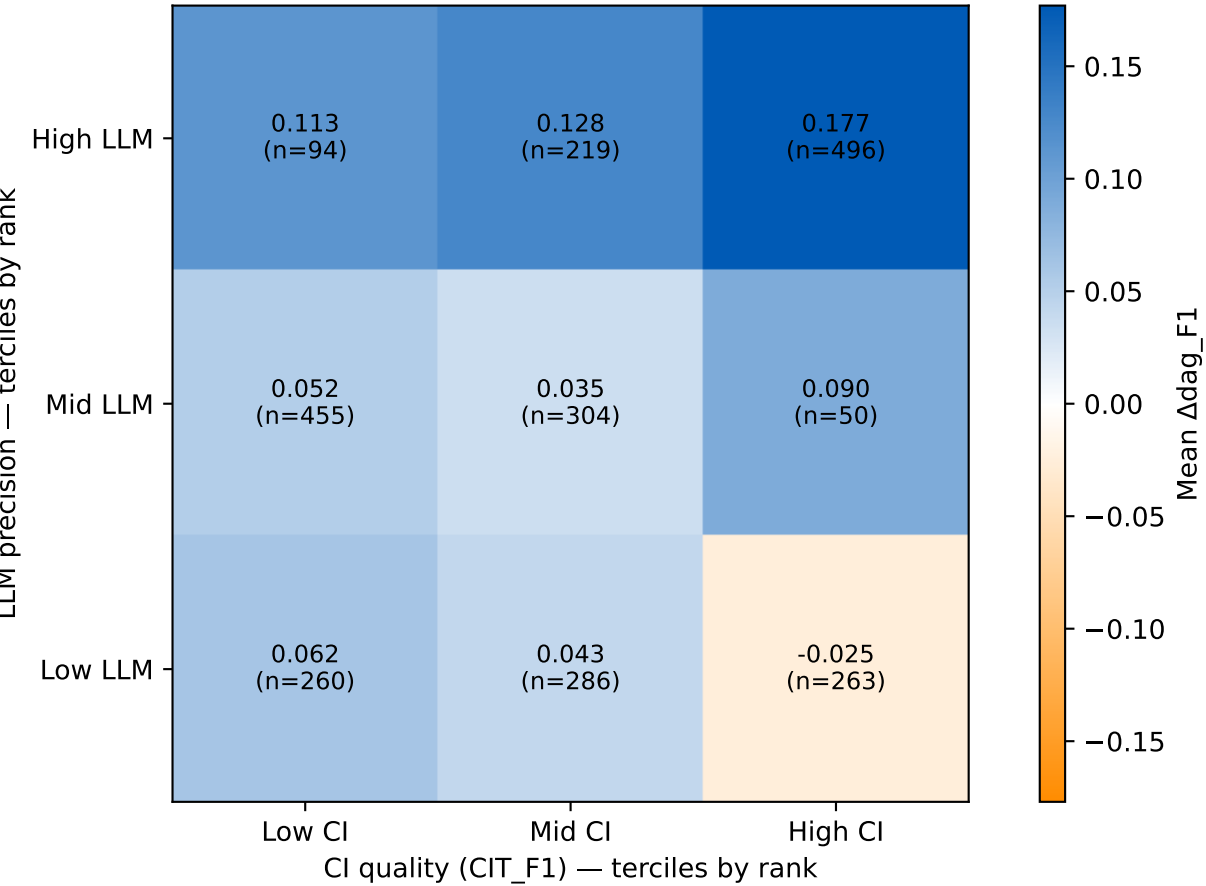
Aggregated 3×3 Heatmaps — Rank-based Low/Mid/High

