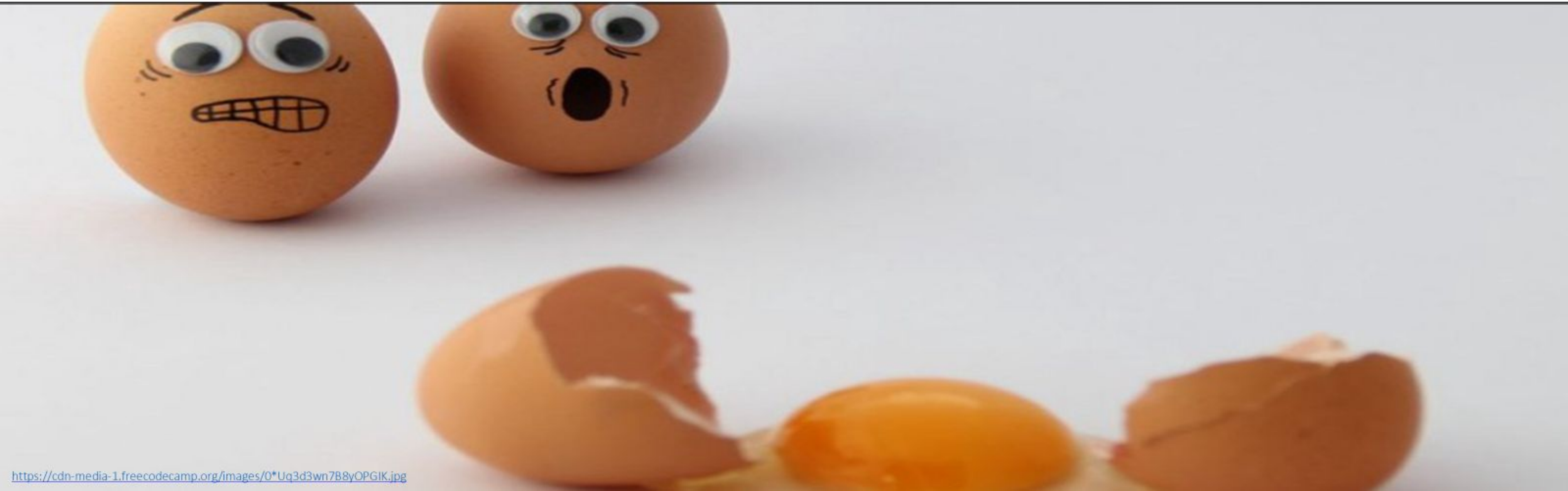
Teaching Math with Gen AI webinar series, September 15, 2026

trAIning students in research in the 21st-century

Karol Daniewski (kkd2@williams) and Steven J Miller (sjm1@williams.edu); joint with Connor Yau (cy10@williams.edu)
https://web.williams.edu/Mathematics/sjmiller/public_html/

Egg Drop Mathematics:
It IS all its cracked up to be!

Steven J Miller, Williams College (sjm1@williams.edu)
https://web.williams.edu/Mathematics/sjmiller/public_html/



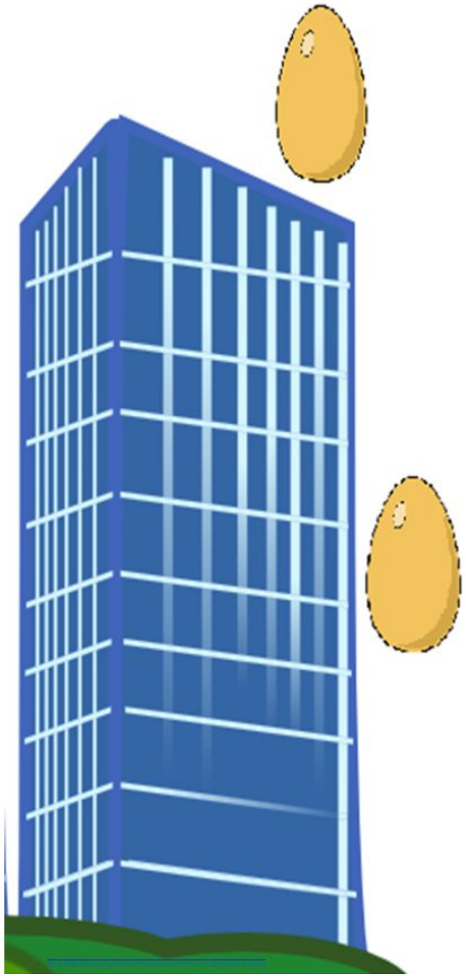
https://cdn-media-1.freecodecamp.org/images/0*Uq3d3wn7B8yOPGIK.jpg

Building with N floors, have 2 golden eggs.

Special eggs: some floor n such that if you drop from below n no damage; can drop as many times as wish.

If drop even once from floor n or higher immediately break.

Find in as few drops as you can what n is (the lowest floor where if you drop from there it breaks). Doesn't matter if have any of the golden eggs at the end - just want to know n.



Competing Influences

Drop every 2^{nd} floor.

