Issue Resolution vs Overthinking Score Reasoning Models Trend o1 high FC $R^2 = 0.808$ p = 0.006 $\beta = -11.476$ Sonnet 40 Non-Reasoning Models Trend 3.5 FC $R^2 = 0.727$ p = 0.031DS-R1 $\beta = -25.802$ 30 GPT-4o-FC o1 high DS-V3 o1 low o1-mini 20 10 DS-R1-32B Sky-T1 GPT-4o-mini-FC QwQ Qwen2.5-32B 0.0 0.5 2.0 2.5 1.0 1.5 3.0 3.5 Overthinking Score (Avg)

Resolved Issues (%)