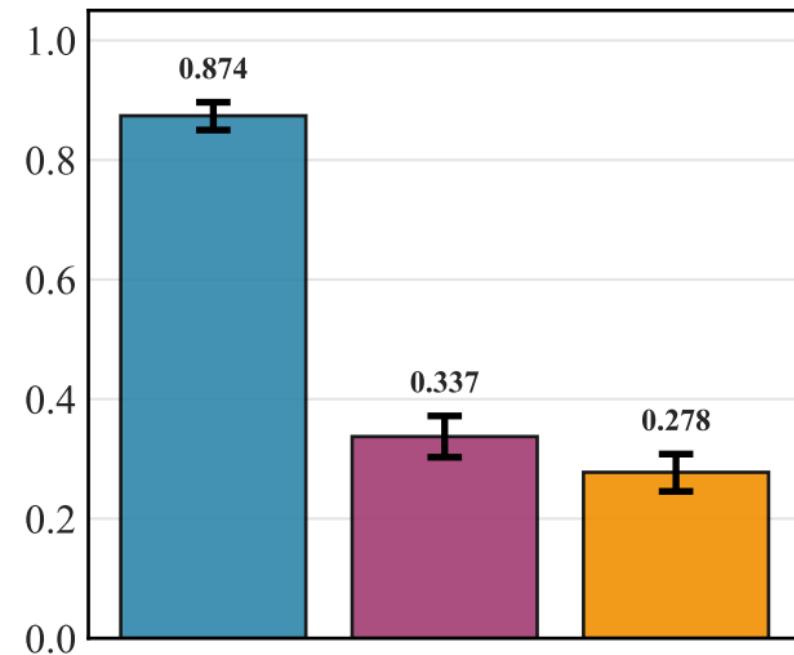
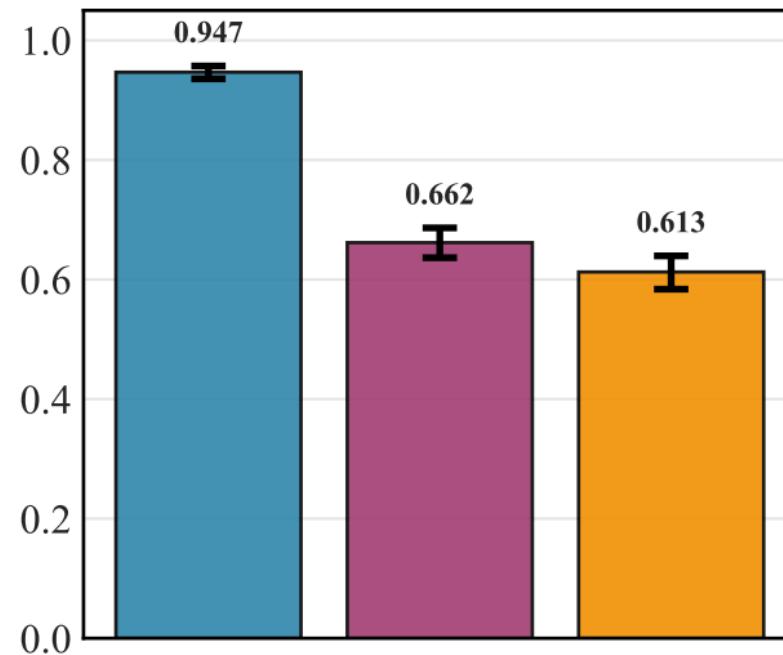


Test Set Performance (95% Bootstrap CI)

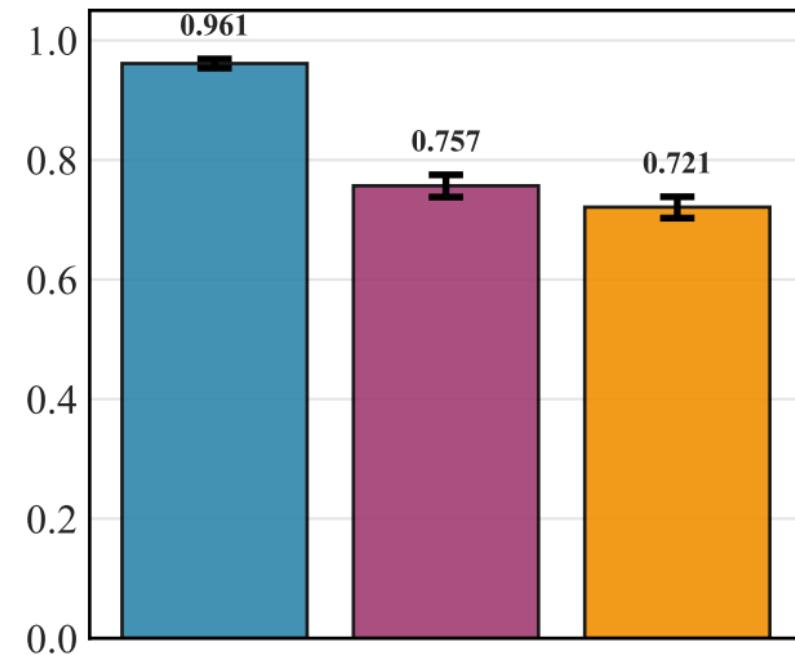
Word Accuracy



1 - CER



F1 Score



LLaMA-3.2-11B



Qwen-2.5-7B



Qwen-2.5-3B