pass@k Comparison: AIME 0.575 Nemotron-Research-Reasoning-Qwen-1.5B 0.567 → DeepSeek-R1-Distill-Qwen-1.5B 0.557 0.545 0.55 -0.532 0.516 0.497 0.50 0.472 0.458 0.449 0.45 -0.439 0.435 0.426 0.411 bass@k 0.393 0.365 0.35 0.30 0.25 0.20 2 3 10 9 k (number of samples)