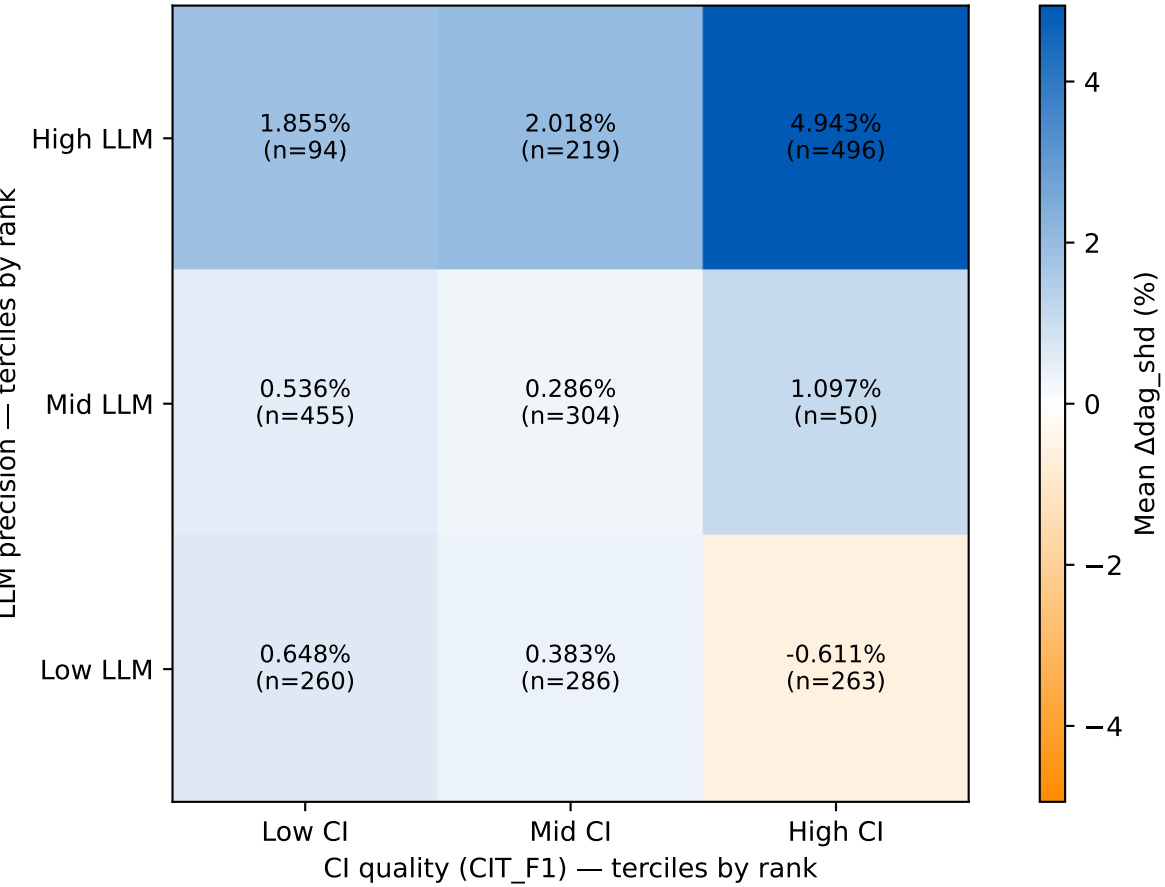
Each axis binned into terciles by rank (ensures ~equal counts per bin)

