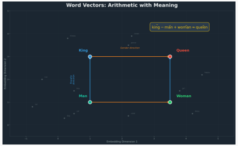
Words as Vectors

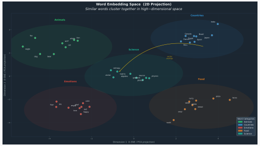
AI Connection



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

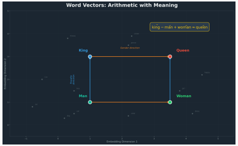
P1 Words as Vectors AI CONNECTION

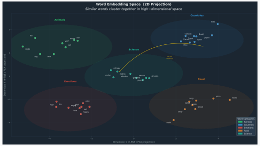




$$\vec{king} - \vec{man} + \vec{woman} \approx \vec{queen}$$

P1 Words as Vectors AI CONNECTION





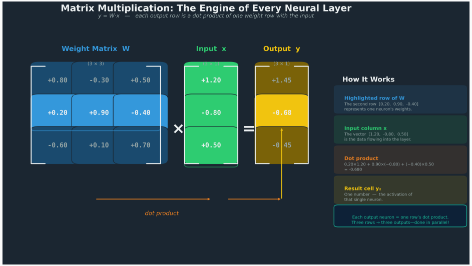
$$\vec{king} - \vec{man} + \vec{woman} \approx \vec{queen}$$

Mikolov et al., 2013 — Word2Vec: meaning encoded as geometry. GPT-3 uses 12,288-dimensional vectors; modern LLMs use comparable or larger spaces.

The Engine: Matrix Multiplication

P1

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

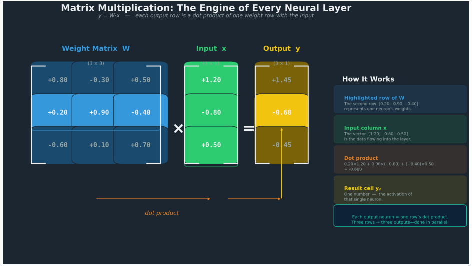


Matrix multiplication visualized

P1 The Engine: Matrix Multiplication

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

Every layer: multiply input vector by weight matrix, add bias. A neural network is this operation repeated hundreds of times.



Matrix multiplication visualized

The Engine: Matrix Multiplication

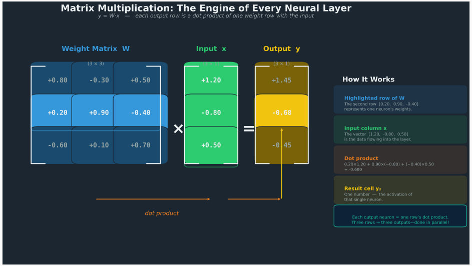
P1

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

Every layer: multiply input vector by weight matrix, add bias. A neural network is this operation repeated hundreds of times.

∼1.8T

PARAMETERS IN GPT-4 (ESTIMATED)



Matrix multiplication visualized

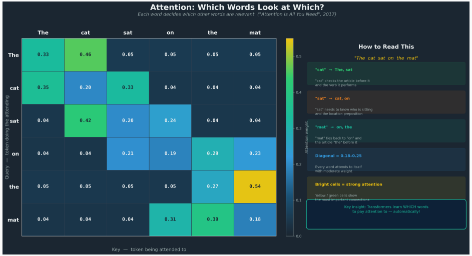
Attention = Three Matrix Multiplies

BREAKTHROUGH

P1

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

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

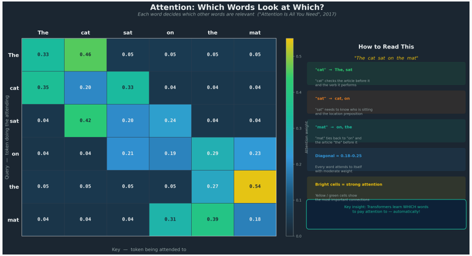


Attention weight heatmap

P1 Attention = Three Matrix Multiplies BREAKTHROUGH

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

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



Attention weight heatmap

Three matrices. That’s all attention is. Cayley’s 1858 invention, applied to language — and it changed the world.

