AI for Mathematics

Week 1: From Counting to AI

Ma Jiajun

School of Math, XMU / Dept. of Math, XMUM

April 7, 2026

"Shock! Shock!"

– Donald Knuth, March 2026

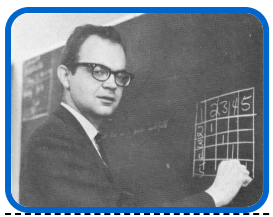
A 30-Year Problem, Solved in 1 Hour

Donald Knuth (b. 1938)

Turing Award (1974). Author of *The Art of Computer Programming* and creator of **TeX** – the typesetting system used by mathematicians worldwide.

- Worked on a graph theory conjecture for **30 years**
- In March 2026, **Claude AI** solved it in ~1 hour
- Claude recognized the structure as a **Cayley digraph** – the key insight
- Knuth verified, wrote the proof, named the paper **"Claude's Cycles"**

Knuth’s Reaction



2024 (age 86)

> *"It seems I’ll have to revise my opinions about generative AI one of these days."*
>
> *– Donald Knuth, "Claude's Cycles", March 2026*

60 years of work – from *The Art of Computer Programming* to being surprised by AI.

AI Wins Gold at the Math Olympiad

The **International Mathematical Olympiad (IMO)** – the most prestigious math competition in the world.

- **2024**: AI wins **Silver** medal (28/42, gold cutoff was 29!)
- **2025**: Three AI systems win **Gold** simultaneously
  - OpenAI – natural language proofs
  - Google Gemini – solved within 4.5 hours
  - **Harmonic Aristotle** – formally verified in Lean, **zero possibility of error**
- From Silver to Gold in just one year

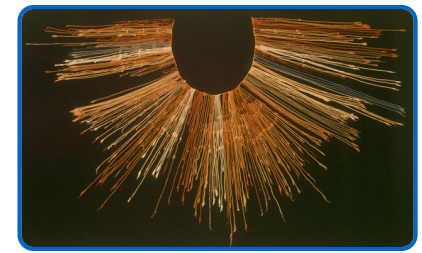
> **How did we get here?** To understand, let's go back to the very beginning of mathematics.

Part I

The Story of Mathematics and Proof

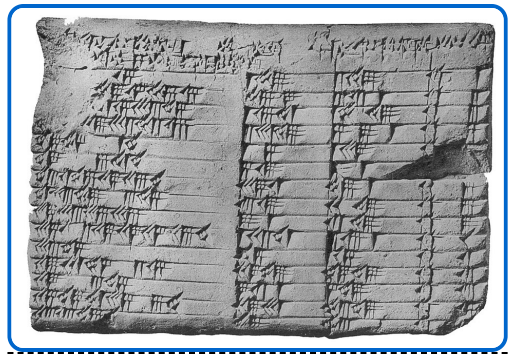
The Beginning of Mathematics: Counting

- **Tally marks** (~30,000 BC) – the oldest mathematical activity
- The **Ishango bone** (left) – found in Congo, marks carved into bone
- **Quipu** (right) – Inca knotted strings for recording numbers
- Humans needed to count: livestock, harvests, debts, time

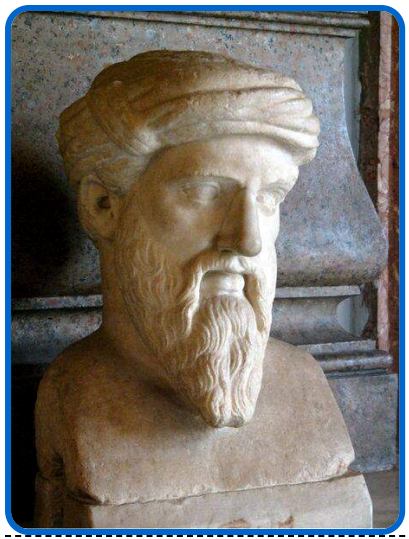


The Beginning of Mathematics: Geometry

- **Ancient Egypt** – measuring land after Nile floods ("geometry" = earth-measuring)
- **Babylon** – astronomical calculations, quadratic equations on clay tablets
- The **Plimpton 322** tablet (right, ~1800 BC) – lists Pythagorean triples!
- Mathematics was **empirical** – practical results, but no "why"
  - Egyptians knew 3-4-5 triangles were right angles
  - But never asked: *why does this work for all right triangles?*



