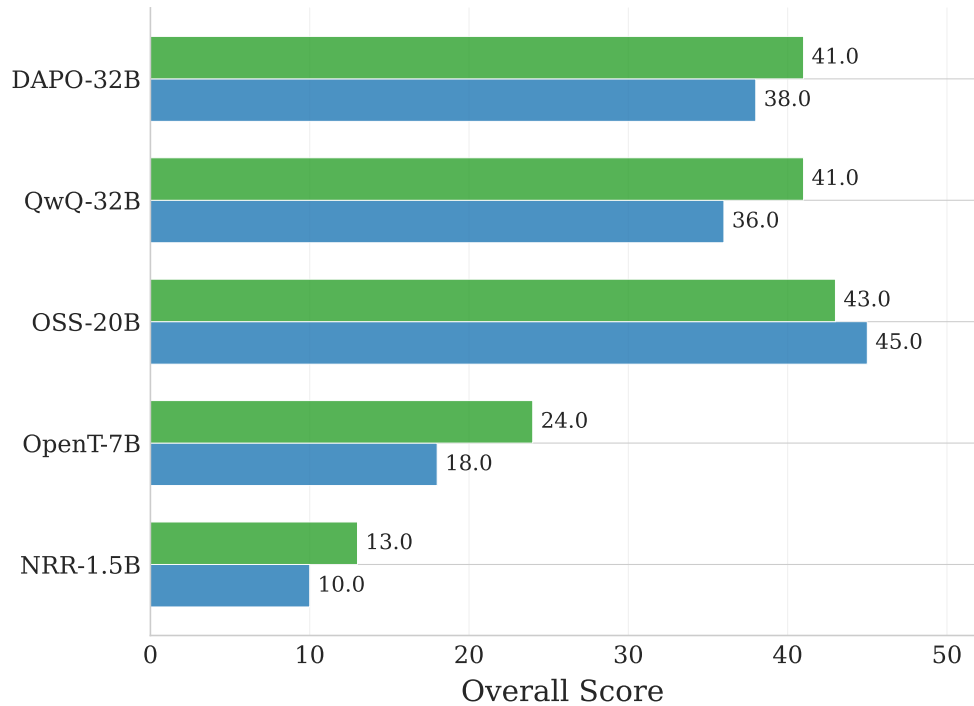
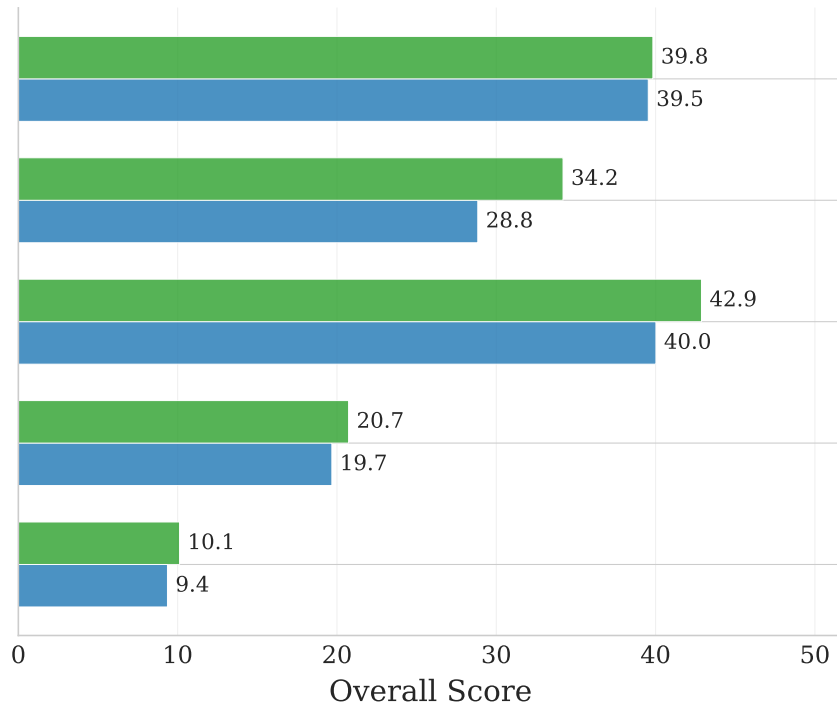


Chain-of-Thought Reasoning Performance Comparison

Sample Dataset



Full Dataset



Empty CoT Default CoT