pass@k Comparison: HUMAN_EVAL 0.902 Nemotron-Research-Reasoning-Qwen-1.5B 0.897 0.90 -→ DeepSeek-R1-Distill-Qwen-1.5B 0.891 0.890 0.886 0.884 0.879 0.877 0.872 0.863 0.847 0.85 0.818 pass@k - 08.0 0.75 0.727 698 0.70 2 3 10 9 k (number of samples)