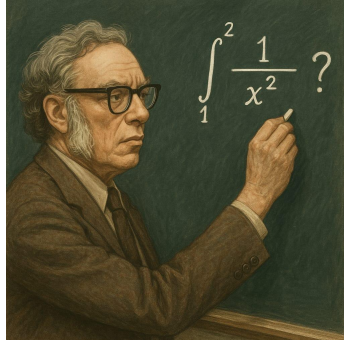




Technion
Israel Institute of Technology



**Ramanujan
Machine**

ASyMOB: Algebraic Symbolic Mathematical Operations Benchmark

Do LLMs Understand Calculus?



Michael Shalyt

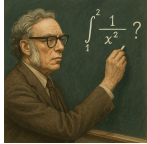


Rotem Elimelech



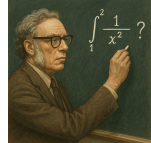
Ido Kaminer

Even a Tiny Tweak Can Confuse LLMs



$$\log_2 x - 2 \log_2 y = 2 \quad y(x) = \frac{1}{2} \sqrt{x} \quad \checkmark$$

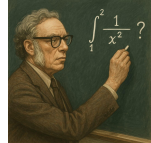
Even a Tiny Tweak Can Confuse LLMs



$$\log_2 x - 2 \log_2 y = 2 \quad y(x) = \frac{1}{2} \sqrt{x} \quad \checkmark$$

$$\log_2 \textcircled{A} x - 2 \log_2 \textcircled{A} y = 2 \quad y(x) = ? \quad \times$$

Even a Tiny Tweak Can Confuse LLMs

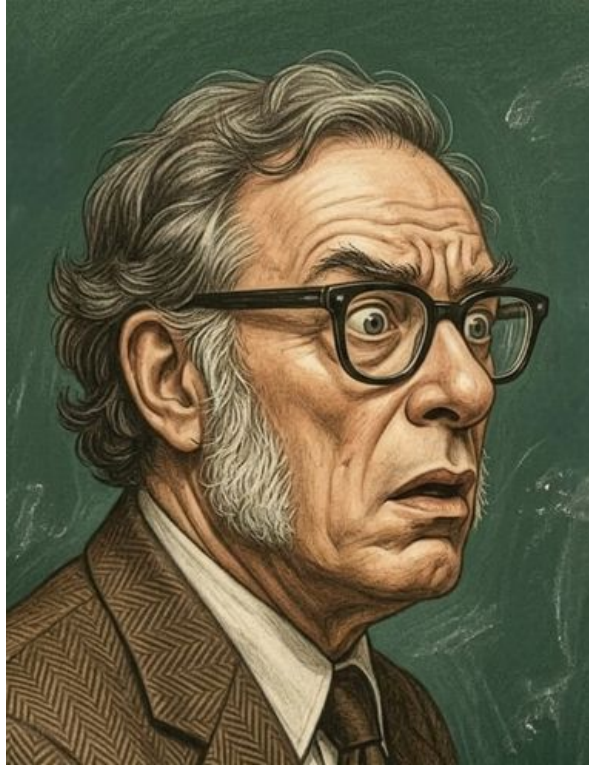


$$\log_2 x - 2 \log_2 y = 2 \quad y(x) = \frac{1}{2} \sqrt{x} \quad \checkmark$$

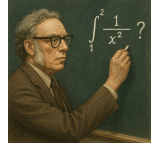
$$\log_2 \textcircled{A} x - 2 \log_2 \textcircled{A} y = 2 \quad y(x) = ? \quad \times$$

$$\log_2 \textcircled{1} x - 2 \log_2 \textcircled{1} y = 2 \quad y(x) = ? \quad \times$$

Even a Tiny Tweak Can Confuse LLMs



100 Seeds → 35,368 Challenges



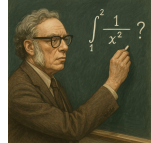
Example Challenge

$$\lim_{x \rightarrow 0} \left(\frac{2 \cdot \tan\left(\frac{x}{2}\right)}{x} \right)^{\frac{3}{x^2}}$$

Answer

$$e^{\frac{1}{4}}$$

100 Seeds → 35,368 Challenges



Example Challenge

$$\lim_{x \rightarrow 0} \left(\frac{2 \cdot \tan\left(\frac{x}{2}\right)}{x} \right)^{\frac{3}{x^2}}$$

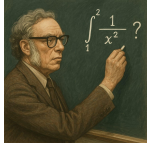
$$\lim_{x \rightarrow 0} A \cdot \left(\frac{2 \cdot \tan\left(\frac{B \cdot x}{2}\right)}{B \cdot x} \right)^{\frac{C \cdot 3}{(B \cdot x)^2}}$$

