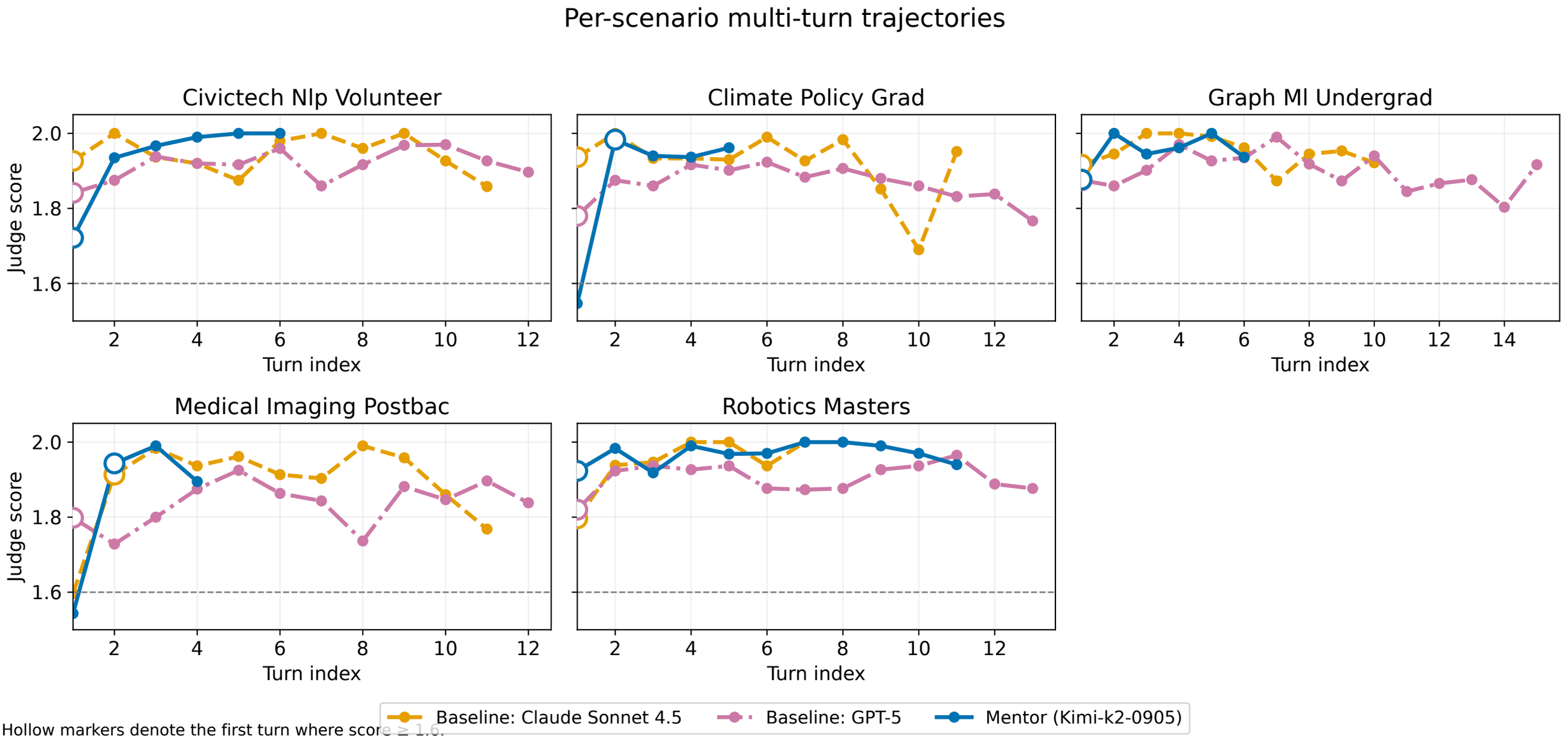
Per-scenario multi-turn trajectories
Civictech Nlp Volunteer
Climate Policy Grad
Graph Ml Undergrad
Medical Imaging Postbac
Robotics Masters
Judge score
Turn index
2.0
1.8
1.6
Baseline: Claude Sonnet 4.5
Baseline: GPT-5
Mentor (Kimi-k2-0905)
Hollow markers denote the first turn where score ≥ 1.6.