## Normalized confusion matrix (mistral-small-latest) action -0.39 0.02 0.00 0.05 0.01 0.01 0.10 0.10 0.01 0.05 0.20 0.03 0.03 delivery -0.08 0.02 0.05 0.50 0.00 0.02 0.06 0.05 0.05 0.16 0.00 0.03 0.00 entrance -0.05 0.00 0.00 0.03 0.58 0.11 0.03 0.03 0.00 0.05 0.00 0.13 0.00 exit -0.03 0.01 0.00 0.00 0.08 0.71 0.03 0.08 0.00 0.04 0.00 0.01 0.00 True label interaction -0.06 0.00 0.00 0.00 0.00 <mark>0.87</mark> 0.00 0.00 0.00 0.06 0.00 0.00 movement -0.08 0.00 0.00 0.00 0.06 0.14 0.14 0.53 0.03 0.03 0.00 0.00 0.00 music -0.03 0.01 0.00 0.09 0.02 0.01 0.00 0.00 0.77 0.02 0.03 0.02 0.02 narration -0.14 0.06 0.00 0.08 0.00 0.00 0.03 0.00 0.00 0.47 0.06 0.17 0.00 setting -0.05 0.00 0.00 0.04 0.00 0.05 0.02 0.09 0.02 0.04 0.00 0.68 0.02 action movement interaction MUSIC narration Predicted label