Performance distribution by dataset across all agents

Combined performance score

0.829   0.824   0.707   0.629   0.567

healthcare   eligibility   financial   nasa-FD001   nasa-FD002