

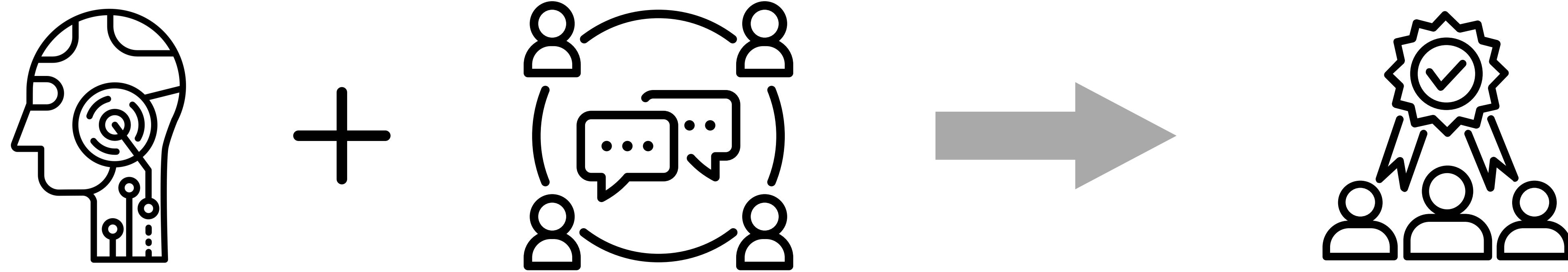
MultICAT: Multimodal Communication Annotations for Teams

<https://multicat.lab.pyarelal.xyz>

*Adarsh Pyarelal, John Culnan, Ayesha Qamar, Meghavarshini Krishnaswamy,
Yuwei Wang, Cheonkam Jeong, Chen Chen, Md Messal Monem Miah,
Shahriar Hormozi, Jonathan Tong, Ruihong Huang*

