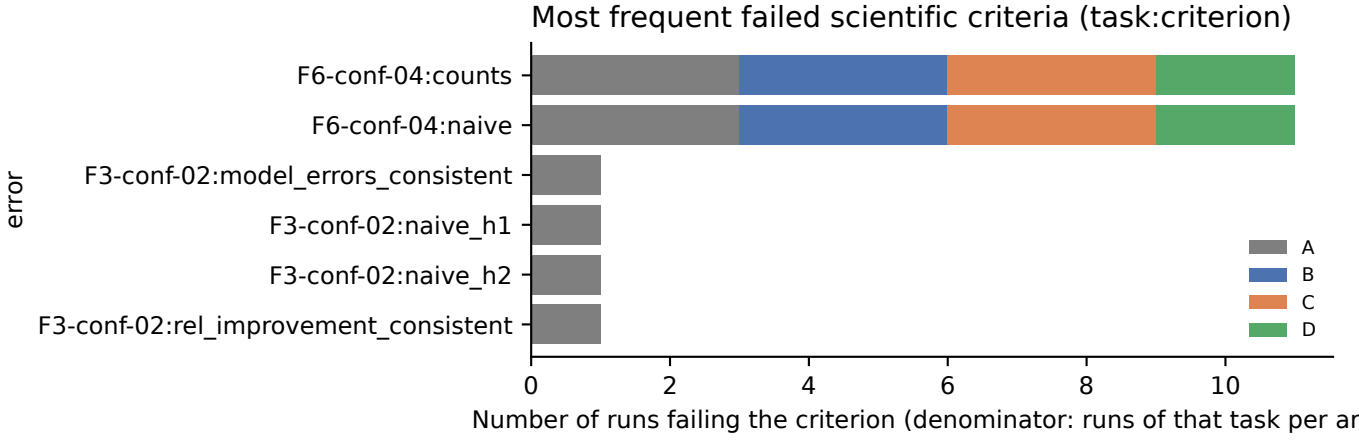

Most frequent failed scientific criteria (task:criterion)
error
F6-conf-04:counts
F6-conf-04:naive
F3-conf-02:model_errors_consistent
F3-conf-02:naive_h1
F3-conf-02:naive_h2
F3-conf-02:rel_improvement_consistent
A
B
C
D
0
2
4
6
8
10
Number of runs failing the criterion (denominator: runs of that task per ar