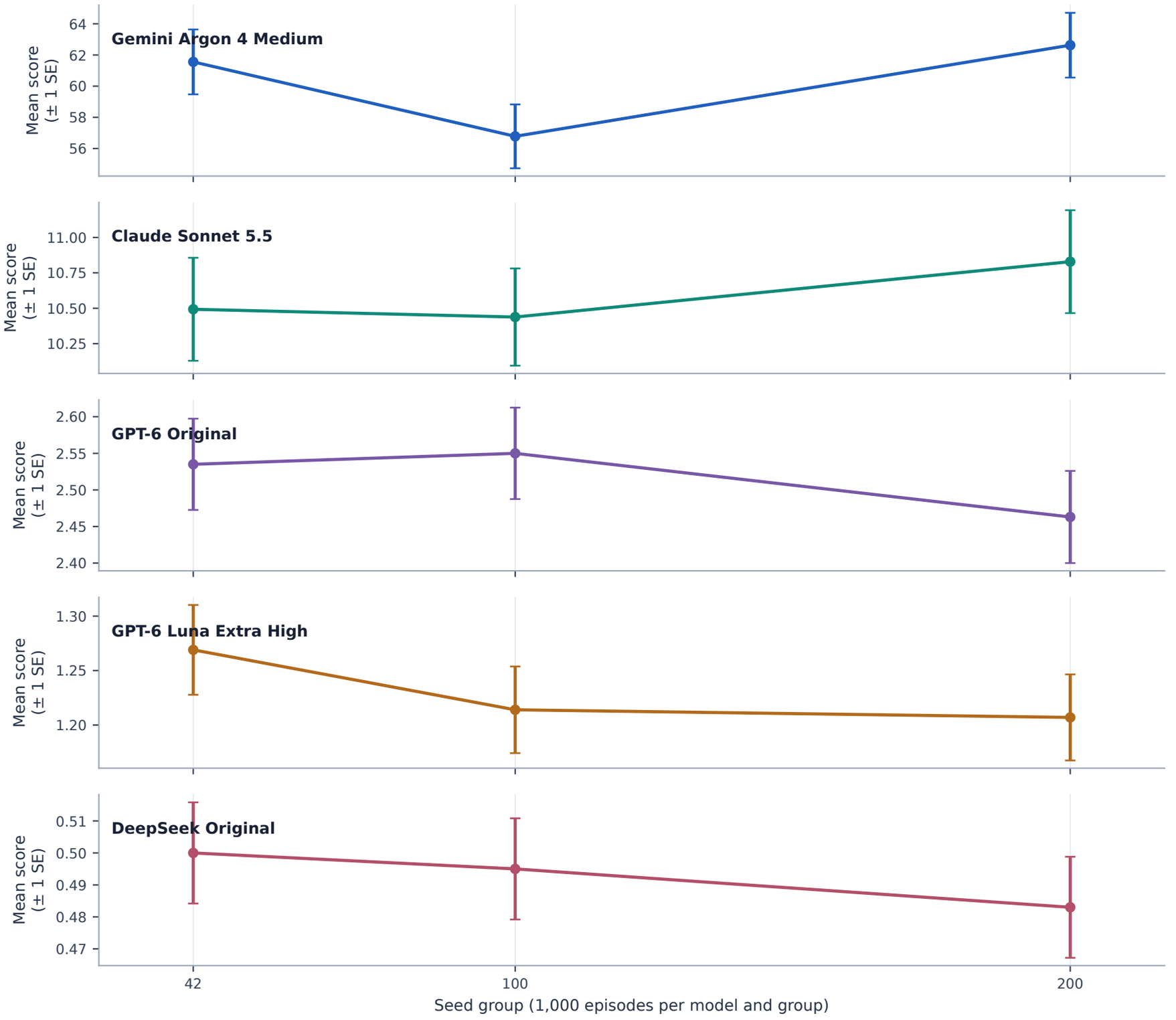
Primary zero-shot performance by seed group



Source: corrected Phase 2 raw episode CSVs. Each point uses 1,000 episodes in the recorded 42, 100, or 200 seed group; this descriptive split does not replace matched-pair analysis.