

Self-correction in Multi-hop Reasoning

samples: 400

