	Claude 3.5 Sonnet	Claude 3 Opus	GPT-4o	Gemini 1.5 Pro
Graduate level reasoning *GPQA, Diamond*	**59.4%*** 0-shot CoT	**50.4%** 0-shot CoT	**53.6%** 0-shot CoT	—
Undergraduate level knowledge *MMLU*	**88.7%**** 5-shot **88.3%** 0-shot CoT	**86.8%** 5-shot **85.7%** 0-shot CoT	— **88.7%** 0-shot CoT	**85.9%** 5-shot —
Code *HumanEval*	**92.0%** 0-shot	**84.9%** 0-shot	**90.2%** 0-shot	**84.1%** 0-shot
Multilingual math *MGSM*	**91.6%** 0-shot CoT	**90.7%** 0-shot CoT	**90.5%** 0-shot CoT	**87.5%** 8-shot
Reasoning over text *DROP, F1 score*	**87.1** 3-shot	**83.1** 3-shot	**83.4** 3-shot	**74.9** Variable shots
Mixed evaluations *BIG-Bench-Hard*	**93.1%** 3-shot CoT	**86.8%** 3-shot CoT	—	**89.2%** 3-shot CoT
Math problem-solving *MATH*	**71.1%** 0-shot CoT	**60.1%** 0-shot CoT	**76.6%** 0-shot CoT	**67.7%** 4-shot
Grade school math *GSM8K*	**96.4%** 0-shot CoT	**95.0%** 0-shot CoT	—	**90.8%** 11-shot

* Claude 3.5 Sonnet scores 67.2% on 5-shot CoT GPQA with maj@32
** Claude 3.5 Sonnet scores 90.4% on MMLU with 5-shot CoT prompting