KEY TAKEAWAYS

Linear Algebra

- Vectors represent meaning — words become geometry
- Matrices transform — every layer is a matrix multiply
- Attention is pure linear algebra — Q, K, V

PILLAR 2

Probability & Statistics

The Language of Uncertainty

P2

Born from Gambling

Blaise Pascal

P2

Born from Gambling

1654 Pascal & Fermat exchange letters about a gambling problem ORIGIN

Blaise Pascal

Fermat

Bayes

Kolmogorov

P2 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

1889 Galton describes the quincunx in *Natural Inheritance* — the bell curve made physical

Blaise Pascal

Fermat

Bayes

Kolmogorov

P2

Born from Gambling

1654 Pascal & Fermat exchange letters about a gambling problem ORIGIN

1763 Bayes' theorem published posthumously

1889 Galton describes the quincunx in *Natural Inheritance* — the bell curve made physical

1933 Kolmogorov writes the axioms — probability becomes rigorous

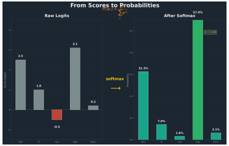
Blaise Pascal

Fermat

Bayes

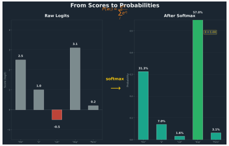
Kolmogorov

Turning Scores into Probabilities



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

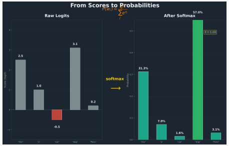
Turning Scores into Probabilities



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

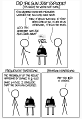
50,000+ words. One probability each. Kolmogorov's axioms in action: all positive, all sum to 1.

P2 Turning Scores into Probabilities



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

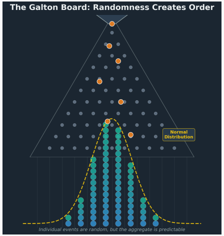
50,000+ words. One probability each. Kolmogorov's axioms in action: all positive, all sum to 1.



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

P2

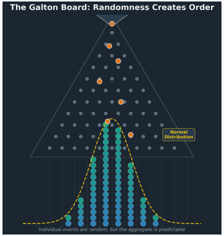
Randomness Creates Order



Galton board (1889)

Individual events are random. Each ball bounces left or right at every peg. Yet the aggregate always forms a bell curve — the Central Limit Theorem made physical.

P2 Randomness Creates Order



Galton board (1889)

Individual events are random. Each ball bounces left or right at every peg. Yet the aggregate always forms a bell curve — the Central Limit Theorem made physical.

LLMs show a similar pattern: each token is sampled from a probability distribution, but the sequence of samples produces coherent text. Randomness at the micro level, structure at the macro level.

KEY TAKEAWAYS

Probability & Statistics

- Probability is the output language of LLMs
- Softmax satisfies Kolmogorov's axioms — rigorously correct
- Randomness at micro level creates structure at macro level

PILLAR 3

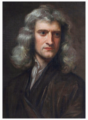
Calculus & Optimization

The Teacher

P3

The Calculus Wars

ORIGIN



Newton (1666)

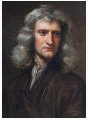
VS

Leibniz (1684)

P3

The Calculus Wars

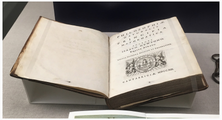
ORIGIN



Newton (1666)

VS

Leibniz (1684)

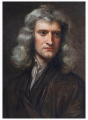


Principia, 1713 ed.

P3

The Calculus Wars

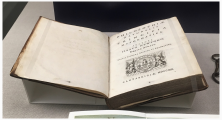
ORIGIN



Newton (1666)

VS

Leibniz (1684)

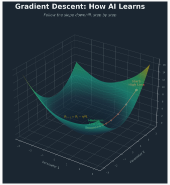


Principia, 1713 ed.