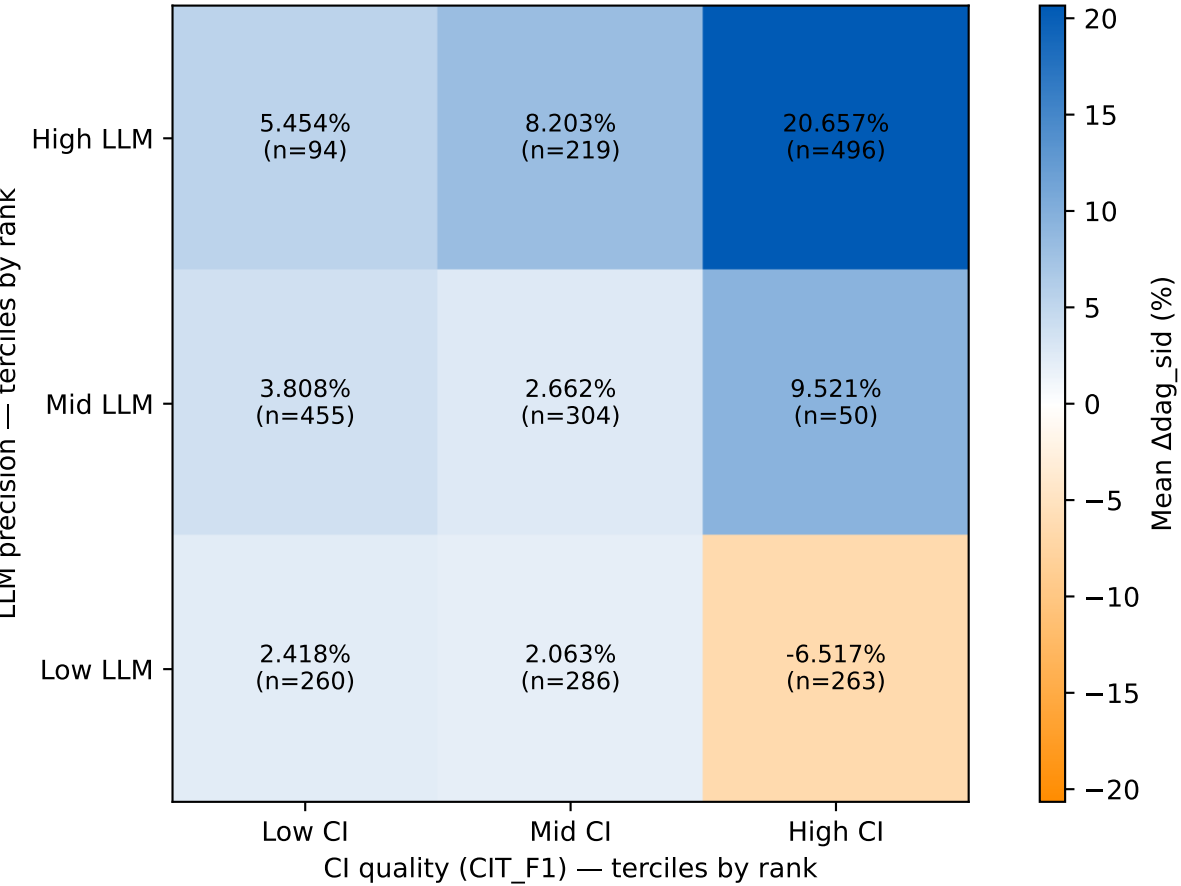
2025-10-02 12:50

DAG Metrics

2025-10-02 12:50

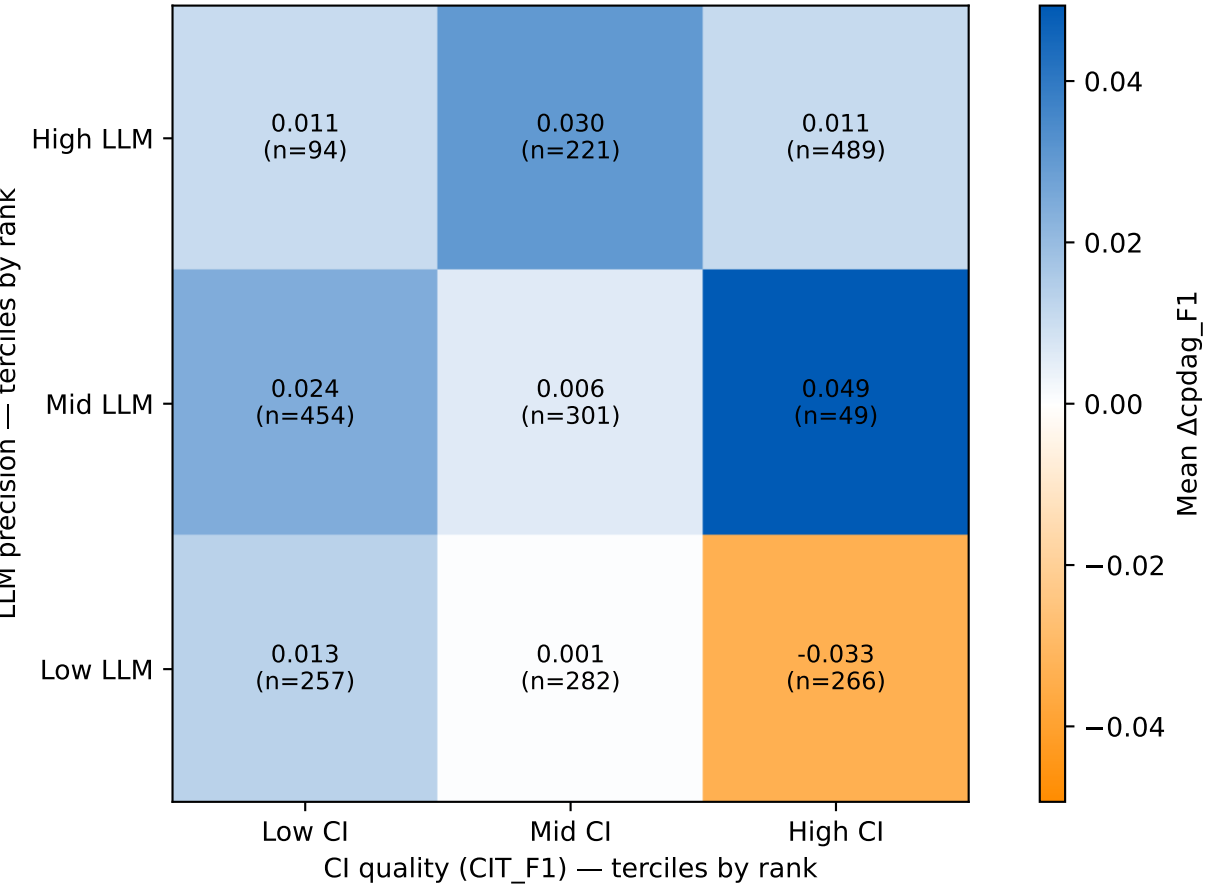

0.113
(n=94)
0.128
(n=219)
0.177
(n=496)
0.052
(n=455)
0.035
(n=304)
0.090
(n=50)
0.062
(n=260)
0.043
(n=286)
-0.025
(n=263)
High LLM
Mid LLM
Low LLM
LLM precision — terciles by rank
Low CI
Mid CI
High CI
CI quality (CIT_F1) — terciles by rank
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
Mean Δdag_F1

