Normalized confusion matrix (gpt-4o) action -0.71 0.03 0.00 0.01 0.01 0.01 0.13 0.01 0.00 0.05 0.03 0.01 0.00 delivery -0.12 0.00 0.00 0.67 0.00 0.02 0.03 0.00 0.03 0.11 0.00 0.00 0.02

exit +0.00 0.00 0.00 0.00 0.12 0.75 0.03 0.10 0.00 0.00 0.00 0.00 0.00 True label interaction -0.03 0.00 0.00 0.00 0.00 <mark>0.97</mark> 0.00 0.00 0.00 0.00 0.00 movement -0.11 0.00 0.00 0.00 0.00 0.14 0.08 0.64 0.00 0.00 0.03 0.00 0.00 narration -0.25 0.06 0.00 0.00 0.00 0.00 0.00 0.03 0.00 0.50 0.06 0.11 0.00 object -0.05 0.02 0.00 0.00 0.00 0.00 0.16 0.00 0.00 0.02 0.76 0.00 0.00 setting -0.04 0.00 0.02 0.00 0.00 0.00 0.02 0.14 0.00 0.07 0.02 0.70 0.00 action action movement interaction narration Music

Predicted label