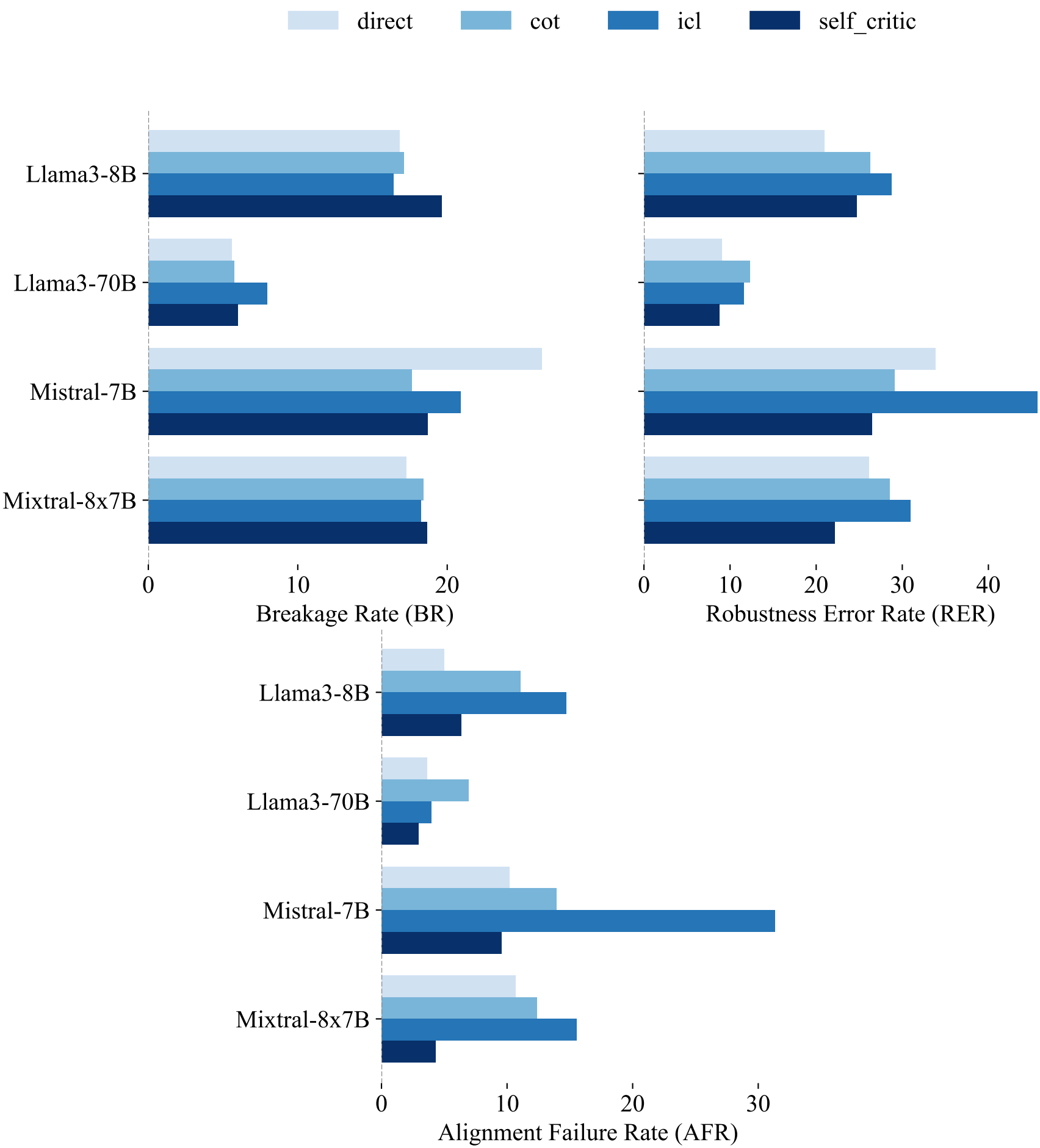
direct
cot
icl
self_critic
Llama3-8B
Llama3-70B
Mistral-7B
Mixtral-8x7B
0
10
20
Breakage Rate (BR)
0
10
20
30
40
Robustness Error Rate (RER)
Llama3-8B
Llama3-70B
Mistral-7B
Mixtral-8x7B
0
10
20
30
Alignment Failure Rate (AFR)