Both invented calculus independently. Newton accused Leibniz of plagiarism, then secretly wrote the Royal Society report exonerating himself. Modern historians agree: both were right. But we use Leibniz's notation: $\frac{dy}{dx}$

P3

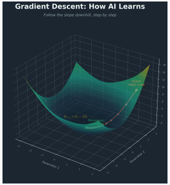
Gradient Descent: How AI Learns



Augustin-Louis Cauchy (1847)

P3

Gradient Descent: How AI Learns



Augustin-Louis Cauchy (1847)

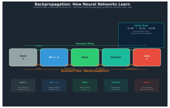
Blindfolded on a hill. Feel the slope under your feet. Step downhill. Repeat. That is gradient descent — and Cauchy invented it in 1847 for solving systems of equations.

P3

Backpropagation = The Chain Rule

BREAKTHROUGH

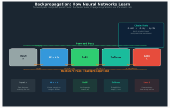
$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial w}$$



Geoffrey Hinton

P3 Backpropagation = The Chain Rule BREAKTHROUGH

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



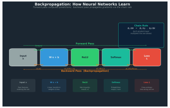
Geoffrey Hinton

NOBEL PHYSICS 2024

Backpropagation = The Chain Rule

BREAKTHROUGH

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



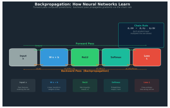
Geoffrey Hinton

NOBEL PHYSICS 2024

The chain rule: derivatives flow backward through every layer

P3 Backpropagation = The Chain Rule BREAKTHROUGH

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



Geoffrey Hinton

NOBEL PHYSICS 2024

The chain rule: derivatives flow backward through every layer

1986: Rumelhart, Hinton & Williams publish in *Nature*. 2024: Hinton wins the Nobel Prize for foundational work in machine learning.

KEY TAKEAWAYS

Calculus & Optimization

- Derivatives tell the model which way to adjust
- The chain rule makes it scale to billions of parameters
- 350-year-old math powers every AI training run

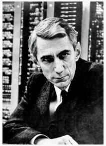
PILLAR 4

Information Theory

The Objective Function

P4

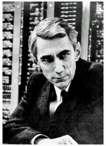
Shannon: Father of Information Theory



Claude Shannon

P4

Shannon: Father of Information Theory



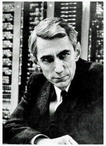
Claude Shannon

ORIGIN

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

P4

Shannon: Father of Information Theory



Claude Shannon

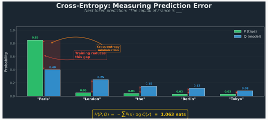
ORIGIN

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

1948: *"A Mathematical Theory of Communication"* — invented the bit as the fundamental unit of information.

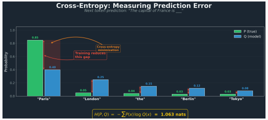
Fun fact: Shannon juggled while riding a unicycle through the halls of Bell Labs.

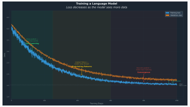
Cross-Entropy: The LLM Loss Function



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

Cross-Entropy: The LLM Loss Function



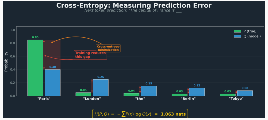


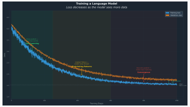
Training loss decreasing over time

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

Cross-Entropy: The LLM Loss Function

P4



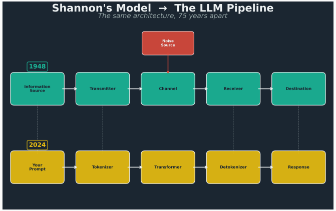


Training loss decreasing over time

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

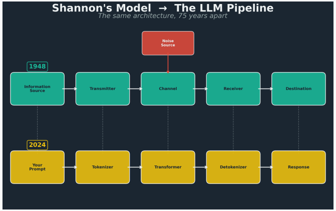
Shannon's 1948 formula **IS** the training objective of every LLM. Minimize surprise: assign high probability to the correct next word.

P4 Shannon's Model → The LLM Pipeline



P4

Shannon's Model $\longrightarrow$ The LLM Pipeline



Shannon (1948): Source $\longrightarrow$ Encoder $\longrightarrow$ Channel $\longrightarrow$ Decoder $\longrightarrow$ Destination

Modern LLM: User $\longrightarrow$ Tokenizer $\longrightarrow$ Transformer $\longrightarrow$ Output Layer $\longrightarrow$ Response

Designed for communication channels. Nearly 80 years later, describes what ChatGPT does with striking precision.

