## Normalized confusion matrix (mistral-large-latest) action -0.64 0.01 0.00 0.01 0.02 0.01 0.10 0.04 0.01 0.01 0.14 0.01 0.00 aggression -0.08 0.75 0.00 0.00 0.00 0.04 0.01 0.00 0.03 0.09 0.00 0.00 delivery -0.30 0.01 0.00 0.50 0.02 0.01 0.07 0.01 0.02 0.03 0.00 0.02 0.00 entrance -0.01 0.00 0.00 0.00 0.87 0.04 0.02 0.02 0.02 0.02 0.00 0.02 0.00 exit -0.02 0.00 0.00 0.00 0.14 0.76 0.00 0.05 0.00 0.01 0.00 0.00 0.00 True label interaction -0.06 0.04 0.00 0.00 0.01 0.00 <mark>0.80</mark> 0.01 0.00 0.00 0.08 0.00 0.00 movement -0.09 0.00 0.00 0.00 0.03 0.15 0.05 0.61 0.00 0.00 0.03 0.01 0.03 music -0.01 0.00 0.00 0.02 0.01 0.00 0.00 0.01 0.88 0.01 0.03 0.02 0.01 narration -0.05 0.04 0.00 0.02 0.01 0.01 0.09 0.03 0.03 0.59 0.08 0.03 0.01 object -0.03 0.02 0.00 0.00 0.01 0.00 0.04 0.00 0.01 0.00 0.86 0.01 0.00 setting -0.03 0.00 0.01 0.00 0.01 0.01 0.01 0.09 0.00 0.02 0.02 0.81 0.00 action movement interaction narration MUSIC apaite delivery

Predicted label