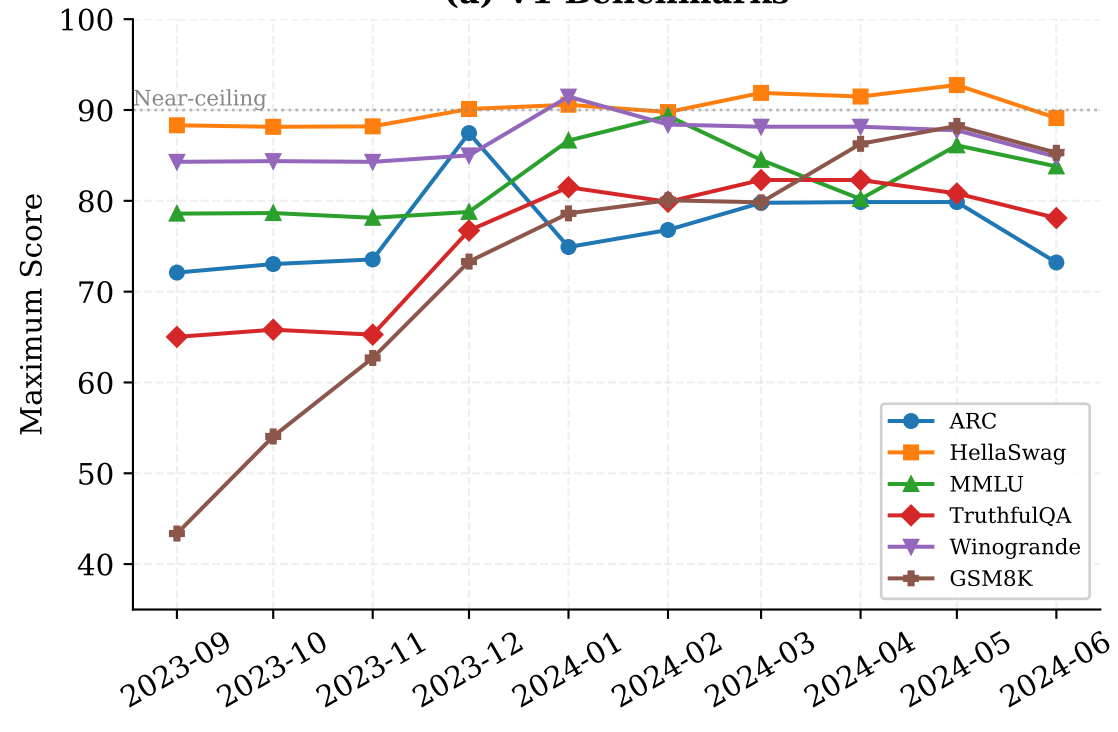**Figure 1: Maximum Score Trajectories Over Time**

(a) V1 Benchmarks

Legend: ARC, HellaSwag, MMLU, TruthfulQA, Winogrande, GSM8K

(b) V2 Benchmarks

Legend: IFEval, BBH, MATH Lvl 5, GPQA, MUSR, MMLU-PRO