Reasoning Models Recover Performance Degradation Compared to Non-Reasoning Models on In-Context Learning Tasks

HELMET In-Context Learning Task: Clinic150 (16K Context)