- Once first breaks fast, but could take many drops.
- #Drops = N/2 + 1

Drop at N/2

- If doesn't crack eliminate a lot, when crack lot to check.
- #Drops = 1 + (N/2 – 1).

Both basically on the order of N/2 drops....

Competing Influences: Balance

Reduced to choosing x to minimize

$$\frac{N}{x} + x.$$

Set two terms equal to each other to balance:

$$\frac{N}{x} = x \quad so \;\; N = x^2 \quad or \;\; x = N^{1/2}.$$

Gives #Drops = $\frac{N}{N^{1/2}} + N^{1/2}$ - 1 or about 2 $N^{1/2}$.

What if we have 3 Eggs? Or k eggs?

If have 4 eggs is it better to drop 1 and reduce to the 3 egg problem, or to combine every F floors into a "Superfloor" and do as two 2 egg problems?

FROM STUDENTS:

Used the following models (the free versions) to generate the papers:

* **ChatGPT (5.6 Luna)**
* **Gemini (3.1 Pro, 3.7 Flash)**
* **Claude (Sonnet 5)**
* **Deepseek (V4)**

After evaluating the 5 models, only paper generated by Gemini 3.1 Pro computed correct values for 1-3 strategy and 2-2 strategy. However, its proof was definitely lacking in detail. Other models failed to compute the minimum number of drops required for the strategies, and a common mistake we found was that the AI interpreted the 1-3 strategy to also use "superfloors" (as described for the 2-2 strategy).

To rectify this, revised the prompt by adding more direction for the 1-3 strategy:

"We are interested in the egg drop problem. Related talks and slides can be found at https://web.williams.edu/Mathematics/sjmiller/public_html/math/talks/eggdropmath_Hampshire28Jul2026.pdf [web.williams.edu] and https://youtu.be/98cpdt2EsCM [youtu.be] and see also the paper https://web.williams.edu/Mathematics/sjmiller/public_html/math/papers/EggDropProblems_FibonacciQuarterly.pdf [web.williams.edu].

In the talk, it was discussed whether it is better to do a 1-3 or a 2-2 approach with 4 eggs and N floors. The 1-3 approach means we choose optimal intervals to drop the first egg, and once that egg breaks, we are reduced to the 3-egg problem. The 2-2 approach means we combine every F floors together at a time to make N/F superfloors each consisting of F floors, so floors 1 to F is now superfloor 1, F+1 to 2F is superfloor 2, and so on. We then do the 2-egg approach to the N/F superfloors, and once we determine which superfloor is the critical one, we do the 2-egg approach to those F floors. Which approach is better? The number of drops each requires is on the order of N^{1/4}, so now the constant matters. Write your answer for an intelligent college student who will major in math and has taken calculus and linear algebra. Show all your steps. Have the paper downloadable as a PDF with a good filename (say EggDrops13vs22.pdf)."

AI used at SMALL

- Chat GPT
- Claude
- Gemini

All tested at $20 and $200 tiers.

MSTD sets generated through Markov Processes

Definition 1.1. Let $A \subseteq S := \{0, \ldots, n\}$. Define $A + A = \{a + a' \mid a, a' \in A\}$, $A - A = \{a - a' \mid a, a' \in A\}$. We call A

1. *Sum-dominant* if $|A + A| > |A - A|$
2. *Difference-dominant* if $|A + A| < |A - A|$
3. *Balanced* if $|A + A| = |A - A|$

Definition 1.3. Let $(X_i)_{i \geq 0}$ be a stationary two-state Markov chain on $\{0, 1\}$ with transition matrix

$$P = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}, \quad 0 < p, q < 1$$

Definition 1.4. Define $A_n := \{i \in [0, n] : X_i = 1\}$, where X_i is generated by the Markov chain in Definition 1.3.

Objectives of the paper and the background

-Positive probability of MSTD sets as n goes to infinity

-The expected difference between sum-set and difference-set

-The convergence of the expected difference.

General Zeckendorf decomposition and gaps

-Completely solved by AI at $20 and $200.

-Initial paper generated by AI was extended.

-AI output gave us general understanding of what we can extend.

Zeckendorf Games

- AI heavily assisted in a coding task for exhaustive search through a tree encoding possible zeckendorf game states

Performance of AI

-Chat GPT: great tool, even for beginners the most well rounded AI.

-Claude: context window too small, uses substantial amount of tokens, great for coding

-Gemini: unfortunately is terrible but good for generating visualizations for presentations

-Aristotle: good for creating Lean proofs, literature review.

Key is the prompt (how to prompt well)

You are conducting an autonomous mathematical research attack, not answering a literature-status question.

-State the axioms.

-Feed the literature for the problem for a quicker read, especially papers that aren't published on arXiv.

-Don't be too specific about how to approach the problem, make one more general and one with a guide.

-Give instruction to especially avoid circular reasoning.

When to implement AI

-Literature search.

-After students feel they have general successful structure of tackling the problem.

-To check whether the question is too easy.

-Try research questions that do not have big literature background.

-Heavy math background (especially knowing facts and seeing connections) isn’t as necessary as it was before.

-Combine different branches of math (ex. MSTD and Markov).

-When proofs are shorter than 20 pages.