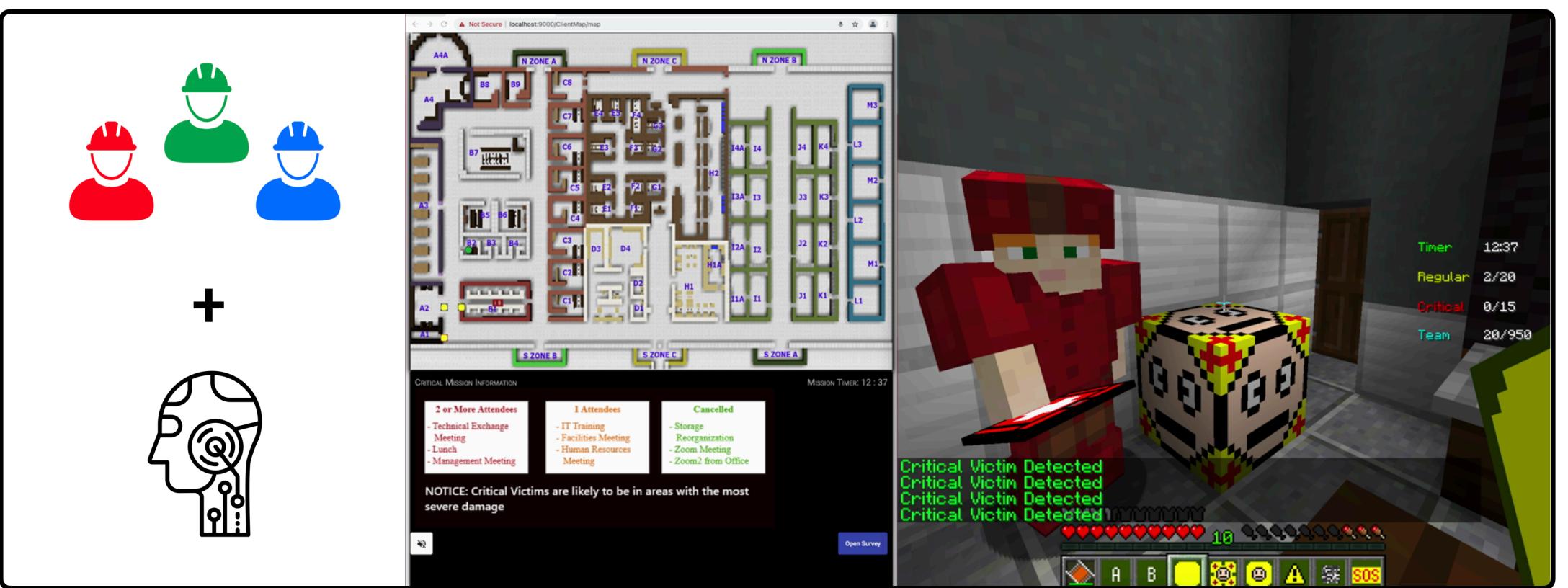


Long-term goal

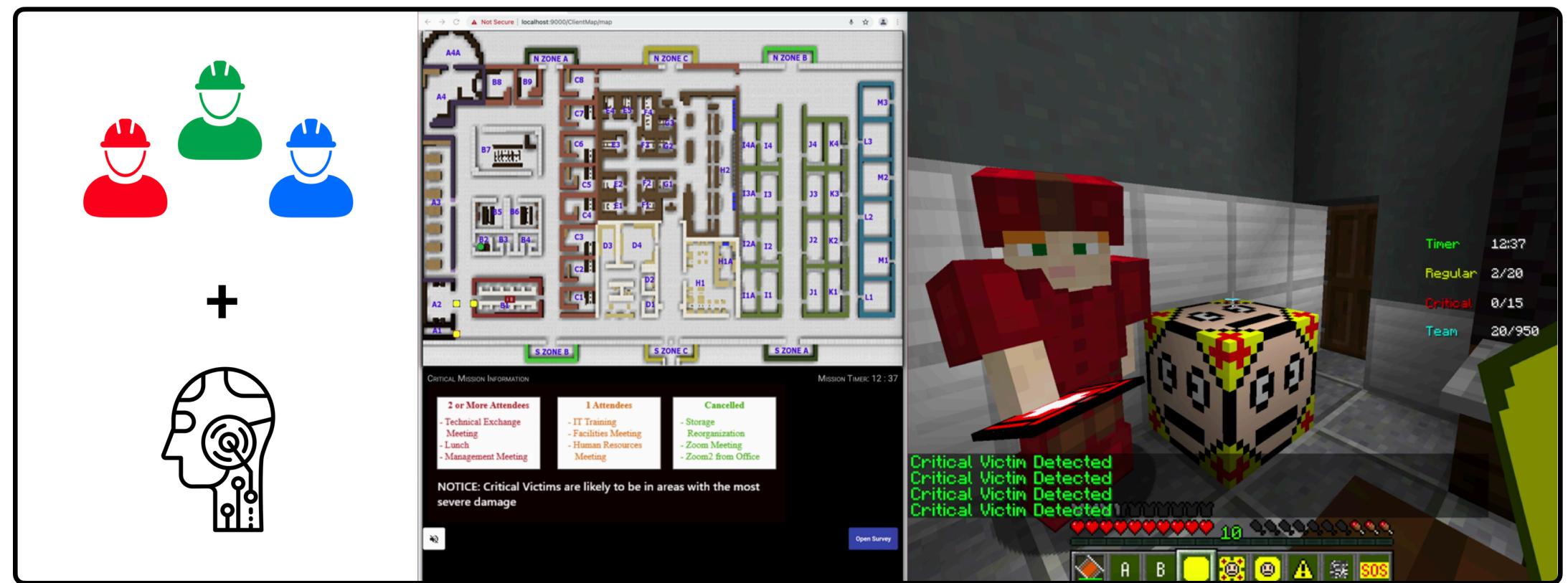


Can we build AI teammates that intervene on team communication to improve team performance?

