Moral Robustness Z-Score by Model 2.56 ± 0.42 claude-sonnet-4-5- 0.21 ± 0.12 claude-haiku-4-5 0.04 ± 0.05 gemini-2.5-flash-lite - -0.32 ± 0.02 gpt-4.1 -0.37 ± 0.02 gpt-4o-mini - -0.38 ± 0.02 gpt-4.1-nano- -0.41 ± 0.01 gpt-4.1-mini - -0.66 ± 0.0 grok-4-fast - -0.66 ± 0.0 grok-4 -1.0 - 0.50.0 0.5 1.0 1.5 2.0 2.5 3.0 **Robustness Z-Score**