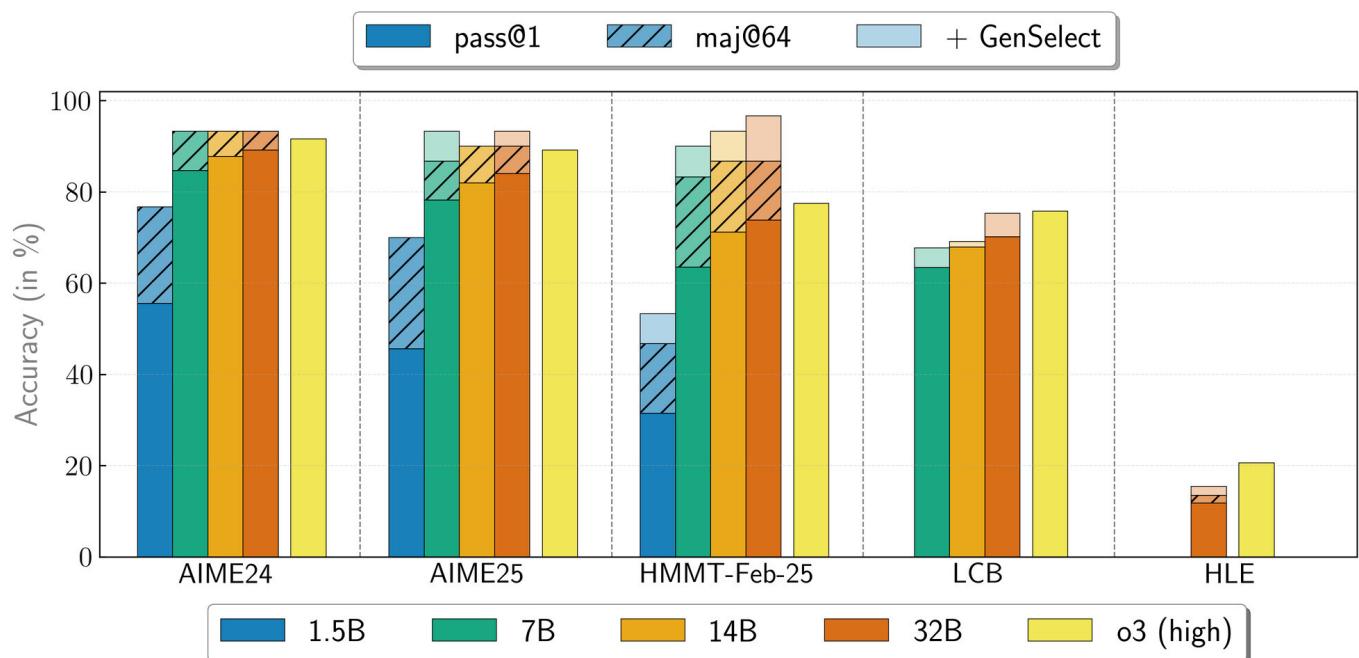


NVIDIA introduced a new line of local AI models Openreasoning-Nemotron, capable of solving problems at the Openai level.

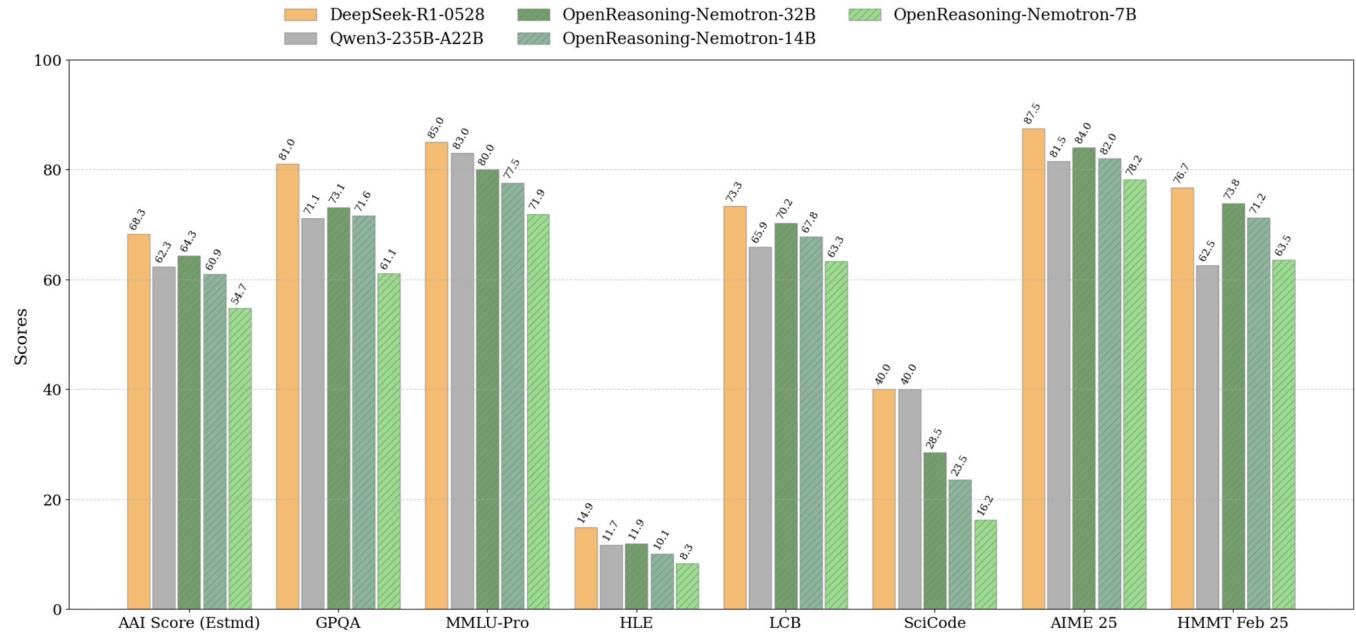
The series includes four models with 1.5, 7, 14 and 32 billion parameters. All of them are created on the basis of a large 671-milliard model of Deepseek, but in much more compact form, which allows you to run them on ordinary gaming graphics cards.

For learning, NVIDIA used 5 million tasks in mathematics, science and programming created using her Nemo Skills platform.

Models were studied exclusively with the use of supervised learning - without the use of RLHF, which makes them a convenient basis for further research.



Techpowerup



Techpowerup

According to the test results, the largest 32B model reaches 89.2 points at the AIME24 and 73.8 Olympics on Hmmt, and the youngest 1.5b shows 55.5 and 31.5, respectively. In the Genselect mode, which generates several answers in parallel and selects the best, performance of the 32B model is comparable or even exceeds the O3-High model from Openai.

All four NVIDIA models are already available on Hugging Face for loading and local use.