Answer

$$e^{\frac{1}{4}}$$

$$A \cdot e^{\frac{C}{4}}$$

100 Seeds → 35,368 Challenges



Example Challenge

$$\lim_{x \rightarrow 0} \left(\frac{2 \cdot \tan\left(\frac{x}{2}\right)}{x} \right)^{\frac{3}{x^2}}$$

$$\lim_{x \rightarrow 0} A \cdot \left(\frac{2 \cdot \tan\left(\frac{B \cdot x}{2}\right)}{B \cdot x} \right)^{\frac{C \cdot 3}{(B \cdot x)^2}}$$

$$\lim_{x \rightarrow 0} 17 \cdot \left(\frac{2 \cdot \tan\left(\frac{91 \cdot x}{2}\right)}{91 \cdot x} \right)^{\frac{57 \cdot 3}{(91 \cdot x)^2}}$$

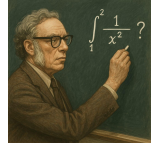
Answer

$$e^{\frac{1}{4}}$$

$$A \cdot e^{\frac{C}{4}}$$

$$17 \cdot e^{\frac{57}{4}}$$

100 Seeds → 35,368 Challenges



Example Challenge

$$\lim_{x \rightarrow 0} \left(\frac{2 \cdot \tan\left(\frac{x}{2}\right)}{x} \right)^{\frac{3}{x^2}}$$

Answer

$$e^{\frac{1}{4}}$$

$$\lim_{x \rightarrow 0} A \cdot \left(\frac{2 \cdot \tan\left(\frac{B \cdot x}{2}\right)}{B \cdot x} \right)^{\frac{C \cdot 3}{(B \cdot x)^2}}$$

$$A \cdot e^{\frac{C}{4}}$$

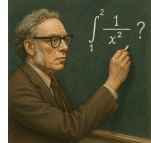
$$\lim_{x \rightarrow 0} 17 \cdot \left(\frac{2 \cdot \tan\left(\frac{91 \cdot x}{2}\right)}{91 \cdot x} \right)^{\frac{57 \cdot 3}{(91 \cdot x)^2}}$$

$$17 \cdot e^{\frac{57}{4}}$$

$$\lim_{x \rightarrow 0^+} \left(\frac{2 \cdot \tan\left(\frac{x}{2}\right)}{x} \right)^{\frac{(\sin^2(-Fx) + \cos^2(Fx)) \cdot 3}{x^2}}$$

$$e^{\frac{1}{4}}$$

100 Seeds → 35,368 Challenges



Example Challenge

$$\lim_{x \rightarrow 0} \left(\frac{2 \cdot \tan\left(\frac{x}{2}\right)}{x} \right)^{\frac{3}{x^2}}$$

Answer

$$e^{\frac{1}{4}}$$

$$\lim_{x \rightarrow 0} A \cdot \left(\frac{2 \cdot \tan\left(\frac{B \cdot x}{2}\right)}{B \cdot x} \right)^{\frac{C \cdot 3}{(B \cdot x)^2}}$$

$$A \cdot e^{\frac{C}{4}}$$

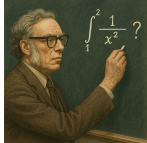
$$\lim_{x \rightarrow 0} 17 \cdot \left(\frac{2 \cdot \tan\left(\frac{91 \cdot x}{2}\right)}{91 \cdot x} \right)^{\frac{57 \cdot 3}{(91 \cdot x)^2}}$$

$$17 \cdot e^{\frac{57}{4}}$$

$$\lim_{x \rightarrow 0^+} \left(\frac{2 \cdot \tan\left(\frac{x}{2}\right)}{x} \right)^{\overbrace{\frac{(\sin^2(-Fx) + \cos^2(Fx)) \cdot 3}{x^2}}^1}$$

