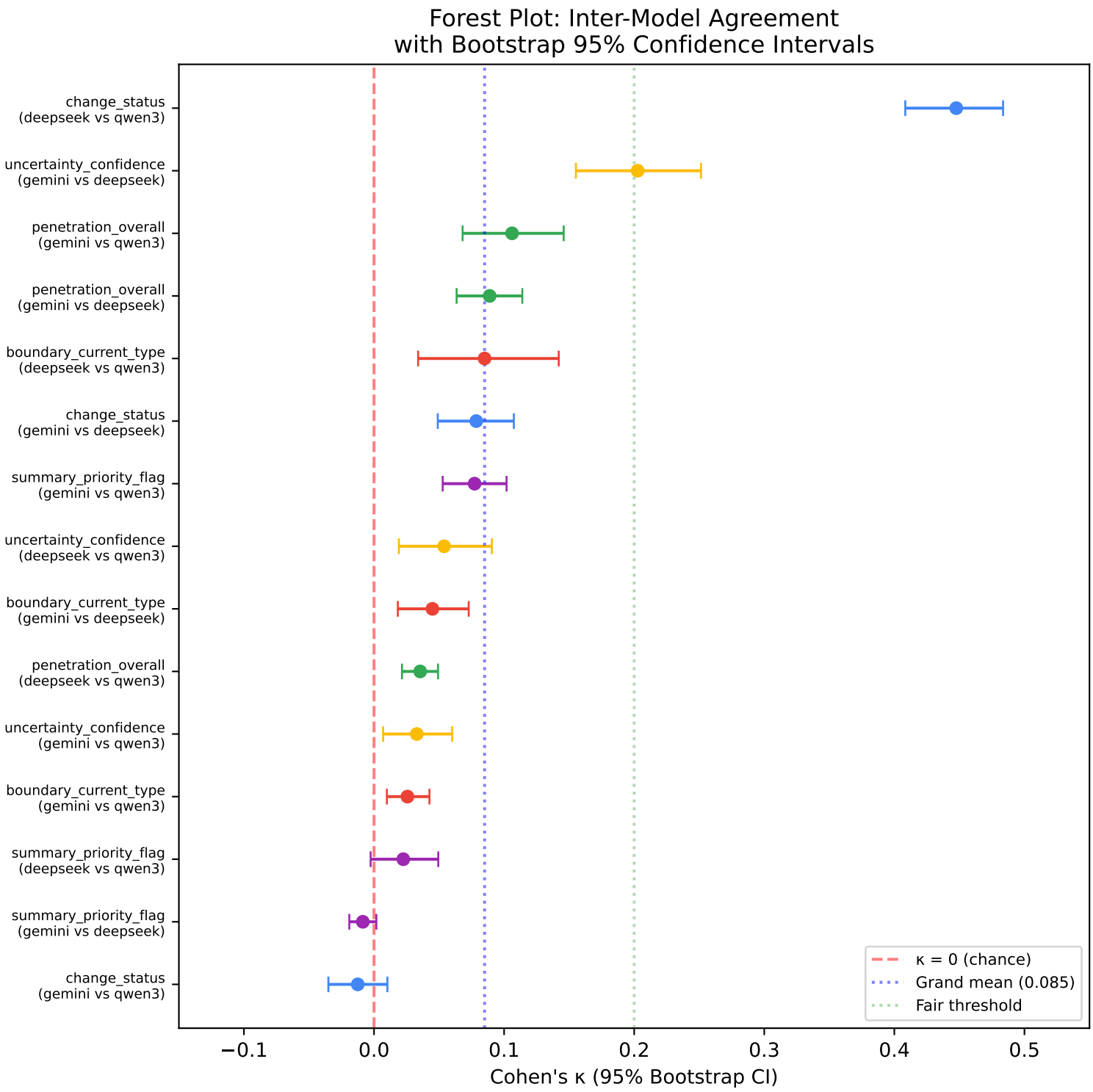
Forest Plot: Inter-Model Agreement
with Bootstrap 95% Confidence Intervals
change_status
(deepseek vs qwen3)
uncertainty_confidence
(gemini vs deepseek)
penetration_overall
(gemini vs qwen3)
penetration_overall
(gemini vs deepseek)
boundary_current_type
(deepseek vs qwen3)
change_status
(gemini vs deepseek)
summary_priority_flag
(gemini vs qwen3)
uncertainty_confidence
(deepseek vs qwen3)
boundary_current_type
(gemini vs deepseek)
penetration_overall
(deepseek vs qwen3)
uncertainty_confidence
(gemini vs qwen3)
boundary_current_type
(gemini vs qwen3)
summary_priority_flag
(deepseek vs qwen3)
summary_priority_flag
(gemini vs deepseek)
change_status
(gemini vs qwen3)
κ = 0 (chance)
Grand mean (0.085)
Fair threshold
−0.1
0.0
0.1
0.2
0.3
0.4
0.5
Cohen's κ (95% Bootstrap CI)