## Normalized confusion matrix (gpt-4o) action -0.38 0.01 0.00 0.01 0.03 0.01 0.26 0.04 0.01 0.03 0.18 0.03 0.01 delivery -0.14 0.02 0.00 0.53 0.03 0.03 0.09 0.00 0.06 0.09 0.00 0.00 0.00 exit +0.03 0.00 0.00 0.00 0.11 0.82 0.01 0.03 0.00 0.00 0.00 0.00 0.00 True label interaction -0.00 0.13 0.00 0.03 0.00 0.03 <mark>0.81</mark> 0.00 0.00 0.00 0.00 0.00 0.00 movement -0.03 0.00 0.00 0.00 0.03 0.22 0.11 0.56 0.00 0.00 0.03 0.00 0.03 music -0.01 0.01 0.00 0.01 0.02 0.00 0.00 0.01 0.88 0.01 0.03 0.02 0.01 narration -0.06 0.06 0.00 0.00 0.00 0.03 0.03 0.06 0.00 0.58 0.08 0.11 0.00 object -0.00 0.00 0.00 0.00 0.00 0.00 0.05 0.00 0.00 0.00 0.94 0.02 0.00 setting -0.02 0.00 0.00 0.02 0.00 0.04 0.00 0.12 0.00 0.02 0.00 0.79 0.00 toward -0.00 0.01 0.01 0.01 0.07 0.00 0.04 0.01 0.00 0.00 0.01 0.04 0.81 action adjoestion movement interaction narration MUSIC apaite Predicted label