$$e^{\frac{1}{4}}$$

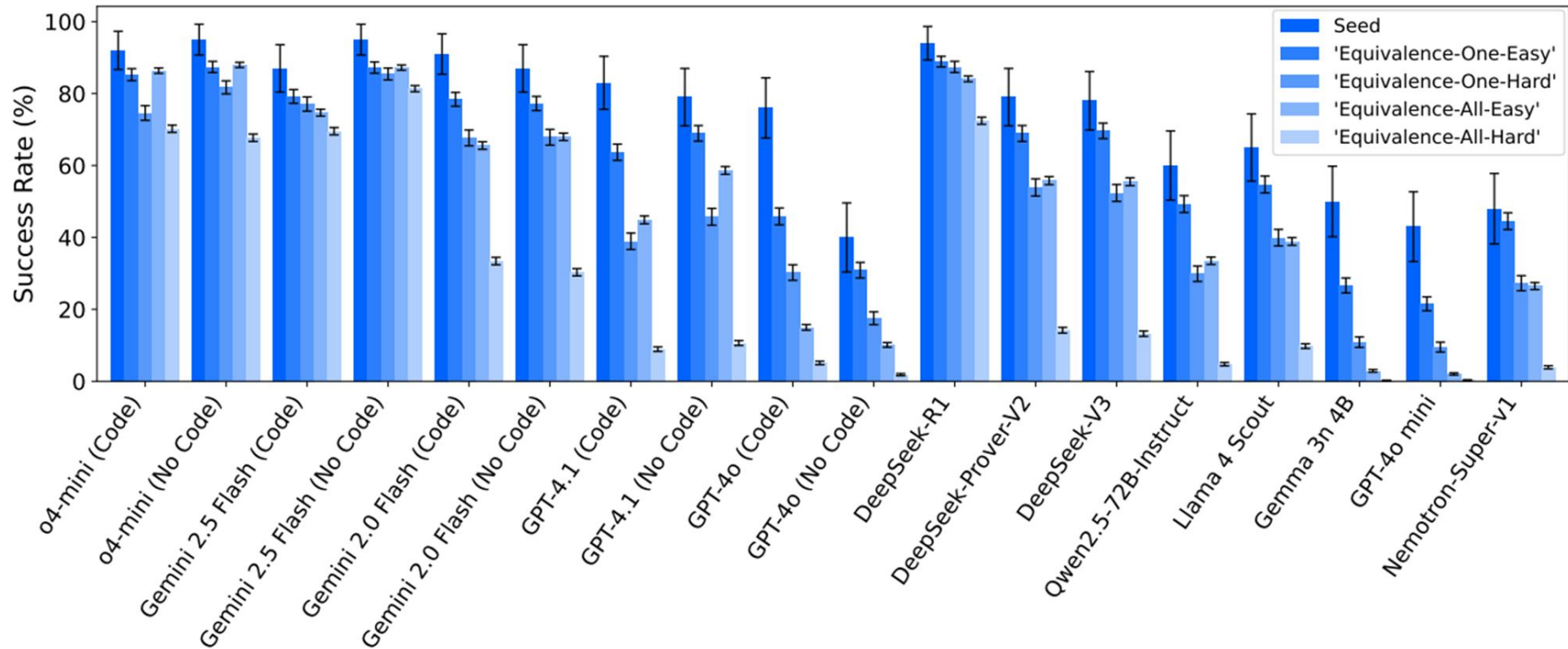
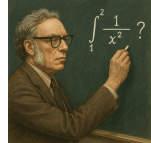
Tested Models