ASIST Study 3 Dataset



ASIST Study 3 Dataset



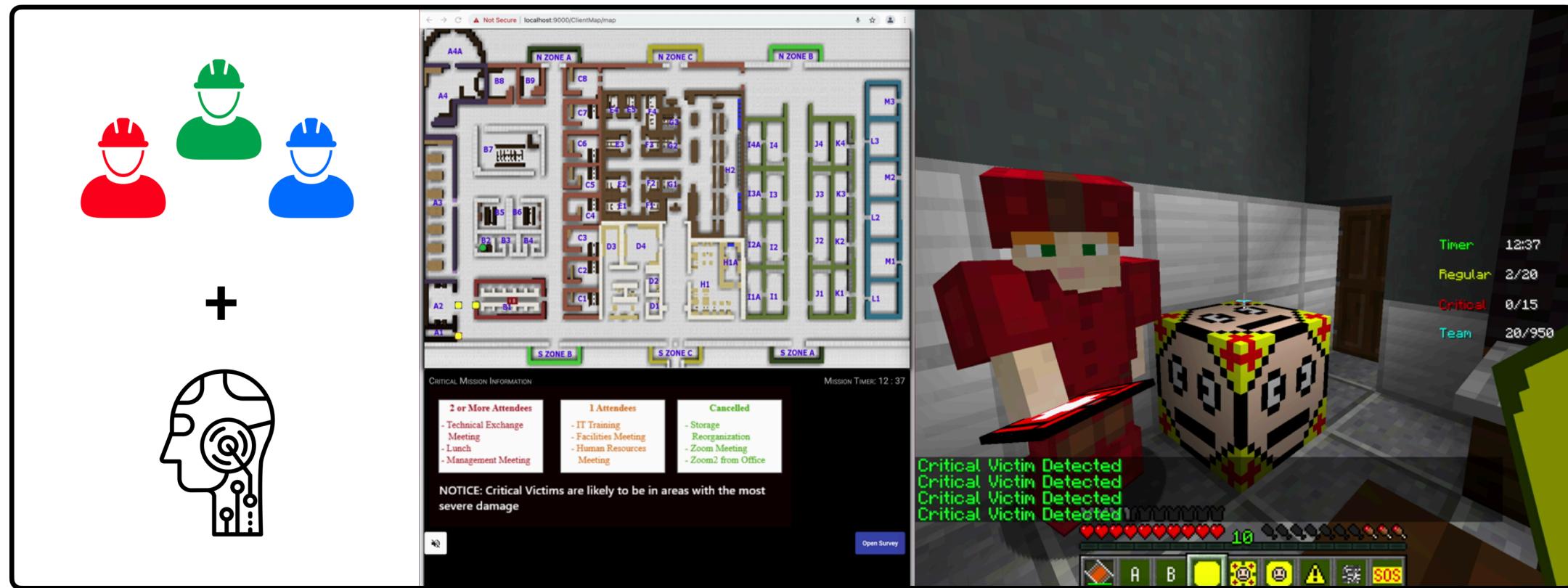
Original UUID

06e5da76-fff9-4d4f-9956-b4b4e7aeeae5

Start time	Speaker	Trial	ASR transcription	End time
16:16	E000651	T000604	C I was gonna go over there	16:18



ASIST Study 3 Dataset



Original UUID

06e5da76-fff9-4d4f-9956-b4b4e7aeeae5

Start time	Speaker	Trial	ASR transcription	End time
16:16	E000651	T000604	C I was gonna go over there	16:18

