

**Reproducibility Statement.** We provide an anonymized artifact containing: (i) the scenario generator and the 54 scenario files (3 domains  $\times$  3 difficulties  $\times$  6 each), (ii) the mentor decision graphs (public/oracle variants) specified in a JSON schema, (iii) prompt templates for all methods (CoT, self-consistency, textual memo, graph\_grounded), (iv) the learned per-mentor, per-domain utility weights, and (v) result logs with per-decision Monte Carlo traces. All methods use the same GPT-3.5-class LLM backbone with fixed generation settings (temperature=0.2, top\_p=1.0, n=1; self-consistency uses  $k=5$  independent chains at temperature=0.7). For Mentor-Mind, feasible actions are scored via Monte Carlo with  $S=400$  samples/action; CVaR uses  $\alpha=0.10$ , and mean-CVaR uses mixing  $\lambda=0.30$  where specified. Random seeds are fixed and reported (e.g., {20250913, 101, 2024, 7, 11}); folds for weight learning are deterministic (3-fold, 6 scenarios/fold/domain) with a grid step of 0.1 under sign priors. Metrics are computed as defined: *Alignment* (mentor agreement), *Regret*  $= \mathbb{E}[U(a^*) - U(\hat{a})]$  estimated via Monte Carlo (small negative estimates may occur from sampling variance), and *HCVR* (hard-constraint violation rate). We report paired  $t$ -tests on scenario-level outcomes. Scripts to regenerate all tables and figures from logs are included; runs do not require specialized hardware (we used a single GPU for prompting and CPU for simulation; wall-clock times are logged).