The Birth of Proof



Pythagoras (~570–495 BC)

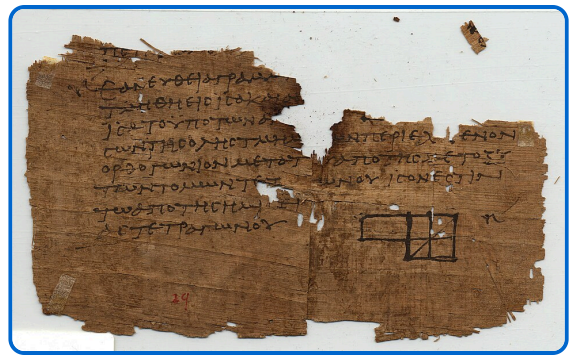
Greek mathematician and philosopher. Founded a school where mathematics was studied as a path to truth. The **Pythagorean theorem** $a^2 + b^2 = c^2$ is named after him.

- **Ancient Greece** – a revolution: started asking **"why"**
- Discovery of $\sqrt{2}$ being **irrational** – shattered Pythagorean beliefs
  - Legend: Hippasus revealed this secret and was drowned at sea by fellow Pythagoreans
  - A truth so dangerous it cost a life – but you can only find it through a **proof**

Euclid's *Elements*

Euclid of Alexandria (~300 BC)

His *Elements* is the most influential mathematics textbook in history, used for over **2000 years**.



Oldest surviving fragment of the *Elements*

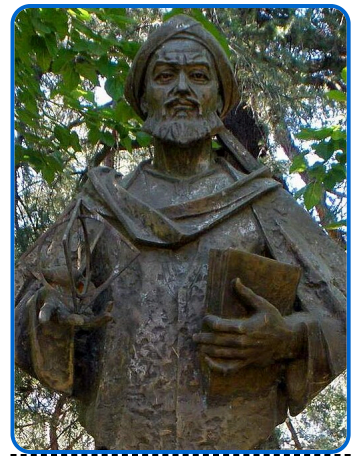
- 5 postulates → 465 propositions
- The first **axiomatic system** in human history
- Established the principle: **Mathematics must be proved.**

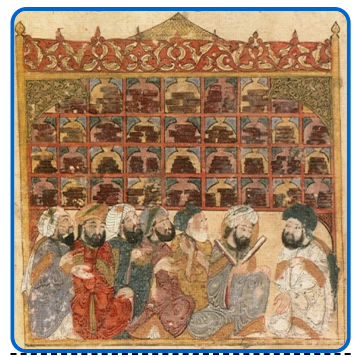
Preserving Mathematical Knowledge

Greek civilization declined – texts scattered and lost.

- **Islamic Golden Age** (8th–14th century)
  - Baghdad **House of Wisdom**: translated Euclid into Arabic
  - **Al-Khwarizmi** (~780–850): "algorithm" + "algebra"
- **China**: *Nine Chapters*, Zu Chongzhi's π
- **India**: invention of **zero**, decimal system

> **Analogy:** Like today's **Mathlib** – preserving all math knowledge for future generations and AI.





Proofs Become Too Complex

Mathematics exploded after the Renaissance…

- **Calculus** – Newton, Leibniz (1680s)
- **Analysis** – Euler, Gauss, Riemann…
- Proofs grew **longer** and **harder to verify**:

Proof	Length
Fermat's Last Theorem	**100+ pages**, 7 years
Finite simple groups	**10,000+ pages**, decades

How can we be sure these are correct?