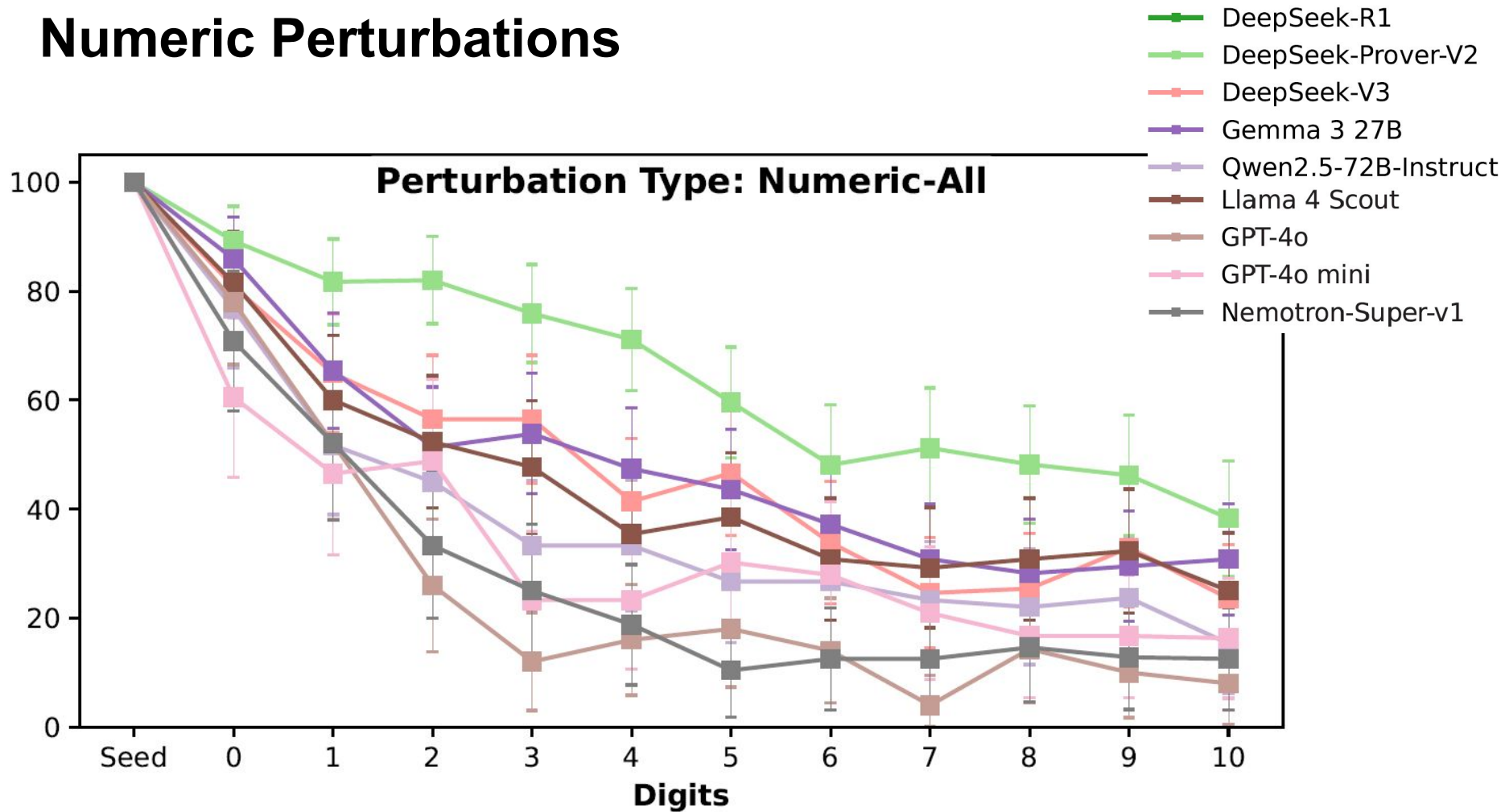
- Gemini 2.5 Flash (Code)
- - Gemini 2.5 Flash (No Code)
- Gemini 2.0 Flash (Code)
- - Gemini 2.0 Flash (No Code)
- o4-mini (Code)
- - o4-mini (No Code)
- GPT-4.1 (Code)
- - GPT-4.1 (No Code)

- DeepSeek-R1
- DeepSeek-Prover-V2
- DeepSeek-V3
- Gemma 3 27B
- Qwen2.5-72B-Instruct
- Llama 4 Scout
- GPT-4o
- GPT-4o mini
- Nemotron-Super-v1

