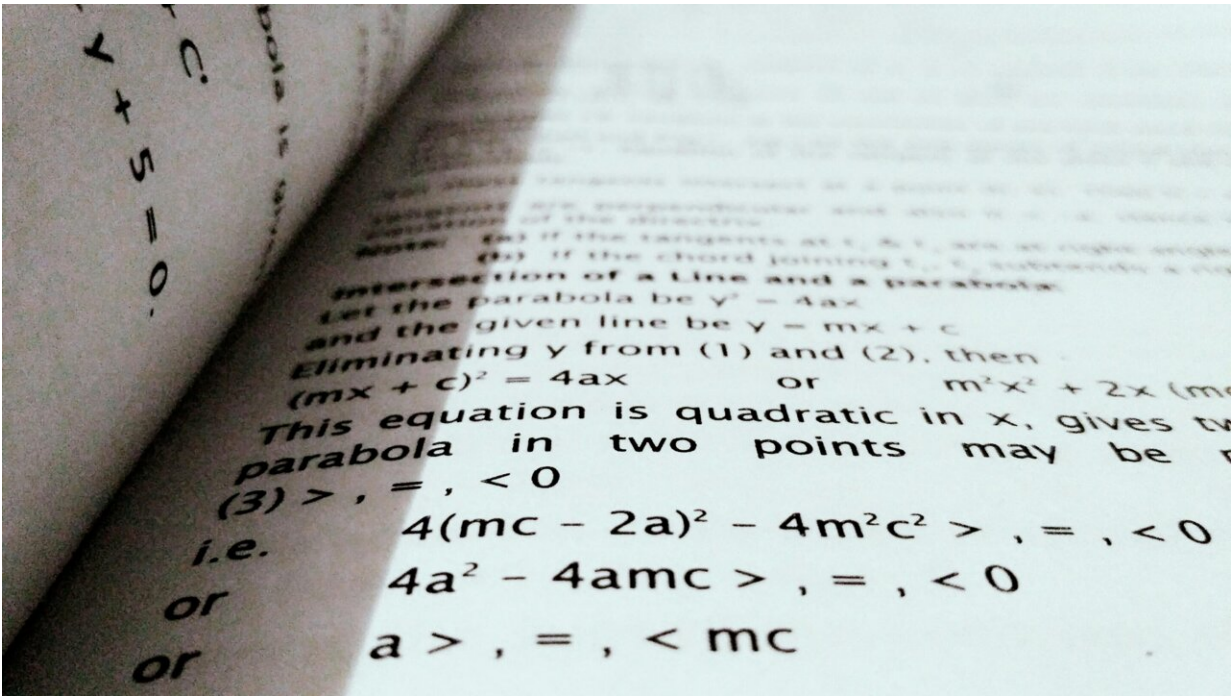


# Microsoft's small language model outperforms larger models on standardized math tests

March 8 2024, by Bob Yirka



Credit: Deepak Gautam from Pexels

A small team of AI researchers at Microsoft reports that the company's Orca-Math small language model outperforms other, larger models on standardized math tests. The group has published a [paper](#) on the *arXiv* preprint server describing their testing of Orca-Math on the Grade

School Math 8K (GSM8K) benchmark and how it fared compared to well-known LLMs.

Many popular LLMs such as ChatGPT are known for their impressive conversational skills—less well known is that most of them can also solve math word problems. AI researchers have tested their abilities at such tasks by pitting them against the GSM8K, a dataset of 8,500 grade-school math word problems that require multistep reasoning to solve, along with their correct answers.

In this new study, the research team at Microsoft tested Orca-Math, an AI application developed by another team at Microsoft specifically designed to tackle math word problems, and compared the results with larger AI models.

Microsoft points out on its [Research Blog post](#) that there is a major difference between popular LLMs such as ChatGPT and Orca-Math. The former is a large language model and the latter is a small language model—the difference is in the number of parameters that are used; typically in the thousands or a few million for SLMs, rather than the billions or trillions used by LLMs. Another difference is that, as its name suggests, Orca-Math was designed specifically to solve [math](#) problems; thus, it cannot be used to carry on conversations or answer random questions.

Orca-Math is relatively large compared to other SLMs, with 7 billion parameters, but still much smaller than most of the well-known LLMs. However, it still managed to score 86.81% on the GSM8k, close to GPT-4-0613, which got 97.0%. Others, such as Llama-2, did not fare nearly as well, with scores as low as 14.6%.

Microsoft reveals that it was able to garner such a high score by using higher-quality training data than is available to general-use LLMs and

because it used an interactive learning process the AI team at Microsoft has been developing—a process that continually improves results by using feedback from a teacher. The team at Microsoft concludes that SLMs can perform as well as LLMs on certain applications when developed under specialized conditions.

**More information:** Arindam Mitra et al, Orca-Math: Unlocking the potential of SLMs in Grade School Math, *arXiv* (2024). [DOI: 10.48550/arxiv.2402.14830](https://doi.org/10.48550/arxiv.2402.14830)

Orca-Math: [www.microsoft.com/en-us/research/blog/odel-specialization/](https://www.microsoft.com/en-us/research/blog/odel-specialization/)  
[twitter.com/Arindam1408/status/1764761895473762738](https://twitter.com/Arindam1408/status/1764761895473762738)

© 2024 Science X Network

Citation: Microsoft's small language model outperforms larger models on standardized math tests (2024, March 8) retrieved 25 May 2024 from <https://techxplore.com/news/2024-03-microsoft-small-language-outperforms-larger.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.