Normalized confusion matrix (mistral-large-latest) action -0.44 0.01 0.00 0.01 0.01 0.01 0.24 0.10 0.01 0.01 0.14 0.01 0.00 aggression -0.03 0.79 0.00 0.00 0.00 0.04 0.04 0.04 0.00 0.03 0.08 0.00 0.00 delivery -0.17 0.02 0.00 0.50 0.02 0.01 0.11 0.06 0.04 0.06 0.00 0.00 0.00 entrance -0.00 0.00 0.00 0.00 0.88 0.04 0.01 0.02 0.01 0.02 0.00 0.02 0.00 exit +0.00 0.00 0.00 0.00 0.14 0.80 0.01 0.05 0.00 0.00 0.00 0.00 0.00 True label interaction -0.04 0.10 0.00 0.00 0.01 0.01 <mark>0.75</mark> 0.02 0.00 0.00 0.07 0.00 0.00 movement -0.04 0.00 0.00 0.00 0.02 0.19 0.07 0.65 0.00 0.00 0.03 0.01 0.00 music -0.00 0.00 0.00 0.01 0.02 0.01 0.00 0.01 0.88 0.01 0.03 0.02 0.01 narration -0.04 0.07 0.00 0.01 0.02 0.03 0.11 0.03 0.05 0.54 0.07 0.04 0.00 object -0.02 0.02 0.00 0.00 0.01 0.00 0.05 0.01 0.01 0.00 0.86 0.01 0.00 setting -0.02 0.00 0.00 0.00 0.02 0.03 0.01 0.11 0.00 0.00 0.01 0.82 0.00 toward $+0.00\,0.00\,0.00\,0.02\,0.00\,0.01\,0.04\,0.01\,0.02\,0.00\,0.01\,0.05\,0.84$ actionession movement interaction narration MUSIC apaite delivery Predicted label