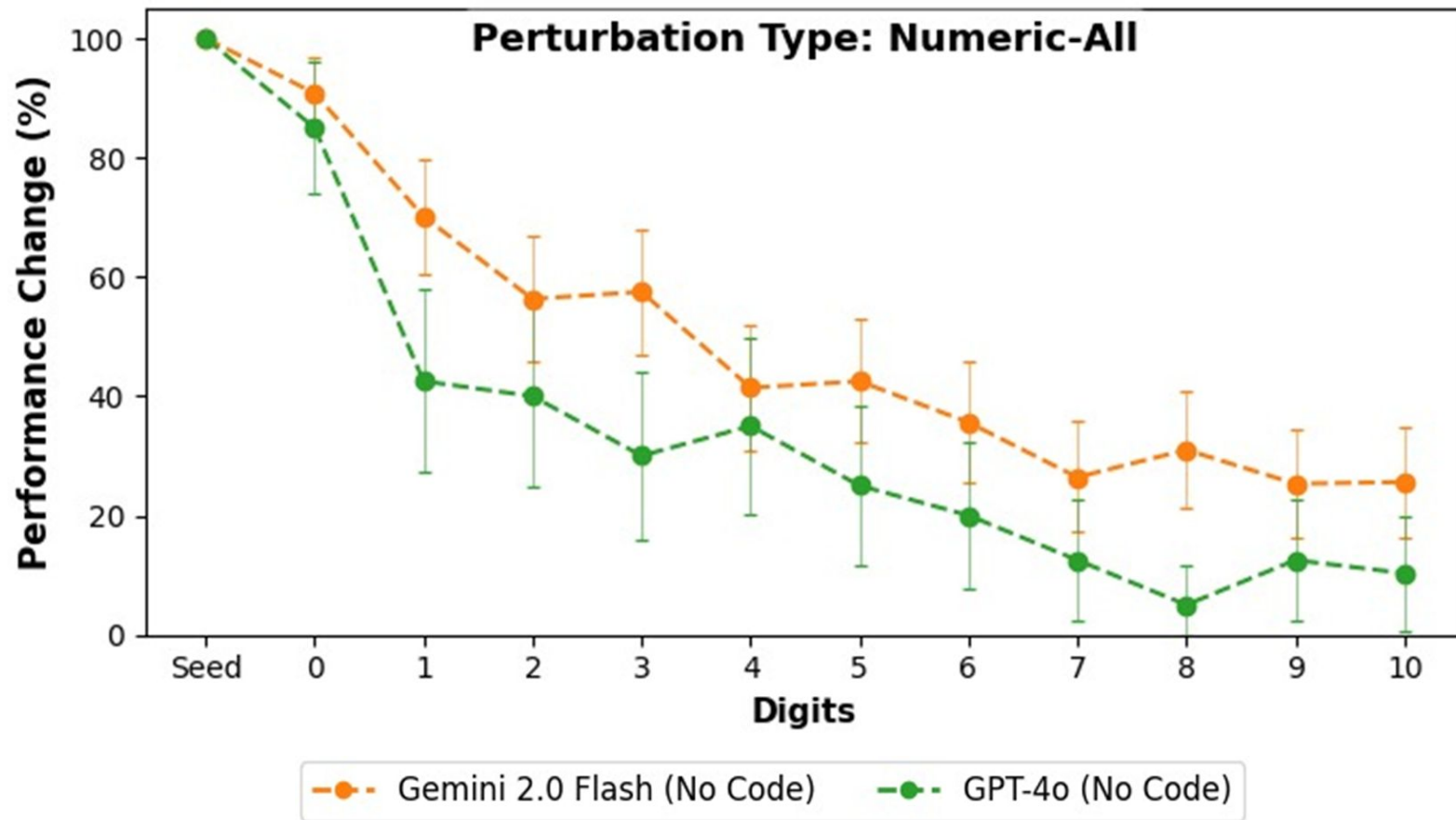
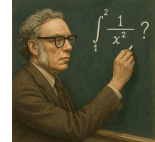
Equivalence Perturbations



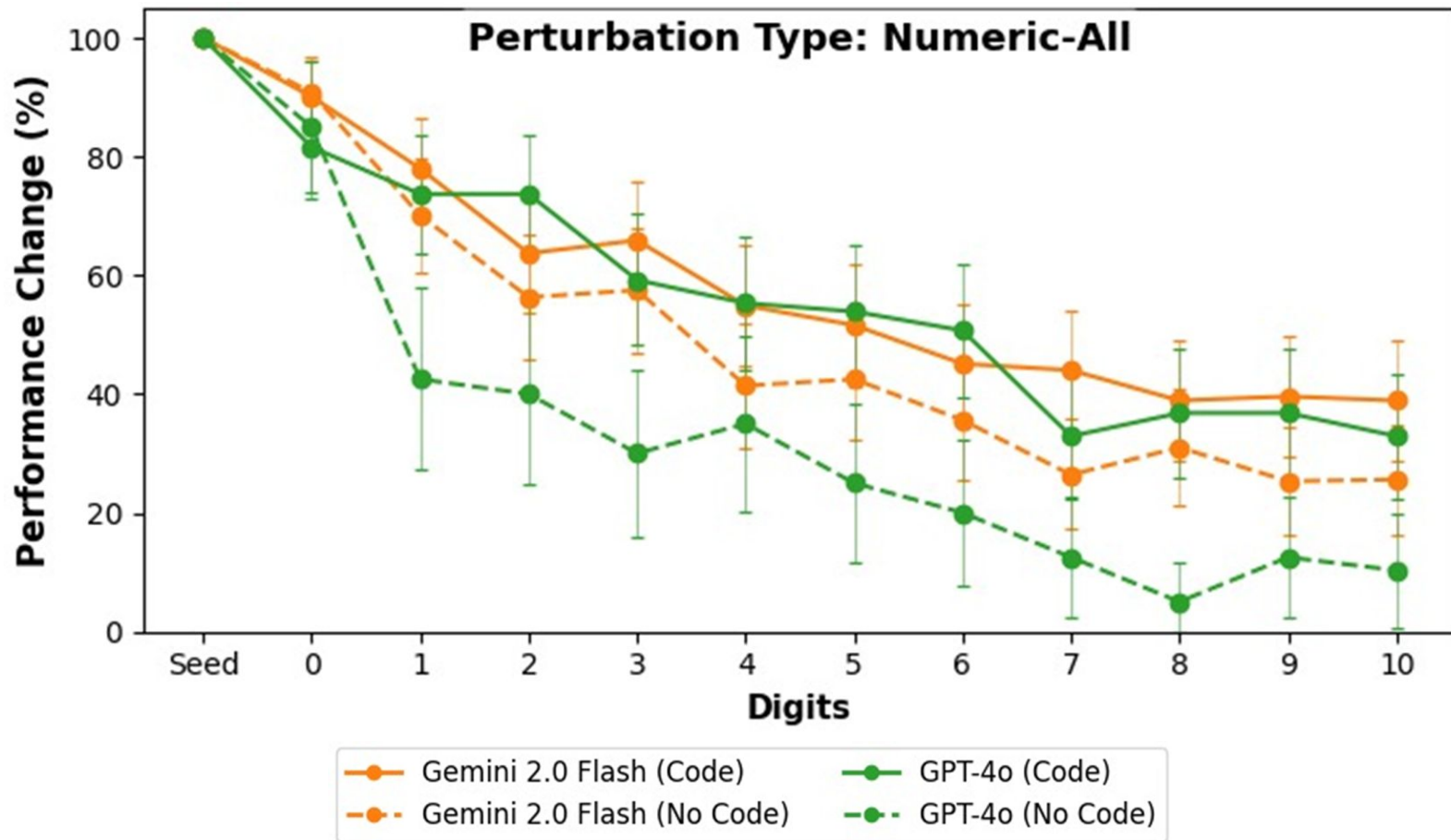
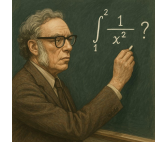
Numeric Perturbations



