

Grouped Comparison: Adjusted Balanced Accuracy (Harmonic Mean), Cost & Latency
(Initial Prompt vs Initial Prompt Simple)

