Normalized confusion matrix (gpt-4o) action -0.33 0.01 0.00 0.02 0.00 0.01 0.35 0.06 0.00 0.05 0.14 0.02 0.01 aggression -0.08 0.72 0.00 0.00 0.00 0.00 0.08 0.01 0.00 0.01 0.09 0.00 0.00 delivery -0.09 0.01 0.02 0.58 0.01 0.01 0.11 0.02 0.03 0.10 0.00 0.00 0.01 entrance -0.01 0.00 0.00 0.01 0.75 0.06 0.02 0.03 0.02 0.08 0.00 0.02 0.00 exit +0.01 0.01 0.00 0.00 0.13 0.79 0.01 0.04 0.00 0.02 0.00 0.00 0.00 True label interaction -0.04 0.03 0.00 0.00 0.01 0.00 <mark>0.78</mark> 0.04 0.00 0.02 0.07 0.01 0.00 movement -0.02 0.01 0.01 0.00 0.03 0.16 0.08 0.58 0.01 0.05 0.01 0.00 0.04 narration -0.01 0.03 0.00 0.01 0.00 0.00 0.02 0.00 0.00 0.86 0.04 0.04 0.00 setting -0.01 0.00 0.01 0.00 0.01 0.01 0.02 0.09 0.01 0.05 0.02 0.79 0.00 action action movement interaction narration Music Predicted label