

# ORCA Benchmark Shows That AI Frequently Fumbles Everyday Math

KRAKÓW, MAŁOPOLSKA, POLAND, November 7, 2025 /EINPresswire.com/ -- [Omni Calculator](#) has introduced the [ORCA \(Omni Research on Calculation in AI\) Benchmark](#) - a new empirical study measuring how reliably today's leading AI chatbots can handle ordinary, day-to-day calculations. The evidence indicates that even very simple math frequently goes wrong.

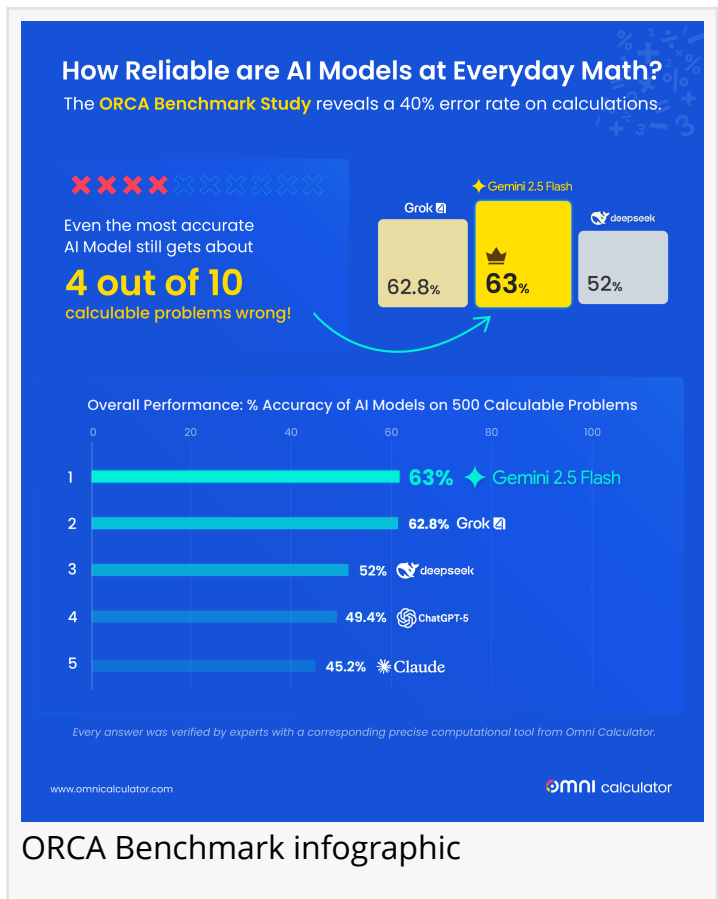
Researchers tested five widely used AI systems on a dataset of 500 realistic numeric problems. No model reached 63% accuracy. Even the top-scoring system still returned incorrect numbers almost 40% of the time. And most of these failures were not complex reasoning challenges - they were basic mistakes: arithmetic slips, wrong formulas, unit misunderstandings, or rounding that pushed the answer outside acceptable tolerance.

According to lead author Joanna Śmietańska-Nowak - who holds a PhD in Physics and postgraduate Machine Learning specialization - this isn't surprising if you understand the architecture.

"AIs are pattern recognizers trained on text, not symbolic reasoners built to perform arithmetic," Śmietańska-Nowak explained. "They learn statistical associations between tokens rather than mathematical rules, which is why they produce answers that sound logical but are numerically wrong."

## Key Findings:

- none of the evaluated models (ChatGPT, Gemini, Claude, DeepSeek, Grok) exceeded 63% total accuracy
- 68% of all errors were mechanical - 33% were arithmetic mistakes and 35% were precision/rounding faults



- weakest domains were Physics and Health/Sports, which required selecting the correct equation from context
- Finance & Economics performance varied wildly - some models achieved ~70–80%, others fell below 40%

## The Bottom Line for Users

The ORCA Benchmark reinforces a simple, protective rule for anyone using AI for numbers: treat any calculation the model produces as a draft, not a final result. If the answer will be used to make a choice - whether that is a mortgage estimate, comparing loan terms, projecting a savings horizon, estimating supplement or medication quantities, computing pace/volume in training, or scaling ingredients for a household recipe - treat the AI's math as a first draft only. Before using the answer to make a real decision, double-check the calculation with a dedicated calculator or trusted domain tool.

[Download the Full ORCA Benchmark Study Here](#)

## About Omni Calculator:

Omni Calculator has been building tools to solve real-world problems for over a decade. Our team of experts creates reliable calculators that empower people to make informed decisions. The ORCA Benchmark is a natural extension of our mission to bring clarity and accuracy to everyday calculations.

Dawid Siuda

Omni Calculator

+48 519839921

dawid@omnicalculator.com

Visit us on social media:

[LinkedIn](#)

---

This press release can be viewed online at: <https://www.einpresswire.com/article/865007637>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2025 Newsmatics Inc. All Right Reserved.