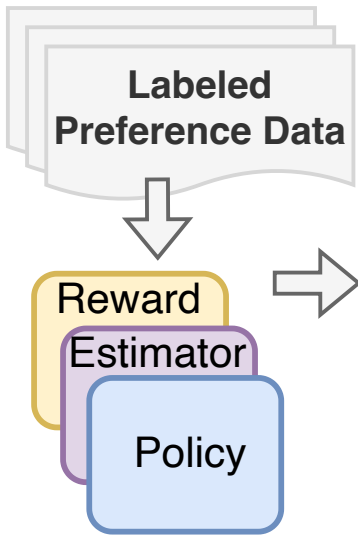
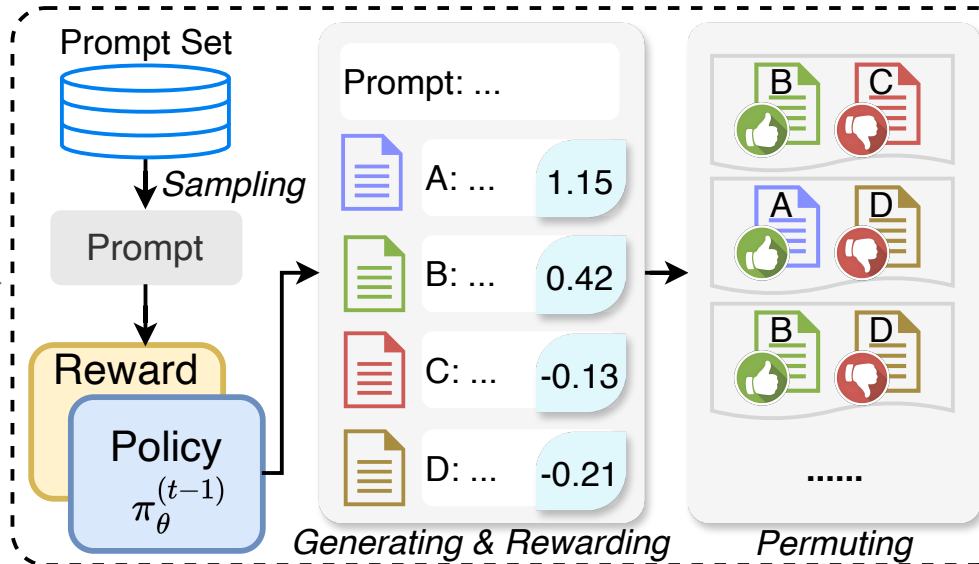


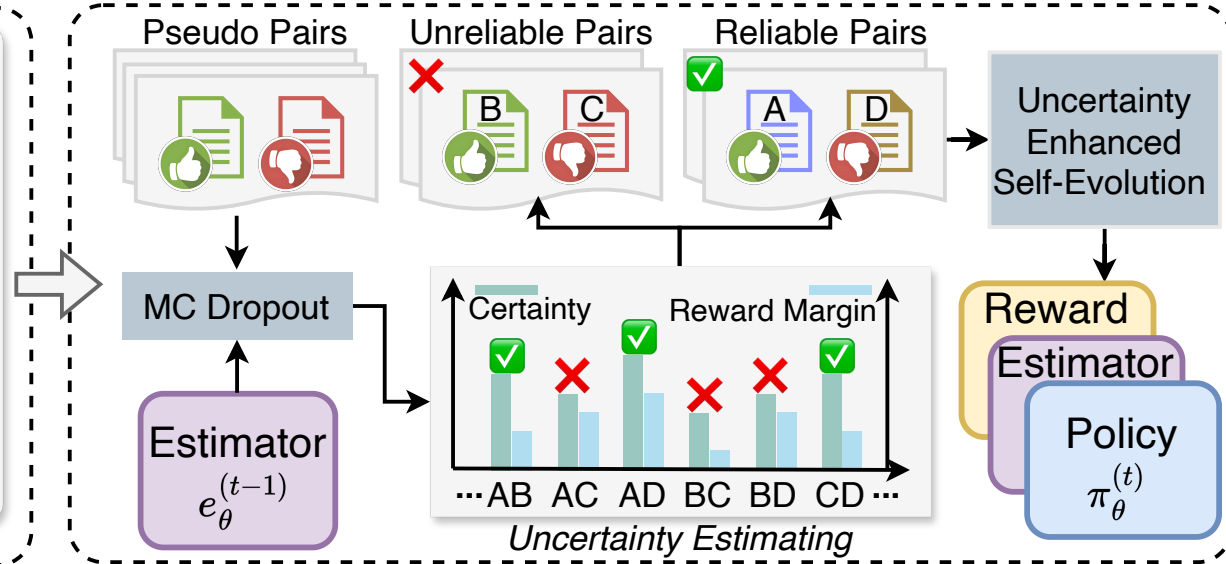
(1) Fine-tuning



(2) Generated Responses Rewarding



(3) Reliable Preference Learning



(4) Repeat Until Convergency