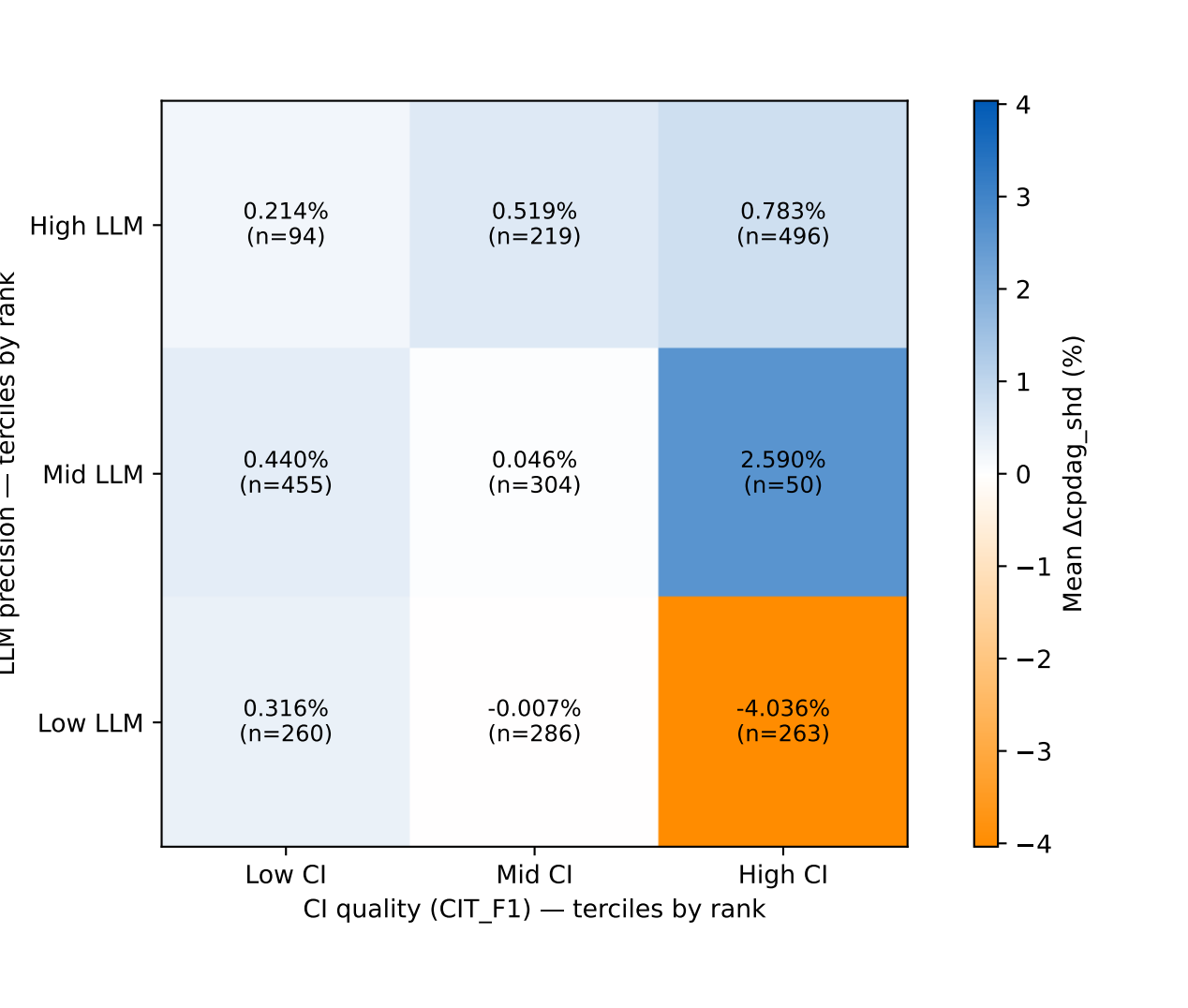

1.855%
(n=94)
2.018%
(n=219)
4.943%
(n=496)
0.536%
(n=455)
0.286%
(n=304)
1.097%
(n=50)
0.648%
(n=260)
0.383%
(n=286)
-0.611%
(n=263)
High LLM
Mid LLM
Low LLM
LLM precision — terciles by rank
Low CI
Mid CI
High CI
CI quality (CIT_F1) — terciles by rank
4
2
0
−2
−4
Mean Δdag_shd (%)

