Normalized confusion matrix (llama-3) action -0.45 0.01 0.04 0.05 0.16 0.01 0.02 0.13 0.02 0.00 0.09 0.02 0.01 aggression -0.41 0.45 0.00 0.00 0.01 0.00 0.00 0.07 0.01 0.00 0.04 0.00 0.00 delivery -0.14 0.01 0.06 0.51 0.09 0.00 0.04 0.06 0.05 0.02 0.00 0.02 0.01 entrance -0.00 0.00 0.01 0.01 0.91 0.03 0.00 0.01 0.00 0.00 0.01 0.02 0.01 exit +0.03 0.00 0.00 0.00 0.40 0.53 0.00 0.02 0.00 0.00 0.00 0.00 0.00 True label interaction -0.58 0.05 0.01 0.02 0.06 0.01 0.15 0.07 0.00 0.00 0.06 0.00 0.00 movement -0.05 0.00 0.03 0.00 0.24 0.17 0.03 0.45 0.01 0.00 0.01 0.01 0.00 music -0.01 0.00 0.05 0.25 0.04 0.00 0.00 0.02 0.55 0.01 0.01 0.03 0.03 narration -0.17 0.05 0.02 0.03 0.10 0.01 0.17 0.16 0.07 0.03 0.06 0.12 0.00 object -0.28 0.01 0.00 0.17 0.01 0.00 0.02 0.03 0.00 0.00 0.45 0.02 0.00 setting -0.02 0.00 0.01 0.00 0.09 0.01 0.01 0.07 0.01 0.01 0.01 0.77 0.00

interaction

action

movement Music

Predicted label

narration

toward -0.02 0.00 0.08 0.04 0.38 0.00 0.02 0.00 0.00 0.00 0.00 0.02 0.43