

Summary Statistics: Benchmark Quality Improvement

Metric	Initial Verdict	Post-Fix Verdict	Change
Total Tasks Evaluated	65	9	-
Benchmark Defects (Grade=1)	29	0	↓ 29
Agent Issues (Grade=0)	36	9	-
Defect Rate	44.6%	0.0%	↓ 44.6%