

Efficiency vs. Performance (Balanced F1) of gpt-5-mini on Hard2Verify