Newton

Gauss

Andrew Wiles and Fermat's Last Theorem

Andrew Wiles (b. 1953) – proved **Fermat's Last Theorem** in 1995, a conjecture open for **358 years**. Over 100 pages – by proving the modularity theorem for semistable elliptic curves, connecting many deep areas of mathematics.

FLT: No positive integers a, b, c satisfy $a^n + b^n = c^n$ for $n > 2$.

Fermat claimed a proof in 1637. It took **358 years**. Can a computer **verify** such a proof?

*Kevin Buzzard's **FLT Project** is formalizing Wiles's proof in Lean. When proposed, many thought it was hopeless – but with AI, it now looks very achievable.*

Can Machines Help Us

Verify Proofs?

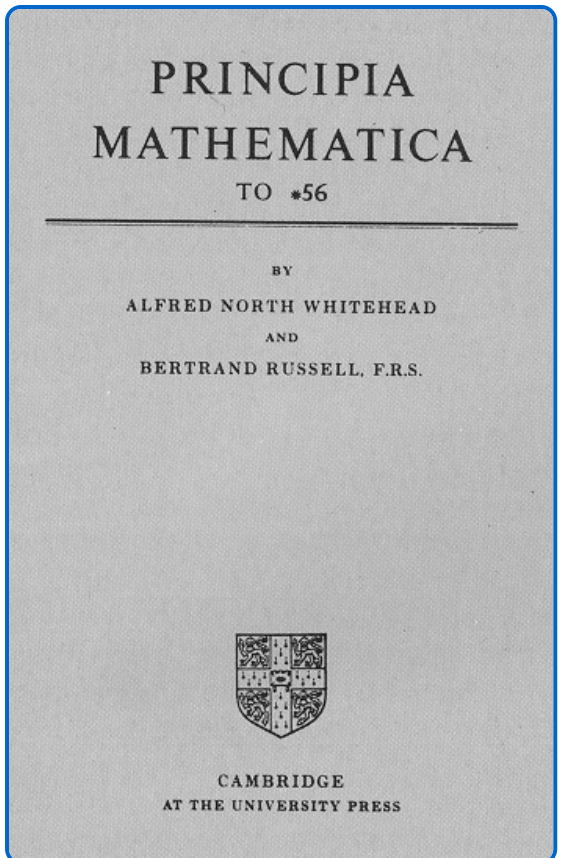
Hilbert's Dream

David Hilbert (1862–1943)

One of the most influential mathematicians in history. Proposed 23 famous problems in 1900 that shaped 20th-century mathematics.

- **Hilbert's Program (1920s)**: Can we create a mechanical procedure to verify all of mathematics?
- The dream of **complete formalization** of mathematics

Principia Mathematica


PRINCIPIA
MATHEMATICA
TO *56

BY
ALFRED NORTH WHITEHEAD
AND
BERTRAND RUSSELL, F.R.S.

CAMBRIDGE
AT THE UNIVERSITY PRESS

Russell & Whitehead (1910–1913)

- 362 pages to prove that $1 + 1 = 2$
- Every mathematical statement broken down into pure logic
- Heroic but impractical – far too tedious for humans

> Could a **machine** do this tedious work instead?

Gödel's Answer: No.

Kurt Gödel (1906–1978)

Austrian-American logician. His Incompleteness Theorems are among the most important results in mathematical logic.

- **Incompleteness Theorems (1931):**
  - Any sufficiently strong formal system has statements that are **true but unprovable**
- The dream of *complete* mechanization was impossible
- But… *partial* mechanization? That's a different story.

Weyl's Prophecy

Hermann Weyl (1885–1955)

German-American mathematician. Made fundamental contributions to physics, topology, and group theory.

