Normalized confusion matrix (gpt-4o-mini) action -0.32 0.02 0.01 0.03 0.02 0.01 0.25 0.12 0.00 0.04 0.13 0.02 0.02 aparte -0.00 0.00 0.86 0.00 0.00 0.00 0.00 0.07 0.00 0.07 0.00 0.00 0.00 delivery -0.13 0.04 0.04 0.46 0.02 0.00 0.13 0.03 0.05 0.09 0.00 0.00 0.00 entrance -0.02 0.00 0.01 0.00 0.80 0.03 0.02 0.01 0.02 0.09 0.00 0.01 0.00 exit -0.02 0.01 0.00 0.00 0.20 0.71 0.02 0.02 0.00 0.01 0.01 0.00 0.00 True label interaction -0.06 0.10 0.00 0.02 0.01 0.00 <mark>0.67</mark> 0.06 0.00 0.01 0.08 0.00 0.00 movement -0.08 0.00 0.00 0.00 0.05 0.16 0.11 0.55 0.01 0.02 0.01 0.00 0.02 music -0.01 0.00 0.01 0.08 0.01 0.00 0.01 0.00 0.79 0.06 0.02 0.01 0.01 narration -0.06 0.07 0.01 0.00 0.00 0.01 0.07 0.03 0.05 0.68 0.02 0.02 0.00 object -0.07 0.03 0.00 0.05 0.00 0.00 0.08 0.02 0.00 0.02 0.71 0.01 0.00 setting -0.02 0.01 0.01 0.00 0.01 0.01 0.03 0.06 0.01 0.06 0.02 0.78 0.01 toward -0.01 0.00 0.01 0.01 0.04 0.00 0.08 0.00 0.01 0.01 0.00 0.12 0.71 action adjoestion interaction narration MUSIC Predicted label