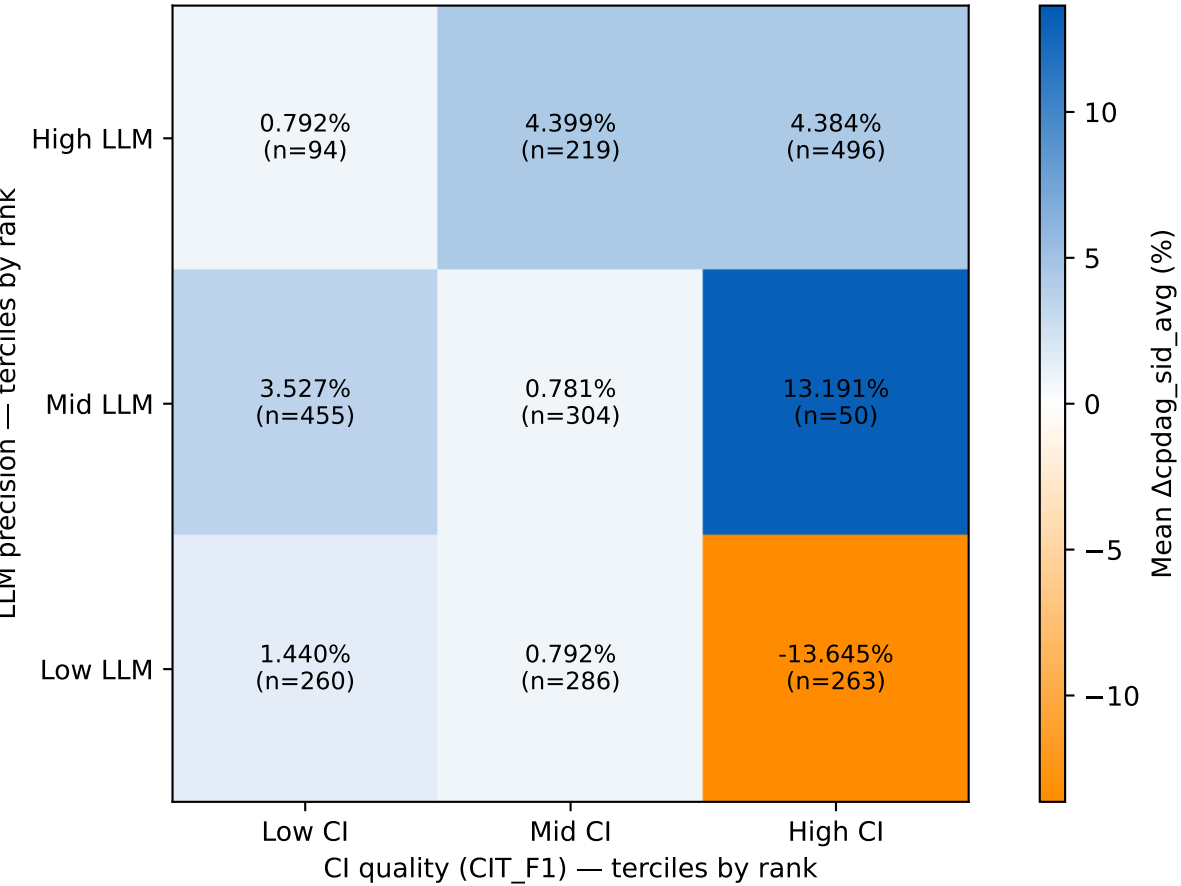

5.454%
(n=94)
8.203%
(n=219)
20.657%
(n=496)
3.808%
(n=455)
2.662%
(n=304)
9.521%
(n=50)
2.418%
(n=260)
2.063%
(n=286)
-6.517%
(n=263)
High LLM
Mid LLM
Low LLM
LLM precision — terciles by rank
Low CI
Mid CI
High CI
CI quality (CIT_F1) — terciles by rank
20
15
10
5
0
−5
−10
−15
−20
Mean Δdag_sid (%)

CPDAG Metrics

2025-10-02 12:50


0.011
(n=94)
0.030
(n=221)
0.011
(n=489)
0.024
(n=454)
0.006
(n=301)
0.049
(n=49)
0.013
(n=257)
0.001
(n=282)
-0.033
(n=266)
High LLM
Mid LLM
Low LLM
LLM precision — terciles by rank
Low CI
Mid CI
High CI
CI quality (CIT_F1) — terciles by rank
0.04
0.02
0.00
−0.02
−0.04
Mean Δcpdag_F1


0.214%
(n=94)
0.519%
(n=219)
0.783%
(n=496)
0.440%
(n=455)
0.046%
(n=304)
2.590%
(n=50)
0.316%
(n=260)
-0.007%
(n=286)
-4.036%
(n=263)
High LLM
Mid LLM
Low LLM
LLM precision — terciles by rank
Low CI
Mid CI
High CI
CI quality (CIT_F1) — terciles by rank
Mean Δcpdag_shd (%)
4
3
2
1
0
−1
−2
−3
−4


0.792%
(n=94)
4.399%
(n=219)
4.384%
(n=496)
3.527%
(n=455)
0.781%
(n=304)
13.191%
(n=50)
1.440%
(n=260)
0.792%
(n=286)
-13.645%
(n=263)
High LLM
Mid LLM
Low LLM
LLM precision — terciles by rank
Low CI
Mid CI
High CI
CI quality (CIT_F1) — terciles by rank
10
5
0
−5
−10
Mean Δcpdag_sid_avg (%)