KEY TAKEAWAYS

Information Theory

- Cross-entropy = minimize surprise — the LLM training objective
- Shannon’s communication model maps remarkably onto the LLM pipeline
- A 1948 formula designed for communication channels trains every modern AI

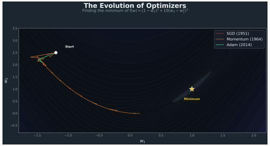
PILLAR 5

Numerical Optimization

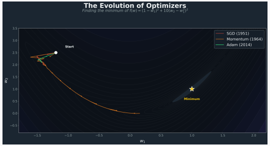
Training at Scale

P5

The Evolution of Optimizers



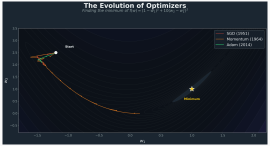
P5 The Evolution of Optimizers



1951 Robbins & Monro invent SGD ORIGIN

P5

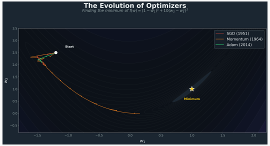
The Evolution of Optimizers



1951 Robbins & Monro invent SGD ORIGIN

1964 Polyak adds momentum — past gradients guide future steps

P5 The Evolution of Optimizers



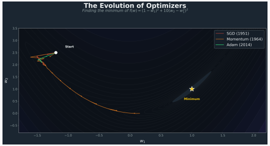
xkcd.com/1838 (CC BY-NC 2.5)

1951 Robbins & Monro invent SGD ORIGIN

1964 Polyak adds momentum — past gradients guide future steps

P5

The Evolution of Optimizers



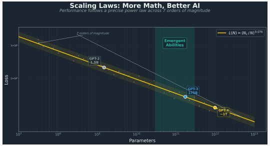
xkcd.com/1838 (CC BY-NC 2.5)

1951 Robbins & Monro invent SGD ORIGIN

1964 Polyak adds momentum — past gradients guide future steps

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

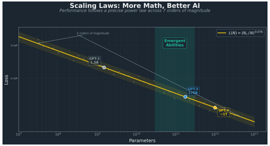
More Math, Better AI: Scaling Laws



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

P5

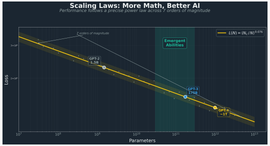
More Math, Better AI: Scaling Laws



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

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

More Math, Better AI: Scaling Laws



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

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

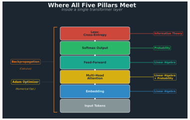
Kaplan et al., 2020: LLM performance follows a power law. Double the parameters → ~5% lower loss. Halving the loss requires ~10,000x more parameters. This equation is why companies spend billions on bigger models.

KEY TAKEAWAYS

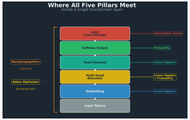
Numerical Optimization

- Adam is the workhorse of modern AI training
- Scaling laws reveal a power law — more parameters, predictably better
- The optimization frontier turns mathematical insight into economic force

Where All Five Pillars Meet



Where All Five Pillars Meet



Linear Algebra Embeddings & attention matrices

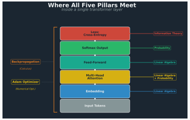
Calculus Backprop via chain rule

Probability Soft-max distributions

Info Theory Cross-entropy loss

Optimization Adam updates weights

Where All Five Pillars Meet



Linear Algebra Embeddings & attention matrices

Calculus Backprop via chain rule

Probability Soft-max distributions

Info Theory Cross-entropy loss

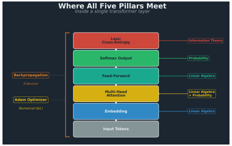
Optimization Adam updates weights

Five branches of pure mathematics, developed over **2000 years**, all running simultaneously in a single forward-backward pass. This is the code of the universe.