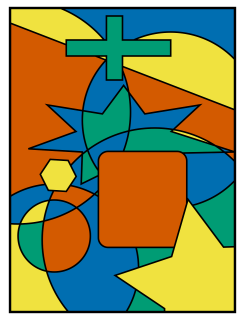
> *"The deeper one digs the spade, the harder the digging gets; maybe it has become too hard for us unless we are given some outside help, be it even by such devilish devices as* ***high-speed computing machines****."*
>
> *— Hermann Weyl, Princeton Bicentennial Conference, 1946*

Computers Enter Mathematics

1956 – Present

The First Steps

- **Logic Theorist (1956)** – Newell, Shaw, Simon
  - First automated theorem prover
  - Proved **38 of 52** theorems from *Principia Mathematica*
- **Automath (late 1960s)** – de Bruijn
  - First **proof checker** – the **Curry-Howard correspondence**: proofs = programs
- **Four Color Theorem (1976)** – Appel & Haken
  - First major **computer-assisted proof**
  - Controversy: *Is a proof you can't check by hand still a proof?*
  - 2005: Gonthier **formalized** it in Coq (6 years)
- **Proof assistants emerge:** Mizar (1973), Coq (1984), Isabelle (1986)
- **QED Manifesto (1994):** Formalize ALL mathematics – too ambitious, dissolved by 1996



The Lean Revolution

Lean: A New Era

Lean – created by Leonardo de Moura in 2013 (Microsoft Research)

- **Lean 4** (2021): reimplemented in Lean itself – fast, modern, programmable
- **Mathlib**: community-built library of formalized mathematics
  - Dec 2025: 250,000+ theorems, 120,000+ definitions
  - The **largest** collection of formalized mathematics ever
- **Why Lean won:**
  - Modern programming language (not just a proof checker)
  - Powerful automation (tactics, metaprogramming)
  - Active open-source community
  - Used by DeepMind, OpenAI, Microsoft, and startups

> Remember the **QED Manifesto** (1994)? Mathlib is making that dream real – 30 years

Formalization is Getting Dramatically Faster

Kepler / Flyspeck (2014) · 11 years

Liquid Tensor (2022) · 6 months

PFR (2023) · 3 weeks

Strong PNT (2025) · 3 weeks

Fields Medal (2025) · 5 days

From 11 years to 5 days. What changed? **AI.**

The People Behind the Formalization Revolution

Peter Scholze

Fields Medal 2018

Liquid Tensor challenge

Timothy Gowers

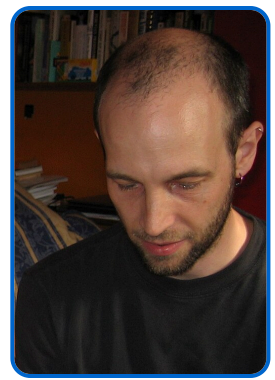
Fields Medal 1998

PFR Conjecture

Terence Tao

Fields Medal 2006

PFR + PNT challenge



Kevin Buzzard

Imperial College

FLT Project + Lean advocacy

> *"Mathematicians learning Lean by doing." – Kevin Buzzard*

AI Joins the Race

2024 – 2026

Math Benchmarks Are Saturating

AI is solving math problems at an accelerating pace.

Benchmark	2021	2024	2026	Status
GSM8K (grade school)	35%	99%	–	**Saturated**
MATH (high school)	~5%	~90%	–	Near saturated
FrontierMath (research)	–	2%	40%	Rapidly rising

AI went from struggling with **grade school math** to tackling **research-level problems** in just a few years. Benchmarks that were "impossible" in 2021 are now **saturated**.

Putnam Competition

The **hardest** undergraduate math exam in North America.

- Human top score (2024): **90/120** · Median: **2/120** (!) · Only **5 humans** ever scored 12/12 in **98 years**

AI System	Score	Notes
DeepSeekMath-V2	**118/120**	Beats human top score
Axiom	**12/12**	Perfect score
Numina-Lean-Agent	**12/12**	Open-source, Lean verified

> Median human: 2/120. AI: 118/120. That's not incremental improvement – it's a **paradigm shift**.

