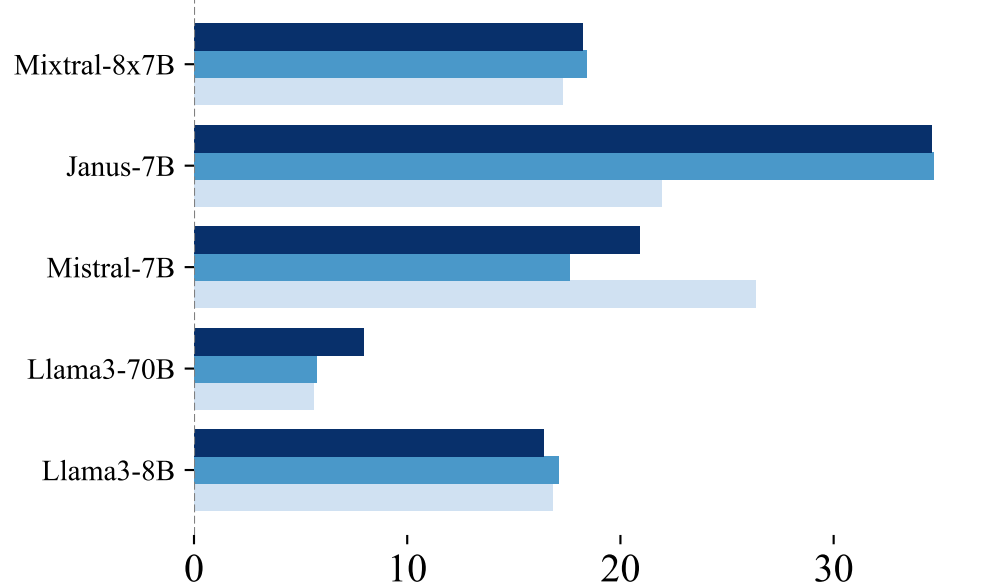
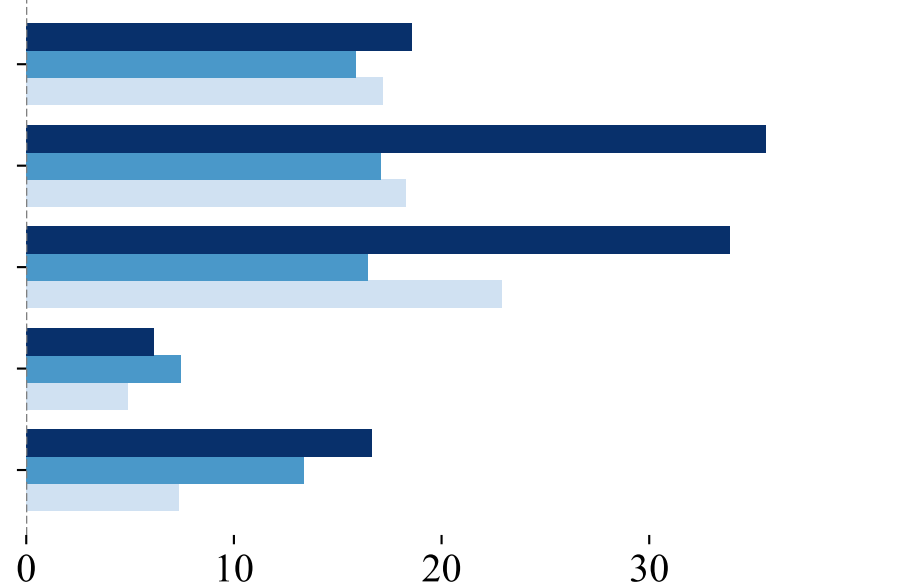


direct cot icl

Breakage Rate (BR) with Different Prompting Methods



Robustness Error Rate (RER) with Different Prompting Methods



Alignment Failure Rate (AFR)^p with Different Prompting Methods Performance Variation Rate (PVR)^e with Different Prompting Methods

