Overall Extraction Accuracy Across Models

Model	Extraction Accuracy
Qwen3-235B-A22B-FP8	0.418
gpt-4.1	0.430
gemini-2.5-flash	0.444
deepseek-v3.1	0.482
kimi-k2	0.483
claude-sonnet-4	0.493

N=4,217 per model\n95% Bootstrap CI\n(10,000 iterations)\nτ* = 0.66

Random (0.5)