## Normalized confusion matrix (gpt-4o) action -0.55 0.01 0.00 0.01 0.03 0.01 0.14 0.13 0.01 0.01 0.10 0.01 0.00 aggression -0.09 0.71 0.00 0.00 0.00 0.00 0.08 0.04 0.00 0.01 0.07 0.00 0.00 delivery -0.29 0.01 0.00 0.46 0.02 0.01 0.08 0.04 0.03 0.03 0.00 0.02 0.00 entrance -0.01 0.00 0.00 0.01 0.84 0.05 0.02 0.02 0.02 0.02 0.00 0.02 0.00 exit +0.01 0.00 0.00 0.00 0.14 0.79 0.01 0.04 0.00 0.00 0.00 0.00 0.00 True label interaction -0.11 0.01 0.00 0.00 0.01 0.01 <mark>0.80</mark> 0.03 0.00 0.00 0.03 0.00 0.00 movement -0.08 0.00 0.00 0.00 0.04 0.18 0.06 0.63 0.00 0.00 0.00 0.00 0.00 music +0.02 0.00 0.00 0.02 0.01 0.00 0.00 0.01 0.87 0.02 0.02 0.02 0.01 narration -0.10 0.03 0.00 0.01 0.05 0.03 0.07 0.06 0.01 0.57 0.03 0.04 0.00 setting -0.04 0.00 0.00 0.00 0.01 0.03 0.02 0.08 0.00 0.01 0.02 0.81 0.00 toward +0.02 0.00 0.00 0.01 0.04 0.00 0.04 0.00 0.01 0.00 0.01 0.15 0.71 action adjoestion interaction narration MUSIC Predicted label