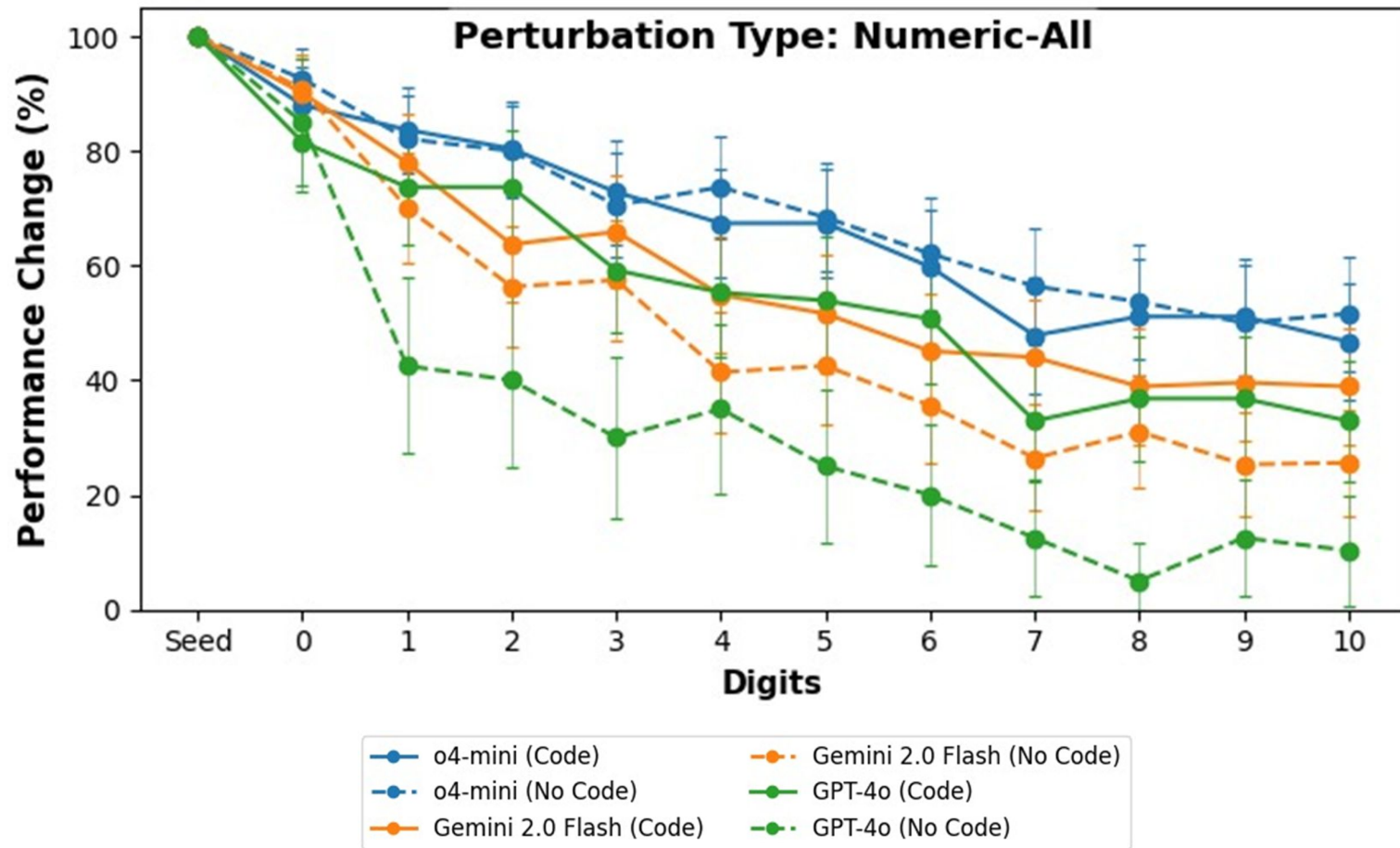
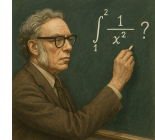
Code Use



