

# LLM vs. Human Evaluation Scores

Pearson  $r = 0.68$ ,  $p = 0.0003$

