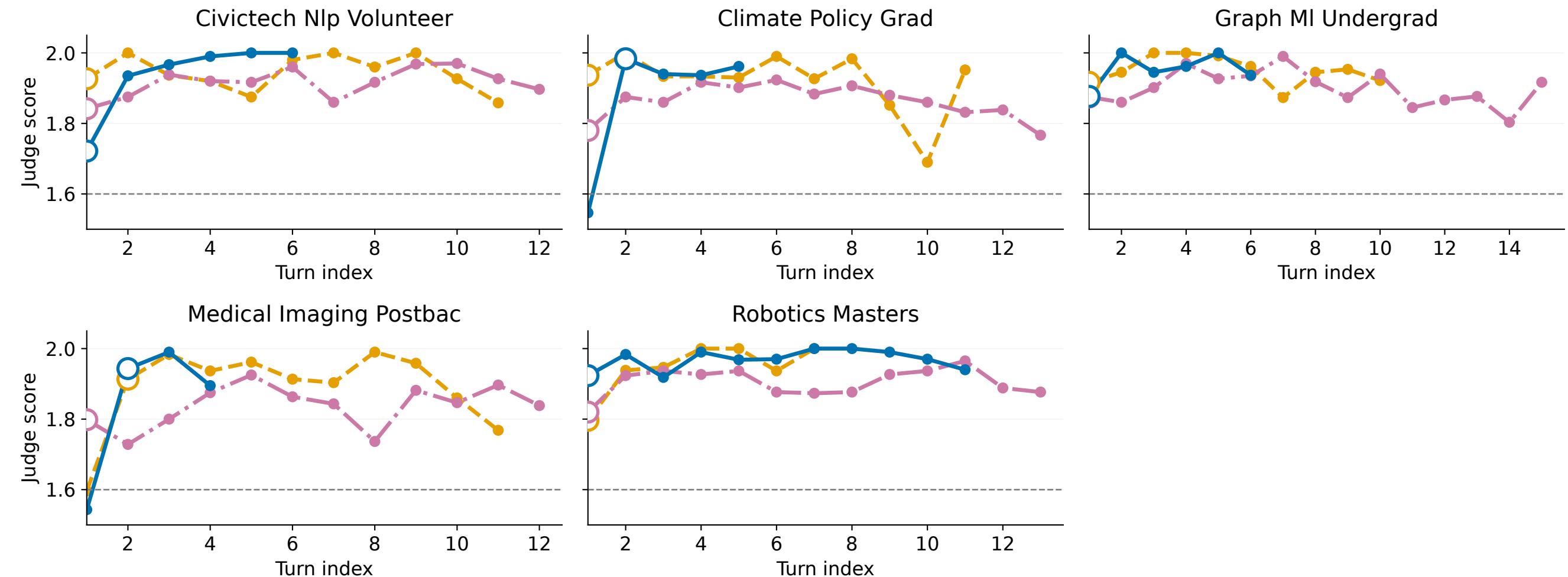


# Per-scenario multi-turn trajectories



Each panel shows one scenario. Lines correspond to Mentor (solid), Claude baseline (dashed), and GPT-5 baseline (dash-dot). Hollow markers indicate the first turn where score  $\geq 1.6$ .

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