What AI Can Actually Do

35/42

IMO Gold Medal — Gemini Deep Think (2025)



Five pillars powering AI capabilities

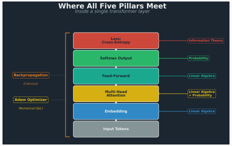
What AI Can Actually Do

35/42

IMO Gold Medal — Gemini Deep Think (2025)

Nobel

Chemistry 2024 — AlphaFold solved protein folding



Five pillars powering AI capabilities

What AI Can Actually Do

35/42

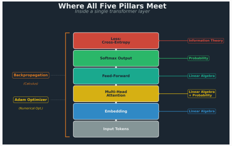
IMO Gold Medal — Gemini Deep Think (2025)

Nobel

Chemistry 2024 — AlphaFold solved protein folding

92%+

HumanEval — Claude coding benchmark (2025)



Five pillars powering AI capabilities

What AI Can Actually Do

35/42

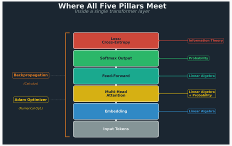
IMO Gold Medal — Gemini Deep Think (2025)

Nobel

Chemistry 2024 — AlphaFold solved protein folding

92%+

HumanEval — Claude coding benchmark (2025)



Five pillars powering AI capabilities

1M+ tokens

Context window — Gemini processes entire codebases at once

What AI Can Actually Do

35/42

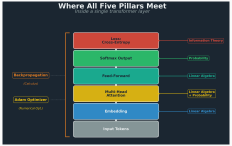
IMO Gold Medal — Gemini Deep Think (2025)

Nobel

Chemistry 2024 — AlphaFold solved protein folding

92%+

HumanEval — Claude coding benchmark (2025)



Five pillars powering AI capabilities

1M+ tokens

Context window — Gemini processes entire codebases at once

∼90%

MMLU — frontier models match expert human level (89.8%)

The Numbers Are Staggering

∼1.7T

ESTIMATED PARAMETERS IN GPT-4 (OPENAI HAS NOT CONFIRMED)

The Numbers Are Staggering

$\sim$1.7T

ESTIMATED PARAMETERS IN GPT-4 (OPENAI HAS NOT CONFIRMED)

$100M+

TRAINING COST — 25,000 GPUS FOR 90 DAYS

The Numbers Are Staggering

$\sim$**1.7T**

ESTIMATED PARAMETERS IN GPT-4 (OPENAI HAS NOT CONFIRMED)

$100M+

TRAINING COST — 25,000 GPUS FOR 90 DAYS

$\sim$**13T**

TRAINING TOKENS ≈ 2,200 WIKIPEDIAS

The Numbers Are Staggering

∼1.7T

ESTIMATED PARAMETERS IN GPT-4 (OPENAI HAS NOT CONFIRMED)

$100M+

TRAINING COST — 25,000 GPUS FOR 90 DAYS

∼13T

TRAINING TOKENS ≈ 2,200 WIKIPEDIAS

900M

WEEKLY CHATGPT USERS (FEB 2026)

The Numbers Are Staggering

$\sim$1.7T

ESTIMATED PARAMETERS IN GPT-4 (OPENAI HAS NOT CONFIRMED)

$100M+

TRAINING COST — 25,000 GPUS FOR 90 DAYS

$\sim$13T

TRAINING TOKENS $\approx$ 2,200 WIKIPEDIAS

900M

WEEKLY CHATGPT USERS (FEB 2026) That is more than 1 in 10 humans on Earth.

The Numbers Are Staggering

$\sim$**1.7T**

ESTIMATED PARAMETERS IN GPT-4 (OPENAI HAS NOT CONFIRMED)

$100M+

TRAINING COST — 25,000 GPUS FOR 90 DAYS

$\sim$**13T**

TRAINING TOKENS $\approx$ 2,200 WIKIPEDIAS

900M

WEEKLY CHATGPT USERS (FEB 2026) That is more than 1 in 10 humans on Earth.

Plot twist: DeepSeek R1 matched GPT-4 for $6 million. Open source. Nvidia lost nearly $600 billion in one day.

Brilliant and Broken

Strawberry: Ask an LLM to count the R's in "strawberry." It says 2. There are 3. The same system that earns an IMO gold medal cannot count letters.

Brilliant and Broken

Strawberry: Ask an LLM to count the R's in "strawberry." It says 2. There are 3. The same system that earns an IMO gold medal cannot count letters.