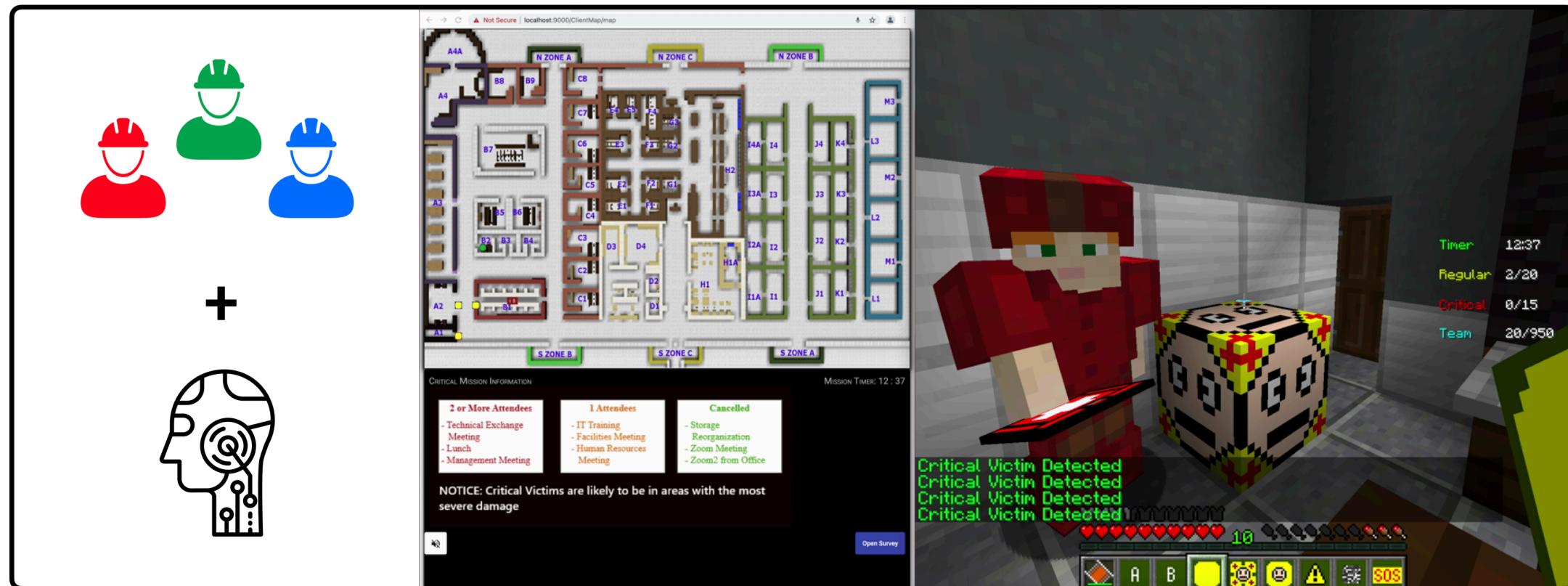


Addressee	Corrected transcription	Emotion	Sentiment
E000649	C8 I was gonna go over there	neutral	neutral

Dialog act	Adjacency pair label	CLC label	CLC checkback score
s	57b	51b.52a	2

MultiCAT Annotations

ASIST Study 3 Dataset



Original UUID

06e5da76-fff9-4d4f-9956-b4b4e7aeeae5

Start time	Speaker	Trial	ASR transcription	End time
16:16	E000651	T000604	C I was gonna go over there	16:18