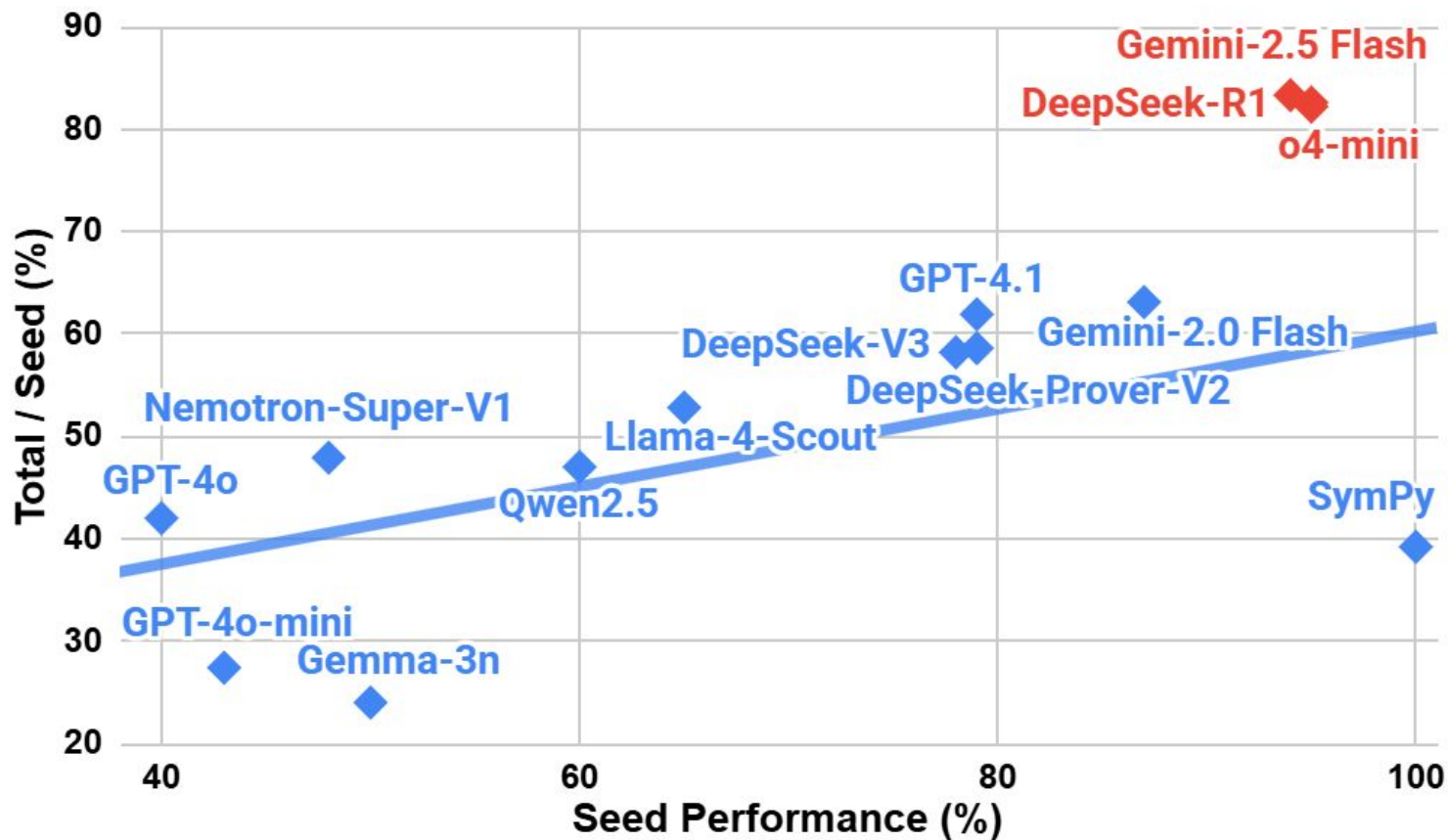
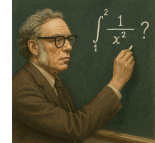
Code Use