Numbers: Many LLMs claim $9.11 > 9.9$ — because they do pattern matching on text, not arithmetic on numbers.

Brilliant and Broken

Strawberry: Ask an LLM to count the R's in "strawberry." It says 2. There are 3. The same system that earns an IMO gold medal cannot count letters.

Numbers: Many LLMs claim $9.11 > 9.9$ — because they do pattern matching on text, not arithmetic on numbers.

Hallucinations: A New York lawyer was fined $5,000 for submitting fabricated court cases generated by ChatGPT. Air Canada had to honor a refund policy its chatbot invented.

Brilliant and Broken

Strawberry: Ask an LLM to count the R's in "strawberry." It says 2. There are 3. The same system that earns an IMO gold medal cannot count letters.

Numbers: Many LLMs claim $9.11 > 9.9$ — because they do pattern matching on text, not arithmetic on numbers.

Hallucinations: A New York lawyer was fined $5,000 for submitting fabricated court cases generated by ChatGPT. Air Canada had to honor a refund policy its chatbot invented.

Why? LLMs are statistical pattern completers, not fact databases. They predict the most likely next token. They have no internal fact-checker and no concept of truth.

The Race: Zero to Gold in 8 Years

The Race: Zero to Gold in 8 Years

2017 *"Attention Is All You Need"* — the transformer is born ORIGIN

The Race: Zero to Gold in 8 Years

2017 *"Attention Is All You Need"* — the transformer is born ORIGIN

2020 GPT-3 launches — 175 billion parameters

The Race: Zero to Gold in 8 Years

2017 *“Attention Is All You Need”* — the transformer is born 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"* — the transformer is born 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"* — the transformer is born 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"* — the transformer is born 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 nearly $600B in one day

The Race: Zero to Gold in 8 Years

2017 *"Attention Is All You Need"* — the transformer is born 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 nearly $600B in one day

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

The Math Behind the Headlines

MBZUAI — Mohamed bin Zayed University of Artificial Intelligence (Abu Dhabi). The world's first graduate-level AI university. Research spans all five pillars.

The Math Behind the Headlines

MBZUAI — Mohamed bin Zayed University of Artificial Intelligence (Abu Dhabi). The world's first graduate-level AI university. Research spans all five pillars.

Falcon LLM — Built by the Technology Innovation Institute, Abu Dhabi. Open-source, now in its third generation (Falcon 3). Every pillar you learned today is inside it: embeddings, softmax, cross-entropy, backprop, Adam.

The Math Behind the Headlines

MBZUAI — Mohamed bin Zayed University of Artificial Intelligence (Abu Dhabi). The world's first graduate-level AI university. Research spans all five pillars.

Falcon LLM — Built by the Technology Innovation Institute, Abu Dhabi. Open-source, now in its third generation (Falcon 3). Every pillar you learned today is inside it: embeddings, softmax, cross-entropy, backprop, Adam.

Dubai Smart City & Abu Dhabi AI Strategy — From autonomous transport to AI-powered government services. The UAE is building its future on mathematical foundations.

The five pillars you just learned are the foundation of all of this.

What YOU Can Do Right Now

What YOU Can Do Right Now

1. **GitHub Student Pack** — free Copilot, free cloud credits
2. **Kaggle Intro to ML** — free course, one weekend
3. **Google Colab** — free GPU in your browser
4. **HuggingFace** — thousands of models, one line of code
5. **Kaggle Competition** — real data, real problems, real community

What YOU Can Do Right Now

1. **GitHub Student Pack** — free Copilot, free cloud credits
2. **Kaggle Intro to ML** — free course, one weekend
3. **Google Colab** — free GPU in your browser
4. **HuggingFace** — thousands of models, one line of code
5. **Kaggle Competition** — real data, real problems, real community

$160K+

ML Engineer average salary (Glassdoor, 2026)

What YOU Can Do Right Now

1. **GitHub Student Pack** — free Copilot, free cloud credits
2. **Kaggle Intro to ML** — free course, one weekend
3. **Google Colab** — free GPU in your browser
4. **HuggingFace** — thousands of models, one line of code
5. **Kaggle Competition** — real data, real problems, real community

$160K+

ML Engineer average salary (Glassdoor, 2026)

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

The tools are free. The courses are free. The barrier has never been lower. What are you doing *this weekend*?

