## Pairwise Comparison: verbal\_openai vs verbal (0.5 Zoom) (Points after semicolon, 87 lines compared, 1388/1707 points shown) 0.5 0.0 0.0 0.1 0.2 0.3 0.4 0.5

verbal\_openai KL Divergence

Line 2 (24 points) Line 4 (23 points) Line 5 (12 points) Line 6 (20 points) Line 7 (8 points) Line 9 (24 points) Line 10 (19 points) Line 13 (24 points) Line 14 (1 points) Line 15 (8 points) Line 16 (24 points) Line 17 (23 points) Line 18 (20 points) Line 19 (6 points) Line 21 (23 points) Line 23 (5 points) Line 25 (19 points) Line 27 (19 points) Line 28 (24 points) Line 30 (24 points) Line 32 (1 points) Line 33 (22 points) Line 34 (22 points) Line 35 (21 points) Line 36 (23 points) Line 37 (23 points) Line 38 (5 points) Line 39 (24 points) Line 40 (22 points) Line 41 (22 points) Line 42 (19 points) Line 43 (19 points) Line 44 (21 points) Line 45 (20 points) Line 46 (23 points) Line 47 (20 points) Line 48 (24 points) Line 49 (24 points) Line 50 (23 points) Line 51 (24 points) Line 52 (22 points) Line 54 (23 points) Line 55 (20 points) Line 56 (14 points) Line 57 (17 points) Line 58 (24 points) Line 59 (23 points) Line 60 (20 points) Line 61 (21 points) Line 64 (24 points) Line 65 (24 points) Line 66 (23 points) Line 67 (22 points) Line 69 (22 points) Line 70 (22 points) Line 71 (15 points) Line 72 (22 points) Line 73 (24 points) Line 74 (24 points) Line 75 (20 points) Line 77 (23 points) Line 78 (23 points) Line 79 (16 points) Line 80 (24 points) Line 81 (14 points) Line 83 (8 points) Line 84 (16 points) Line 85 (23 points) Line 86 (22 points) Line 87 (19 points) y = x (equal performance)

Line 1 (22 points)