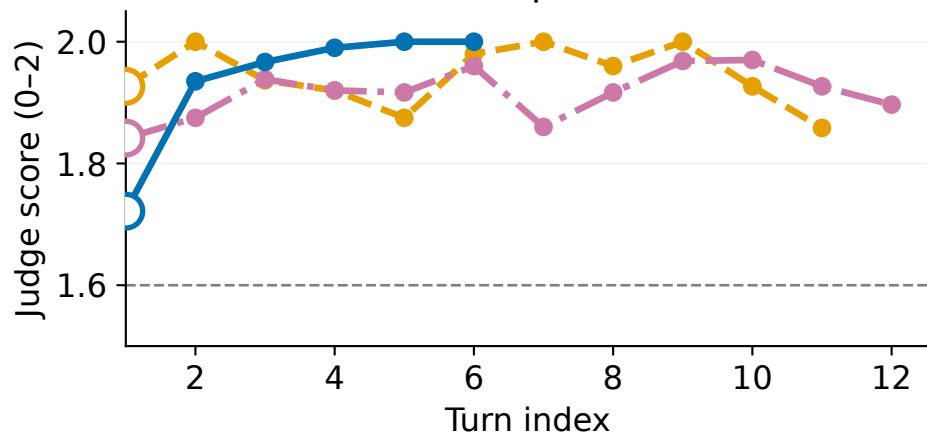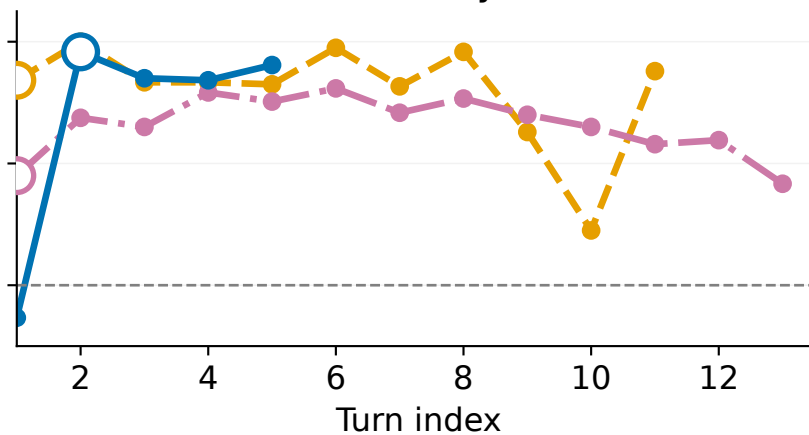Per-scenario multi-turn trajectories

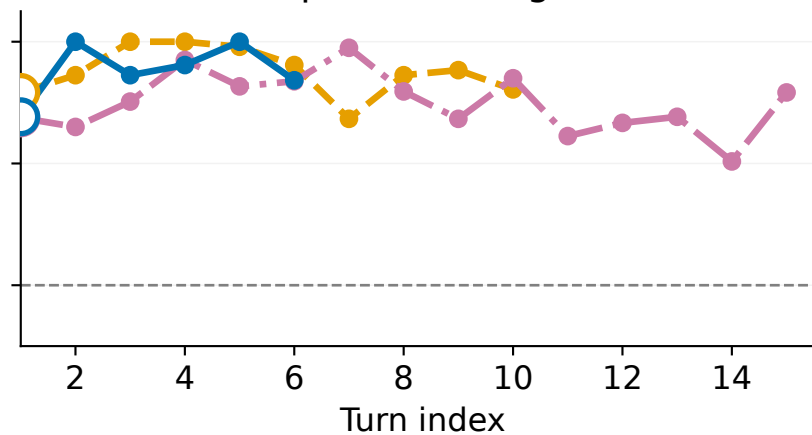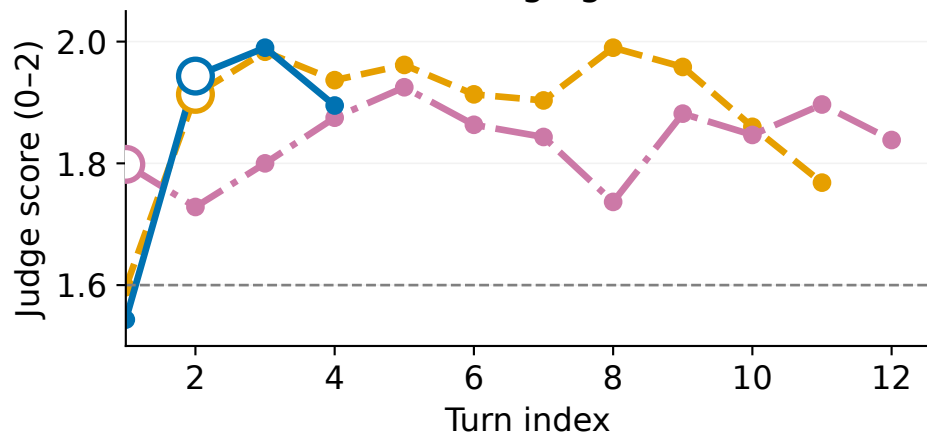Civictech Nlp Volunteer — Climate Policy Grad — Graph Ml Undergrad — Medical Imaging Postbac — Robotics Masters
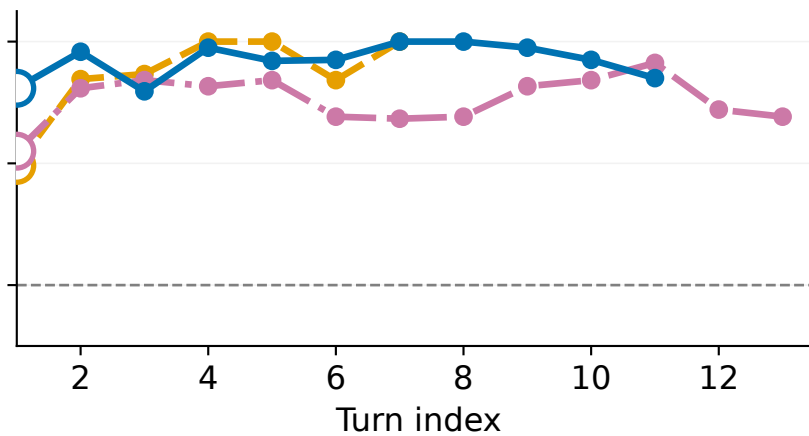
Judge score (0–2) vs Turn index

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