Mathematics Competition Pathways

IMO / EGMO / Math Competitions $\rightarrow$ AI Research

The proof techniques you practice — induction, estimation, combinatorial arguments — train the same mathematical thinking that builds AI. Linear algebra, probability, and optimization *are* competition mathematics.

Mathematics Competition Pathways

IMO / EGMO / Math Competitions $\rightarrow$ AI Research

The proof techniques you practice — induction, estimation, combinatorial arguments — train the same mathematical thinking that builds AI. Linear algebra, probability, and optimization *are* competition mathematics.

MBZUAI Scholarships — Full scholarships for MSc and PhD in AI, Machine Learning, NLP, and Computer Vision. Undergraduate research positions available.

UAE Math Olympiad Team $\rightarrow$ competition experience is valued by every top AI program worldwide.

Mathematics Competition Pathways

IMO / EGMO / Math Competitions $\rightarrow$ AI Research

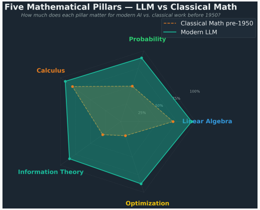
The proof techniques you practice — induction, estimation, combinatorial arguments — train the same mathematical thinking that builds AI. Linear algebra, probability, and optimization *are* competition mathematics.

MBZUAI Scholarships — Full scholarships for MSc and PhD in AI, Machine Learning, NLP, and Computer Vision. Undergraduate research positions available.

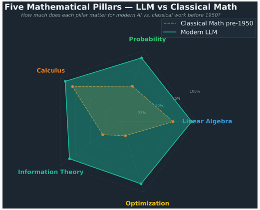
UAE Math Olympiad Team $\rightarrow$ competition experience is valued by every top AI program worldwide.

The math you study for competitions **IS** the math inside AI. The same inequalities that win medals are the same bounds that prove convergence of gradient descent. Your preparation already has a destination.

Five Pillars: The Complete Picture



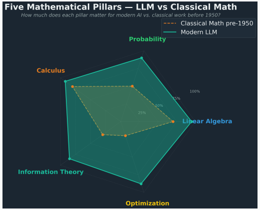
Five Pillars: The Complete Picture



AI/ML Engineer
Data Scientist
Research Mathematician

Quant Analyst
AI Safety Researcher

Five Pillars: The Complete Picture



AI/ML Engineer
Data Scientist
Research Mathematician

Quant 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

Active frontiers: Sparse attention (linear algebra), calibration (probability), second-order methods (calculus), mechanistic interpretability (information theory), distributed optimization at scale.

The Code Is Still Being Written

Active frontiers: Sparse attention (linear algebra), calibration (probability), second-order methods (calculus), mechanistic interpretability (information theory), distributed optimization at scale.

New pillars emerging: Topology, category theory, algebraic geometry, and differential geometry are all finding applications in AI. The mathematical story is far from finished.

The Code Is Still Being Written

Active frontiers: Sparse attention (linear algebra), calibration (probability), second-order methods (calculus), mechanistic interpretability (information theory), distributed optimization at scale.

New pillars emerging: Topology, category theory, algebraic geometry, and differential geometry are all finding applications in AI. The mathematical story is far from finished.

Open problems: Why do LLMs generalize so well? Can we prove convergence guarantees? How do we make AI safe and aligned? These are *mathematical* questions waiting for answers.

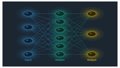
The Code Is Still Being Written

Active frontiers: Sparse attention (linear algebra), calibration (probability), second-order methods (calculus), mechanistic interpretability (information theory), distributed optimization at scale.

New pillars emerging: Topology, category theory, algebraic geometry, and differential geometry are all finding applications in AI. The mathematical story is far from finished.

Open problems: Why do LLMs generalize so well? Can we prove convergence guarantees? How do we make AI safe and aligned? These are *mathematical* questions waiting for answers.

The code of the universe is still being written. The next chapter may be written by someone in this room.



Thank You

Questions?

UAE Mathematics Conference 2026 · Prof. Jörg Osterrieder

The Five Pillars of AI Mathematics

Appendix: Formula Reference