IMO: The Grand Challenge

The **International Mathematical Olympiad** – the ultimate test of mathematical ability.

2020 · IMO Grand Challenge proposed – can AI win gold?

2024 Jan · AlphaGeometry – 25/30 historical problems (gold-level)

2024 Jul · AlphaProof → Silver (28/42, gold cutoff was 29!)

2025 Jul · THREE teams reach Gold simultaneously!

IMO 2025: Gold!

System	Score	Medal	Method
OpenAI	35/42	**Gold**	Informal (natural language)
Gemini Deep Think	35/42	**Gold**	Informal (within 4.5h)
Harmonic Aristotle	5/6	**Gold**	**Formal (Lean-verified!)**
Seed-Prover 1.5	35/42	**Gold**	Formal (16.5h)

- **Aristotle**: first system to produce **formally verified** gold-medal solutions
- All systems struggled on **P6** – requires creative insight
- From **Silver to Gold** in **just one year**

miniF2F: Proof Generation Arms Race

How well can AI **generate formal proofs** in Lean?

System	Pass Rate	Date
DeepSeek-Prover-V1.5	~50%	2024
Goedel-Prover V1	57.6%	Early 2025
DeepSeek-Prover-V2 (671B)	88.9%	Apr 2025
Goedel-Prover-V2 (32B)	**90.4%**	Aug 2025

From 50% to 90% in one year. The 32B model outperforms the 671B model – efficiency is improving too.

AI Discovers New Mathematics

2025 – 2026

AI Is No Longer Just Verifying — It Is Discovering

- **Erdős Problems (Jan 2026)**
  - GPT-5.2 solved Problems #397, #728, #729
  - **Terence Tao** verified personally
- **Claude's Cycles (Mar 2026)** ← our opening story!
  - Knuth's 30-year conjecture → 1 hour
- **DeepMind – Algebraic Geometry**
  - Gemini proved a **new theorem**
  - **Ravi Vakil** (AMS President): "rigorous, correct, and elegant"

Tao verified Erdős proofs

Ken Ono's Story

Ken Ono (b. 1968)

One of the world's top number theorists. Former Thomas Jefferson Professor at University of Virginia.

- In 2025: left academia to become "Founding Mathematician" at **Axiom**, an AI startup
- Axiom's AI solved a 20-year conjecture **he himself couldn't solve**
- Axiom scored **12/12** on Putnam – perfect
- Axiom raised $200M in March 2026

A sign of the times: top mathematicians are moving into AI.

First Proof Challenge (Feb 2026)

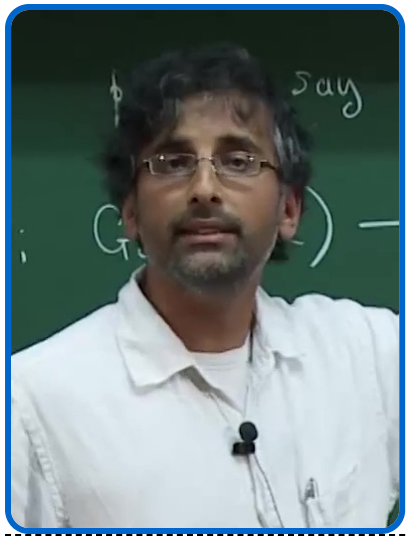
Can AI do **research-level** math on problems it has **never seen**?

- 11 mathematicians proposed 10 **unpublished** research problems
- Solutions had **never appeared online** – zero data contamination
- Results:
  - DeepMind's Aletheia: **6/10**
  - OpenAI: **5/10**
  - Organizers' assessment: **2/10** clearly correct
- AI is approaching the research frontier – **a rapidly closing gap**

> The gap between AI and human mathematicians is **closing fast**. What took decades of training now takes hours of compute.

What Does This Mean?

