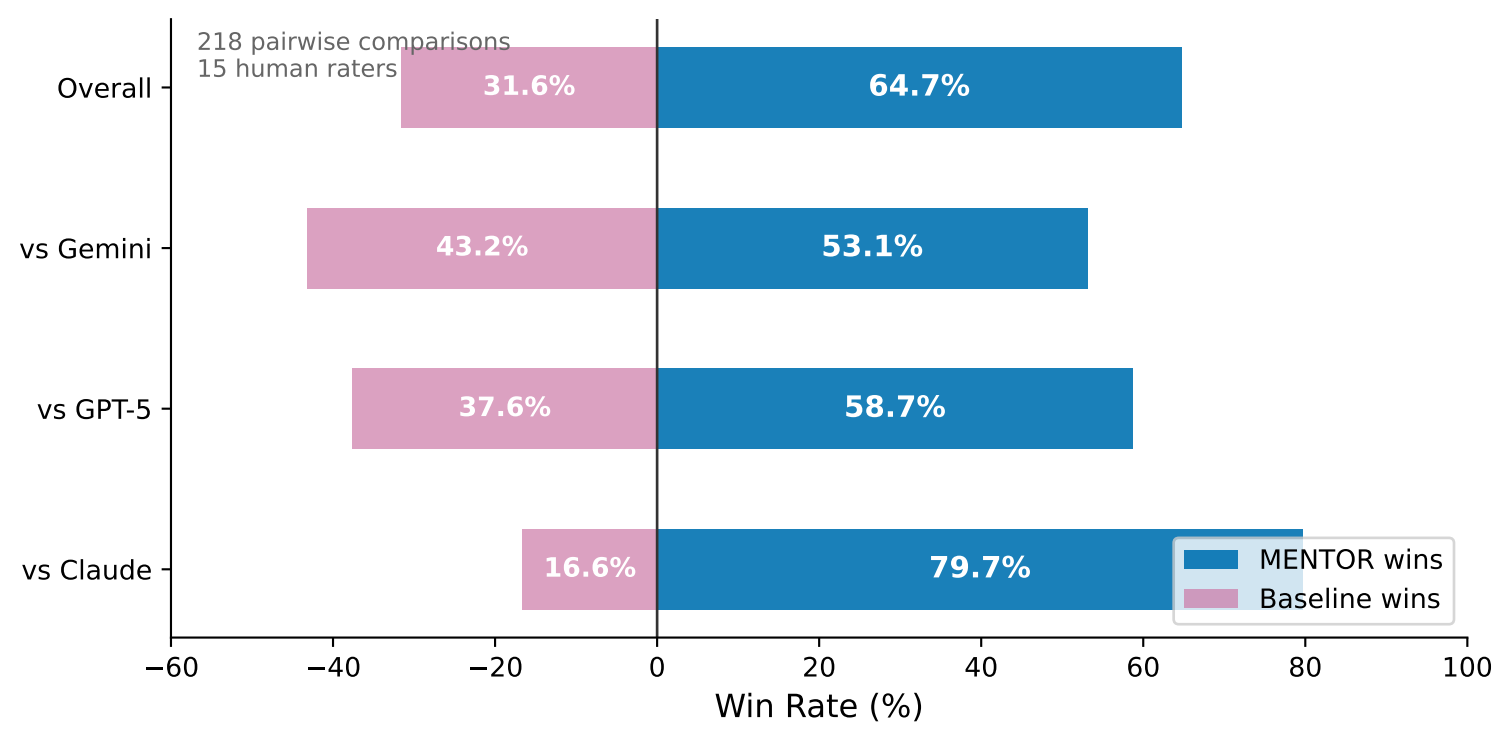
218 pairwise comparisons
15 human raters
Overall
vs Gemini
vs GPT-5
vs Claude
31.6%
64.7%
43.2%
53.1%
37.6%
58.7%
16.6%
79.7%
MENTOR wins
Baseline wins
−60
−40
−20
0
20
40
60
80
100
Win Rate (%)