MMMLU, 1.7B model scale, 1T tokens of Nemotron-CC. Training context lengths from 4k to 16k tokens.

