Normalized confusion matrix (mistral-small-latest) action -0.26 0.01 0.01 0.06 0.02 0.02 0.17 0.12 0.01 0.05 0.17 0.02 0.09 aggression -0.05 0.47 0.00 0.00 0.01 0.00 0.15 0.05 0.04 0.03 0.19 0.01 0.00 delivery -0.11 0.02 0.04 0.42 0.02 0.03 0.04 0.04 0.07 0.12 0.02 0.02 0.06 entrance -0.03 0.01 0.00 0.02 0.69 0.03 0.01 0.02 0.03 0.08 0.04 0.05 0.00 exit $+0.02 \ 0.01 \ 0.00 \ 0.00 \ 0.10 \ 0.69 \ 0.01 \ 0.07 \ 0.01 \ 0.02 \ 0.02 \ 0.02 \ 0.02$ True label interaction -0.02 0.03 0.01 0.03 0.01 0.02 <mark>0.75</mark> 0.03 0.00 0.01 0.08 0.00 0.02 movement -0.08 0.00 0.02 0.01 0.05 0.11 0.13 0.50 0.00 0.03 0.03 0.00 0.03 narration -0.07 0.05 0.00 0.02 0.03 0.03 0.05 0.04 0.03 0.60 0.03 0.05 0.01 object -0.04 0.02 0.00 0.00 0.00 0.01 0.13 0.02 0.01 0.04 0.64 0.03 0.02 setting -0.03 0.01 0.00 0.01 0.01 0.01 0.02 0.11 0.02 0.03 0.03 0.71 0.03 action interaction narration MUSIC apaite delivery Predicted label