## Normalized confusion matrix (llama-3.1) action -0.51 0.01 0.00 0.03 0.04 0.01 0.10 0.18 0.01 0.00 0.07 0.03 0.02 aggression -0.31 0.49 0.00 0.00 0.00 0.00 0.07 0.05 0.00 0.03 0.05 0.00 0.00 delivery -0.13 0.01 0.05 0.45 0.03 0.02 0.08 0.12 0.05 0.03 0.00 0.01 0.02 entrance -0.00 0.00 0.01 0.01 0.88 0.05 0.00 0.01 0.00 0.02 0.01 0.02 0.00 exit -0.02 0.00 0.00 0.00 0.14 0.78 0.00 0.05 0.00 0.00 0.00 0.00 0.00 True label movement -0.05 0.00 0.00 0.00 0.06 0.20 0.00 0.65 0.01 0.00 0.02 0.02 0.00 music -0.01 0.00 0.06 0.09 0.02 0.01 0.00 0.03 0.65 0.02 0.02 0.05 0.04 narration -0.07 0.02 0.01 0.03 0.03 0.02 0.07 0.14 0.03 0.43 0.05 0.10 0.00 object -0.19 0.01 0.00 0.01 0.02 0.00 0.05 0.03 0.00 0.00 0.65 0.02 0.00 setting -0.02 0.00 0.01 0.00 0.01 0.03 0.00 0.09 0.00 0.02 0.01 0.83 0.00 toward -0.02 0.00 0.03 0.01 0.22 0.00 0.03 0.05 0.01 0.00 0.00 0.09 0.55 action adjoestion movement interaction narration aparte Music delivery Predicted label