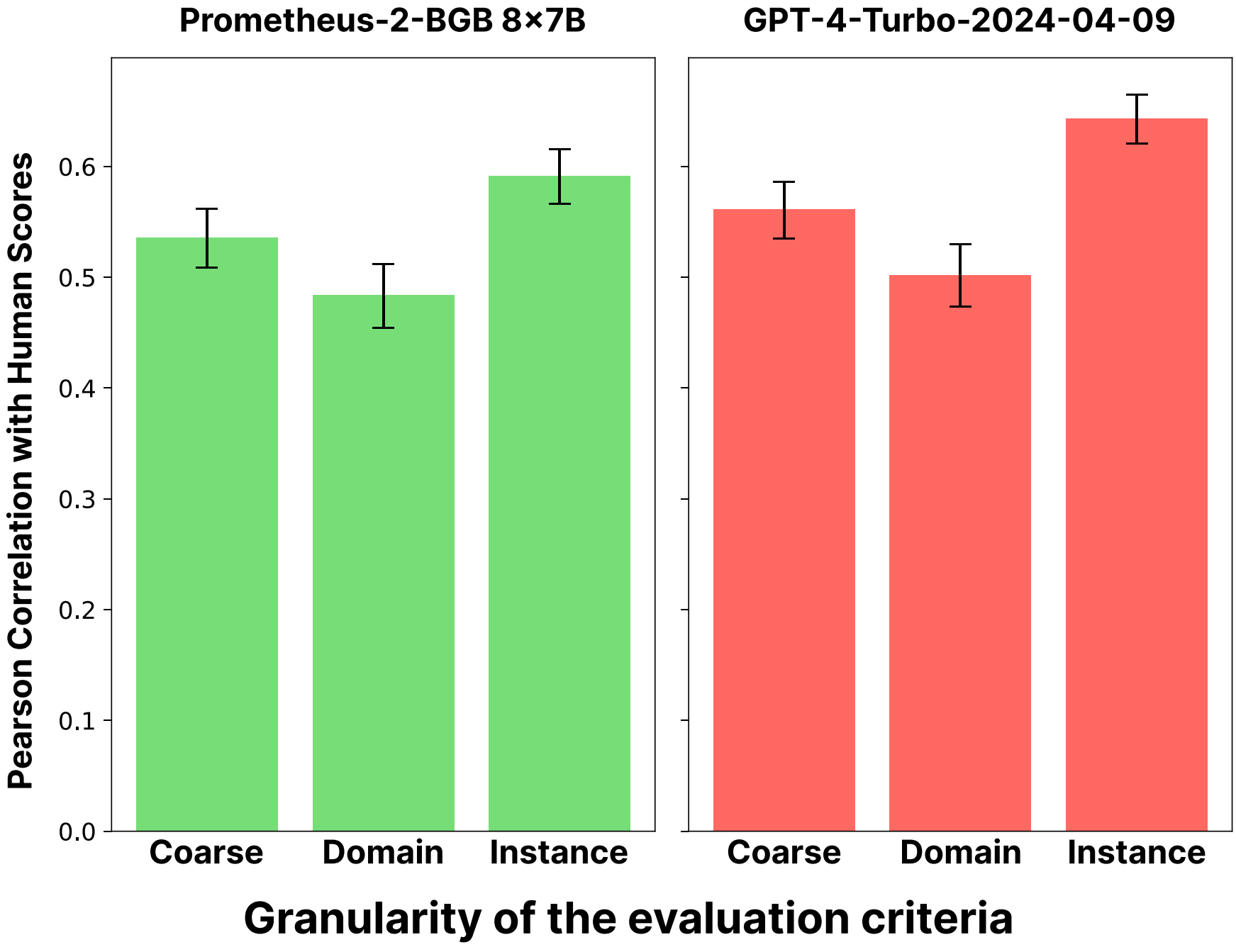

Prometheus-2-BGB 8×7B
GPT-4-Turbo-2024-04-09
Pearson Correlation with Human Scores
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Coarse
Domain
Instance
Coarse
Domain
Instance
Granularity of the evaluation criteria