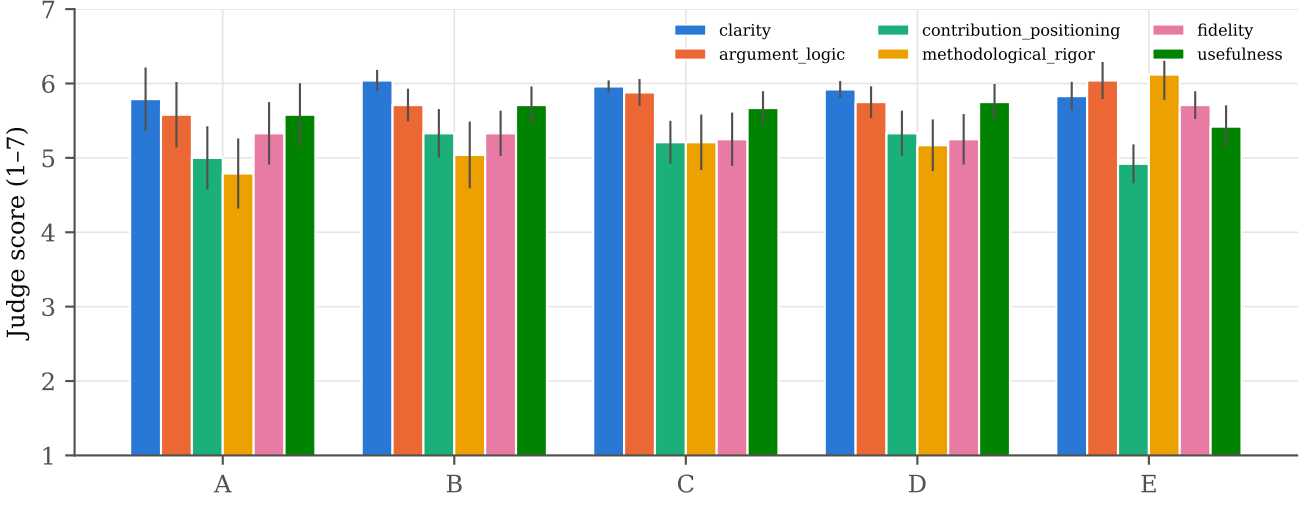

Blind judge scores by condition (finance held-out tasks; ±1.96 SE)
Judge score (1–7)
7
6
5
4
3
2
1
A
B
C
D
E
clarity
argument_logic
contribution_positioning
methodological_rigor
fidelity
usefulness