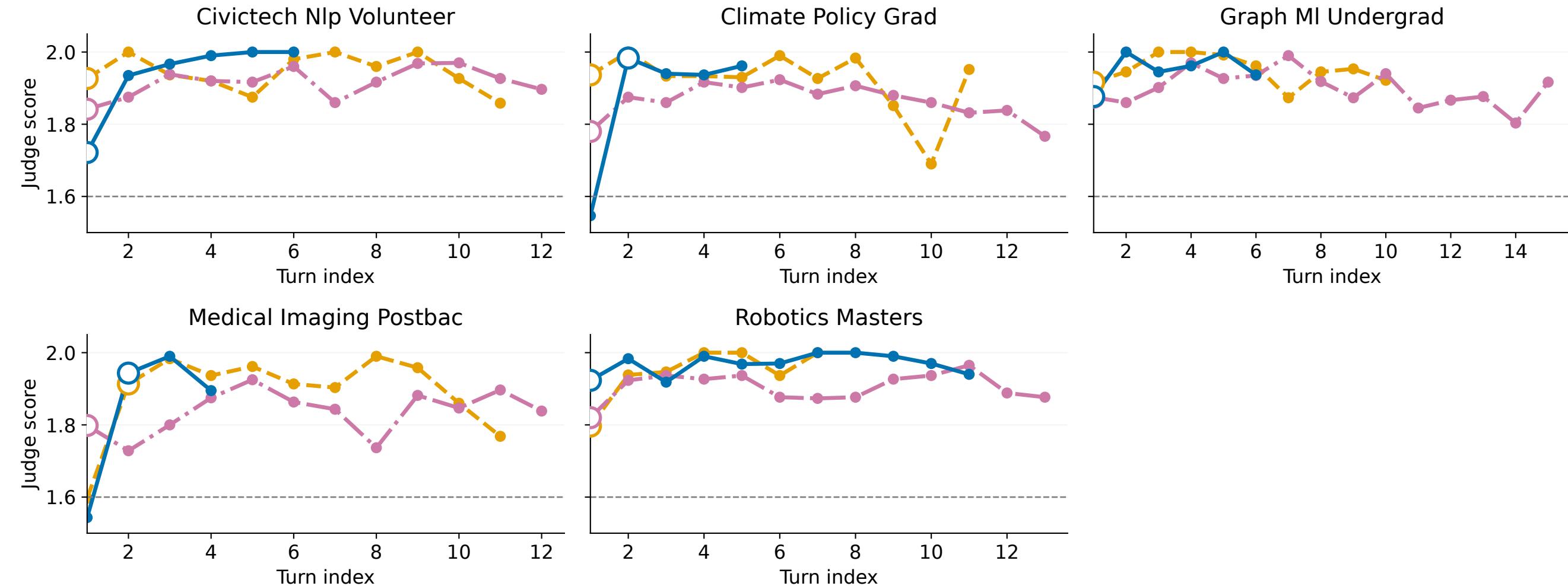


Per-scenario multi-turn trajectories



—●— Baseline: Claude Sonnet 4.5 —●— Baseline: GPT-5 —●— Mentor (Kimi-k2-0905)