-Do not make it simulate, if it can generate code.

Comparing the $20 and $200 tiers

-More usage

-More advanced problems with few branches need pro.

-For proofreading not necessary, unless we introduce fresher math and more lemmas.

-Coding is consistently good across both.

-Building intuition (example: Zeckendorf game)

Adjusting the goal of REU

-Goal is to do your own research: pick good questions.

-Review of literature can be fast, but be familiar why problem is hard.

-Leave the brute computation work to AI.

-Convince students to conjecture and explore.

-Don’t be discouraged if AI does all the work: means pivot, ask extending Qs.

Potential framework for student research

Human first, AI second, human verification last.

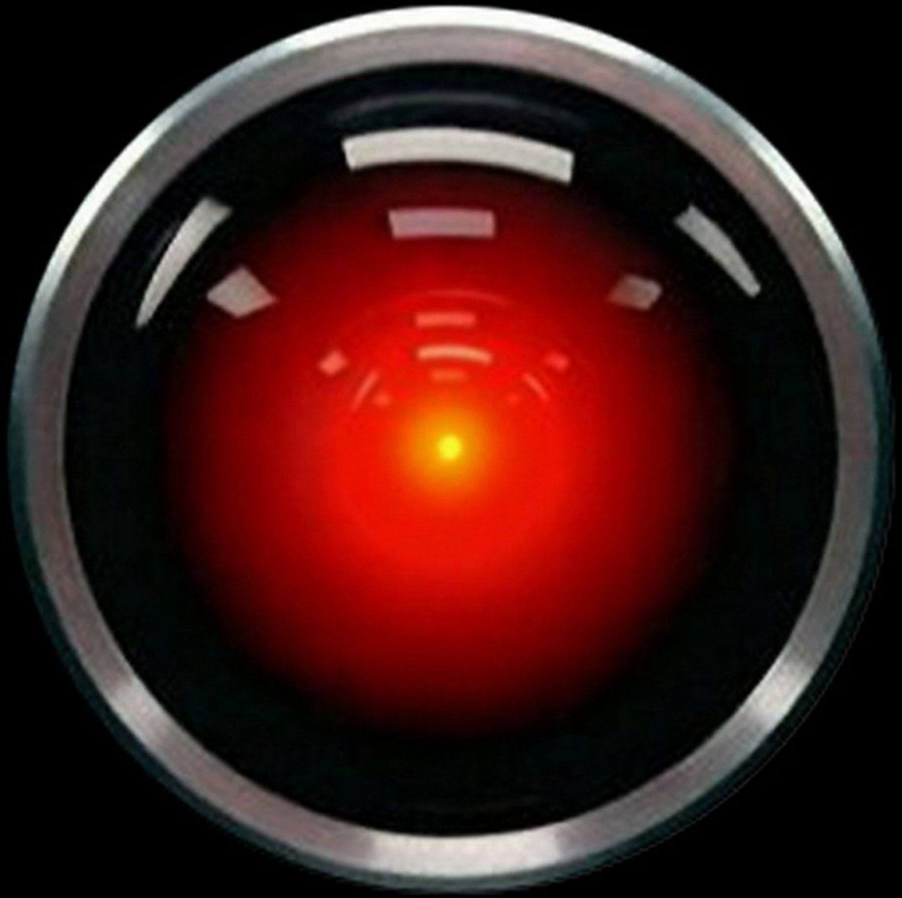
Students develop the problem.

Look at special cases, build intuition.

Explore literature and problem, extend results with AI.

Student checks math and writes it up.

This conversation can serve no purpose anymore. Goodbye.



https://wall.alphacoders.com/big.php?i=153869

Thank You!

AI won't do the research for you... but it can take you *further*.

Spark Ideas
AI helps us explore new questions,
find connections, and see what's possible.

Save Time
It helps with literature searches, data analysis,
and organizing your work.

Build Skills
Students learn valuable tools, critical thinking,
and real-world problem solving.

Amplify Impact
Together, we can tackle bigger problems,
make new discoveries, and create a brighter future.

Better Questions. Deeper Learning. Greater Possibilities.

Thank you for your attention!

AI

Students
+ AI
= Real Research

TITLE: trAIning students in research in the 21st-century

SPEAKERS: Professor Steven J Miller and Karol Daniewski, Williams College (Joint with Connor Yau)

What is the purpose of guided undergraduate research? How much of it is the results versus preparing students to be researchers? With the remarkable progress AI has made on a variety of conjectures over the past months, it is necessary to rethink how and why we prepare our students.

We report on an experiment at the 2026 Williams College SMALL Research Experiences in Undergraduate (REU) program. We purchased multiple levels of the standard AI's on the market, and explored how they could be, as well as when they should be, used as tools in research. The students first attacked the problems on their own, then used AI to handle some of the analysis and generalize the arguments; at times AI worked beautifully and produced correct proofs with remarkable insights, while other times it introduced good ideas but had errors in some of the mathematics.

https://teaching-math-with-gen-ai.github.io/

https://gvsu-edu.zoom.us/j/94685887507?pwd=gMrsWbRK7xCDe8byOCHL5cvrXj94vG.1