↓
Addressee
E000649

↓
Corrected transcription
C8 I was gonna go over there

↓
Emotion
neutral

↓
Sentiment
neutral

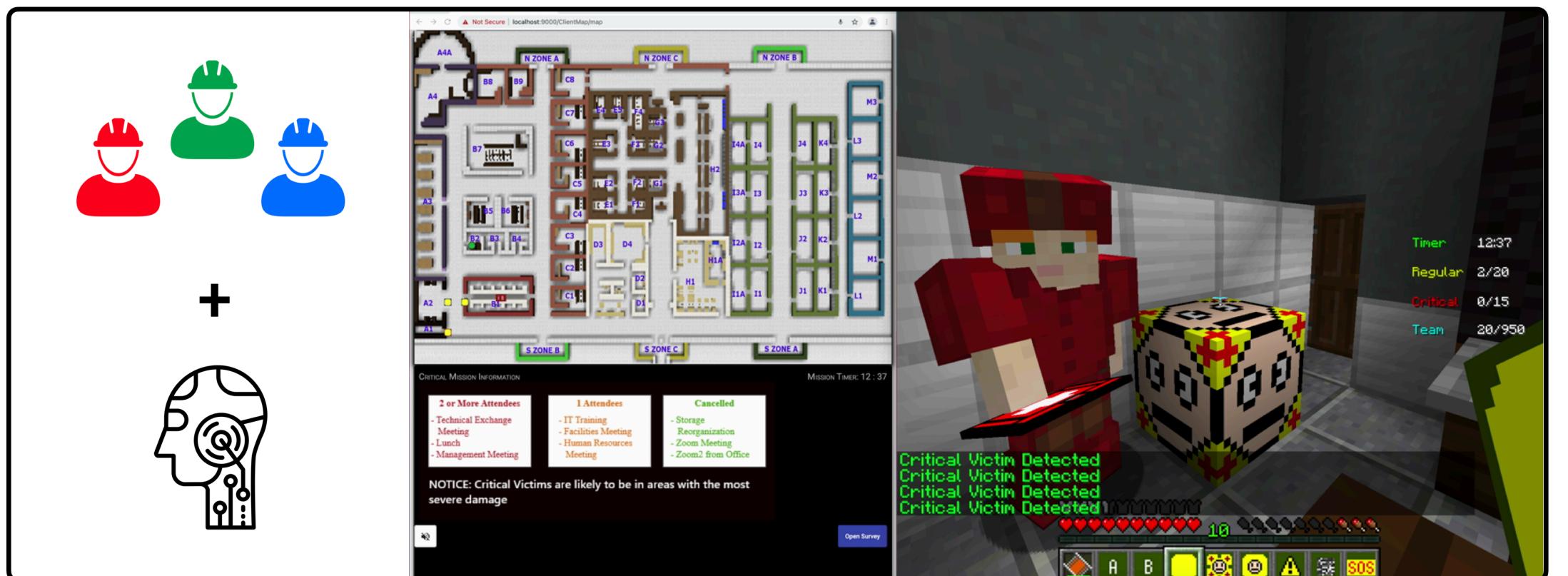
↓
Dialog act
s

↓
Adjacency pair label
57b

↓
CLC label
51b.52a

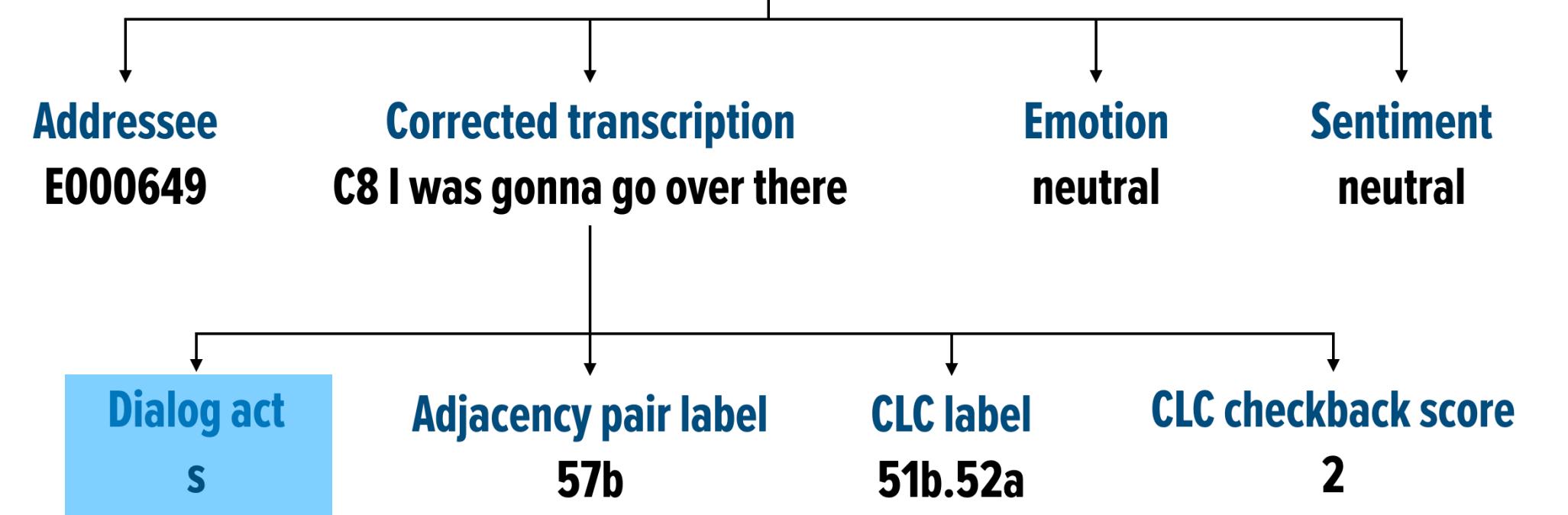
↓
CLC checkback score
2

MultiCAT Annotations



Original UUID
06e5da76-fff9-4d4f-9956-b4b4e7aeaae5

