pass@k Comparison: MBPP 0.764 Nemotron-Research-Reasoning-Qwen-1.5B 0.760 0.755 → DeepSeek-R1-Distill-Qwen-1.5B 0.749 0.75 0.742 0.733 0.724 0.720 0.719 0.712 0.704 0.703 0.694 0.70 -0.681 0.672 0.65 pass@k 0.596 0.60 -0.55 0.50 -2 3 9 10 8 k (number of samples)