Normalized confusion matrix (mistral-large-latest) action -0.35 0.02 0.00 0.02 0.01 0.01 0.14 0.15 0.01 0.04 0.15 0.02 0.07 aggression -0.05 0.72 0.00 0.00 0.00 0.05 0.01 0.00 0.04 0.11 0.01 0.00 delivery -0.17 0.02 0.00 0.47 0.00 0.01 0.09 0.04 0.07 0.10 0.00 0.00 0.01 entrance -0.00 0.00 0.00 0.00 0.86 0.03 0.00 0.03 0.02 0.05 0.00 0.01 0.00 exit +0.00 0.00 0.00 0.00 0.10 0.79 0.02 0.07 0.00 0.01 0.00 0.01 0.00

True label interaction -0.08 0.05 0.00 0.00 0.01 0.00 0.71 0.08 0.01 0.01 0.05 0.00 0.01 movement -0.03 0.01 0.00 0.00 0.02 0.16 0.05 0.66 0.00 0.03 0.02 0.01 0.02 narration -0.02 0.03 0.00 0.00 0.00 0.01 0.02 0.03 0.02 0.79 0.07 0.03 0.00 object -0.04 0.04 0.00 0.00 0.01 0.00 0.09 0.01 0.01 0.02 0.76 0.01 0.00 actionession MUSIC apaite delivery

movement interaction narration Predicted label