Behavioral Consistency (CV): 3-Model Comparison

μ=15.2%

μ=32.2%

μ=47.0%

CV (%)

Claude 4.5
Sonnet

GPT-5
(Snowflake)

Llama-3.1
70B