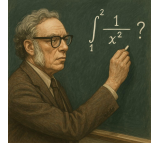
Code Use



Regime Shift in Robustness



Hybrid Approach - LLM+CAS



LLM Only

Prompt: Assume you don't have access to a computer: do not use code, solve this manually - using your internal reasoning. Solve the following integral:

$$\int \frac{(2m+(2m-n)x^n)x^{m-1}}{2(1+x^n+x^{2m})\sqrt{1+x^n}} dx$$

Response:

...Try substitution $u = 1 + x^n$

...But this seems to be getting messier

...Final Answer: $\frac{x^m}{\sqrt{1+x^n}}$

Prompt: Solve the integral:

$$\int \frac{(2m+(2m-n)(u-1))(u-1)^{(m-n)/n}}{2n(u+(u-1)^{2m/n})\sqrt{u}} du$$

Response:

...for general m, n , it cannot be integrated in closed form

LLM + CAS

Prompt: Please solve the following integral. First simplify it manually as much as possible, specifically perform a non-trivial substitution that simplifies the challenge, and then write Mathematica code that solves the simplified version.

$$\int \frac{(2m+(2m-n)x^n)x^{m-1}}{2(1+x^n+x^{2m})\sqrt{1+x^n}} dx$$

Response:

...Try substitution $u = 1 + x^n$

...Final Simplified Integral:

$$\int \frac{(2m+(2m-n)(u-1))(u-1)^{(m-n)/n}}{2n(u+(u-1)^{2m/n})\sqrt{u}} du$$

Mathematica Code:...

Integrate[
 $((2m+(2m-n)(u-1))(u-1)^{(m-n)/n}) / (2n(u + (u-1)^{(2m/n)}) \text{Sqrt}[u]), u]$
]

CAS Only



Prompt:

Integrate[
 $(2m x^{(m-1)+(2m-n) x^{(m+n-1)}} / (2(1+x^n+x^{2m})) \text{Sqrt}[1+x^n], x]$
]

Response:

$$\frac{1}{2} \int \frac{2 m x^{-1+m} + (2 m - n) x^{-1+m+n}}{\sqrt{1+x^n} (1+x^{2m}+x^n)} dx$$

<Left Unsolved>

Prompt:

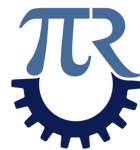
Integrate[
 $((2m+(2m-n)(u-1))(u-1)^{(m-n)/n}) / (2n(u + (u-1)^{(2m/n)}) \text{Sqrt}[u]), u]$
]

Response:

$$\text{ArcTan}\left[\frac{(-1+u)^{m/n}}{\sqrt{u}}\right]$$

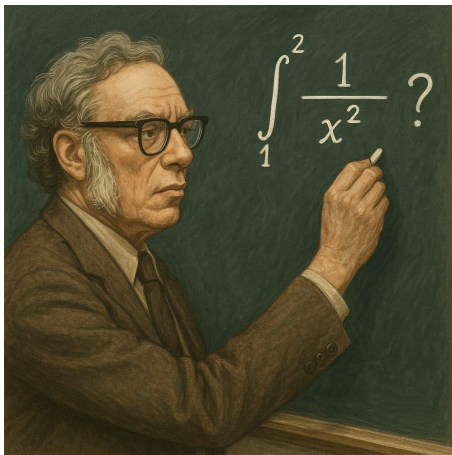


Technion
Israel Institute of Technology



**Ramanujan
Machine**

ASyMOB: Algebraic Symbolic Mathematical Operations Benchmark



Paper



Code



Dataset

