## Normalized confusion matrix (llama-3.1) action -0.60 0.00 0.00 0.02 0.07 0.02 0.02 0.03 0.00 0.02 0.03 0.18 0.00 delivery -0.12 0.00 0.02 0.52 0.06 0.02 0.03 0.00 0.02 0.05 0.00 0.12 0.05 entrance -0.00 0.00 0.00 0.00 0.79 0.05 0.00 0.00 0.00 0.03 0.00 0.13 0.00 exit -0.04 0.00 0.00 0.01 0.16 0.73 0.00 0.04 0.00 0.01 0.00 0.00 0.00 True label movement -0.11 0.00 0.00 0.00 0.00 0.25 0.03 0.53 0.00 0.00 0.03 0.06 0.00 music $\frac{1}{2}$ 0.00 0.00 0.02 0.04 0.05 0.01 0.01 0.00 0.70 0.05 0.01 0.09 0.03 narration -0.14 0.06 0.00 0.00 0.08 0.03 0.08 0.06 0.06 0.22 0.03 0.25 0.00 object -0.63 0.00 0.00 0.06 0.03 0.00 0.05 0.02 0.00 0.00 0.19 0.02 0.00 setting -0.02 0.00 0.02 0.00 0.02 0.04 0.00 0.07 0.00 0.02 0.00 0.82 0.00 toward -0.02 0.00 0.06 0.01 0.23 0.01 0.03 0.01 0.00 0.01 0.00 0.22 0.39 action interaction movement narration aparte delivery MUSIC Predicted label