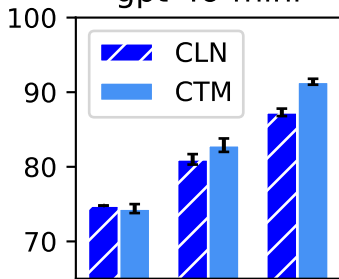
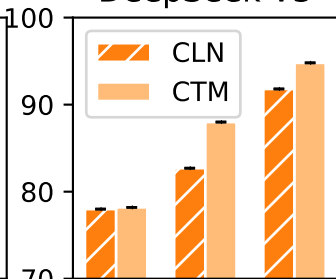


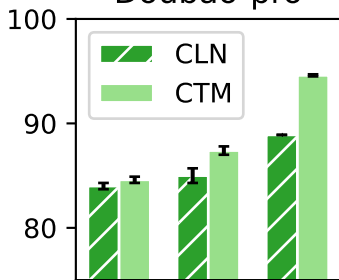
gpt-4o-mini



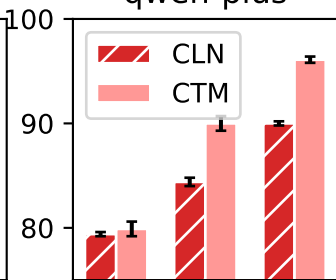
DeepSeek-V3



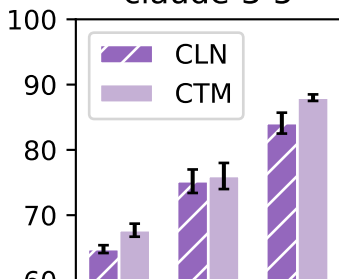
Doubao-pro



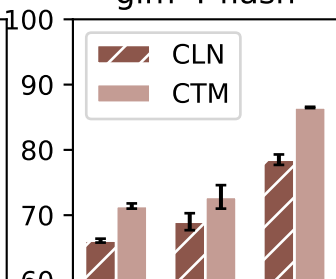
qwen-plus



claude-3-5



glm-4-flash



EdEval  
Auto-Dataset  
LatestEval

EdEval  
Auto-Dataset  
LatestEval