Mathematicians' Perspectives



Akshay Venkatesh (b. 1981) – Fields Medal 2018, IAS Princeton

In his essay *"Some Thoughts on Automation and Mathematical Research"*:

- AI will alter our understanding of what mathematics **is**
- "The easier it gets to solve certain problems, the less we will **value** them"
- Mathematicians will become: **critic, translator, conductor, experimentalist**
- Impact comparable to "the introduction of the computer"

What Does This Mean for **YOU**?

- The **Industrial Revolution** gave everyone physical strength
- The **AI Revolution** gives everyone **intellectual strength**
- Learning to use AI for math = acquiring a superpower
- AI + math skills are in **high demand**: finance, tech, research, data science
  - Companies like Axiom ($200M), Harmonic, Math Inc. are hiring
- You **don’t need to be a math genius** – AI levels the playing field
- But you **DO need to understand** what AI is doing

That's what this course is about.

AI in Math Education

- LLMs as **personal math tutors** – infinite patience, 24/7
- **Lean games** for interactive learning:
  - **Natural Number Game** – proof tactics through play
  - **Abstract Algebra Game** – built at XMUM!
- XMUM students doing **FYP projects** in Lean:
  - Game theory, Coxeter groups, Hecke algebra
  - Joint projects with NUS, PKU, Harvard
- **This course itself** is an example

Your AI toolkit:

- **LLMs** – solve & explain
- **Lean** – verify proofs
- **LLM + Lean** – generate verified proofs
- **AI agents** – discover new math

Course Overview

Weeks	Topic
1–3	LLM for math + Type Theory + Lean
4–8	Reading Lean + LLM proof generation
9–12	Discovery, conjecture + projects
13–14	**Student presentations**

Assessment:

- Assignments: **40%**
- Presentation: **20%**
- Final Project: **40%**

Part II

Let's Try It! – Using LLMs to Solve Math

Live Demo

Let's see what AI can do with math **right now**.

1. A **calculus** problem
2. A **linear algebra** problem
3. A **competition** problem

Using: **Claude** / **ChatGPT** / **DeepSeek**

Key questions:
When does AI get it right?
When does it fail?
How do we **verify**?

Example prompt:

```
"Prove that √2 is irrational. Show every step."
```

Your Turn

If you have a laptop:

1. Go to claude.ai or chatgpt.com
2. Try asking it a math problem **you know the answer to**
3. Does it get it right?
4. Can you find a problem where it **makes a mistake**?

Try these:

- "What is the integral of $x^2 \sin(x)$?"
- "Is 91 a prime number? Explain."
- "Prove that there are infinitely many primes."
- Ask a problem from **your own course**!

Part III

A First Look at Formal Proofs

Why Formal Verification?

LLMs can **hallucinate** – produce confident but **wrong** proofs.

- How do we **know** an AI's proof is correct?
- Answer: **formal verification** – a computer checks every step
- The key idea: **Curry-Howard Correspondence**

Propositions = Types

Proofs = Programs

A proof is correct if and only if the program **type-checks**. The computer says ✅ or ❌ – no ambiguity.

Natural Number Game

An interactive way to learn Lean – through a **game**!

- Created by **Kevin Buzzard** and collaborators
- Learn proof tactics by proving properties of natural numbers
- No installation needed – runs in the **browser**
- You'll learn tactics like `rfl`, `rw`, `induction` – the building blocks of formal proof

https://adam.math.hhu.de/#/g/leanprover-community/NNG4

Let me show you a few levels...

Homework

1. **Play the Natural Number Game**
   - Complete as many levels as you can
   - https://adam.math.hhu.de/#/g/leanprover-community/NNG4
2. **Try using an LLM for a math problem**
   - Pick a problem from your other courses
   - Ask Claude / ChatGPT to solve it
   - Evaluate: is the solution correct? Why or why not?

See You Next Week!

Week 2: Dependent Type Theory + LLM Mathematical Reasoning