## Normalized confusion matrix (mistral-small-latest) action -0.43 0.06 0.01 0.05 0.01 0.00 0.13 0.04 0.00 0.03 0.16 0.05 0.02 delivery -0.02 0.00 0.03 0.56 0.00 0.06 0.02 0.11 0.02 0.14 0.00 0.00 0.05 entrance -0.00 0.00 0.00 0.00 0.74 0.00 0.00 0.08 0.00 0.00 0.00 0.05 0.13 exit -0.00 0.01 0.00 0.00 0.22 0.34 0.01 0.37 0.00 0.00 0.00 0.04 0.00 True label interaction -0.03 0.13 0.00 0.00 0.00 0.00 <mark>0.84</mark> 0.00 0.00 0.00 0.00 0.00 0.00 movement -0.06 0.00 0.00 0.00 0.03 0.25 0.19 0.39 0.03 0.00 0.00 0.03 0.03 music $\frac{1}{2}$ 0.00 0.00 0.00 0.02 0.01 0.00 0.02 0.00 0.81 0.07 0.03 0.02 0.02 narration -0.06 0.06 0.00 0.03 0.03 0.00 0.25 0.03 0.00 0.36 0.00 0.08 0.11 object -0.18 0.03 0.02 0.00 0.02 0.02 0.37 0.02 0.03 0.02 0.27 0.02 0.02 setting -0.05 0.00 0.00 0.00 0.00 0.05 0.00 0.07 0.00 0.05 0.00 0.67 0.11 action deliverytance movement interaction narration MUSIC apaite Predicted label