GPT4V Ratings vs. F1 Score 1.0 GPT4V Regression Line 0.8 -Quality of Ratings
0
7
9 0.2 0.0 + 0.0 0.2 0.4 0.6 0.8 1.0 F1 Score