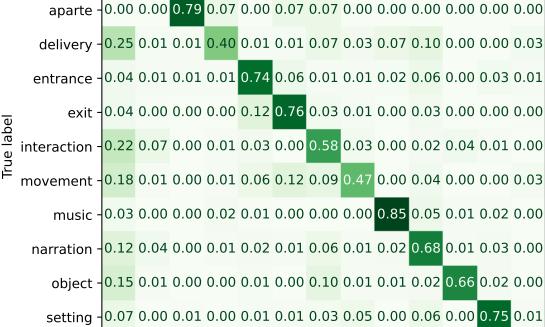
Normalized confusion matrix (gpt-4o) action - 0.42 0.01 0.00 0.03 0.04 0.02 0.15 0.09 0.00 0.06 0.12 0.01 0.04 aggression - 0.27 0.63 0.00 0.00 0.00 0.00 0.04 0.01 0.00 0.01 0.04 0.00 0.00 aparte - 0.00 0.00 0.79 0.07 0.00 0.07 0.07 0.00 0.00 0.00 0.00 0.00 0.00 delivery - 0.25 0.01 0.01 0.40 0.01 0.01 0.07 0.03 0.07 0.10 0.00 0.00 0.03



music - 0.03 0.00 0.00 0.02 0.01 0.00 0.00 0.00 0.05 0.05 0.01 0.02 0.00 narration - 0.12 0.04 0.00 0.01 0.02 0.01 0.06 0.01 0.02 0.68 0.01 0.03 0.00 object - 0.15 0.01 0.00 0.00 0.01 0.00 0.10 0.01 0.01 0.02 0.66 0.02 0.00 setting - 0.07 0.00 0.01 0.00 0.01 0.01 0.03 0.05 0.00 0.06 0.00 0.75 0.01 toward - 0.06 0.00 0.00 0.01 0.07 0.00 0.02 0.01 0.00 0.00 0.00 0.04 0.78 oction parties and selliver transce exist action predicted label