Normalized confusion matrix (gpt-4o) action -0.45 0.01 0.01 0.04 0.00 0.00 0.24 0.05 0.00 0.04 0.13 0.01 0.01 delivery -0.20 0.01 0.03 0.53 0.01 0.01 0.08 0.01 0.03 0.10 0.00 0.00 0.00 entrance -0.05 0.00 0.00 0.01 0.72 0.05 0.02 0.02 0.02 0.08 0.00 0.03 0.01 exit +0.03 0.00 0.00 0.00 0.10 0.71 0.01 0.11 0.00 0.02 0.00 0.00 0.00 True label interaction -0.10 0.02 0.00 0.01 0.01 0.00 <mark>0.79</mark> 0.02 0.00 0.01 0.03 0.01 0.00 movement -0.06 0.00 0.00 0.02 0.02 0.08 0.09 0.64 0.00 0.05 0.02 0.01 0.02 narration -0.06 0.02 0.00 0.01 0.00 0.01 0.04 0.01 0.01 0.79 0.03 0.03 0.00 setting -0.04 0.01 0.01 0.00 0.00 0.01 0.01 0.08 0.00 0.05 0.01 0.79 0.01 action adjoestion movement interaction MUSIC narration Predicted label