Start time	Speaker	Trial	ASR transcription	End time
16:16	E000651	T000604	C I was gonna go over there	16:18

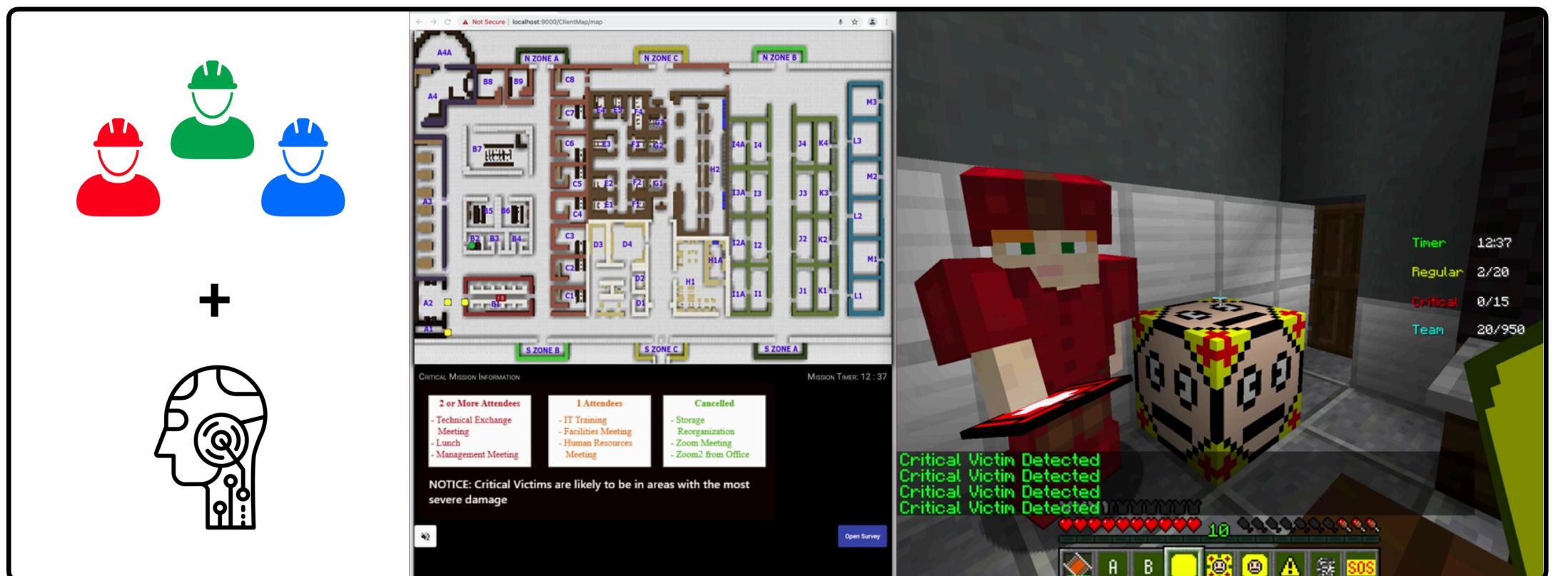


MultiCAT Annotations

Dialog Acts

- A dialog act (DA) is the communicative function underlying a speaker's utterance (Bunt et al., 2020)

- Examples:
 - *Statement*
 - *Filler*
 - *Backchannel*
 - *Disruption*
 - *Question*



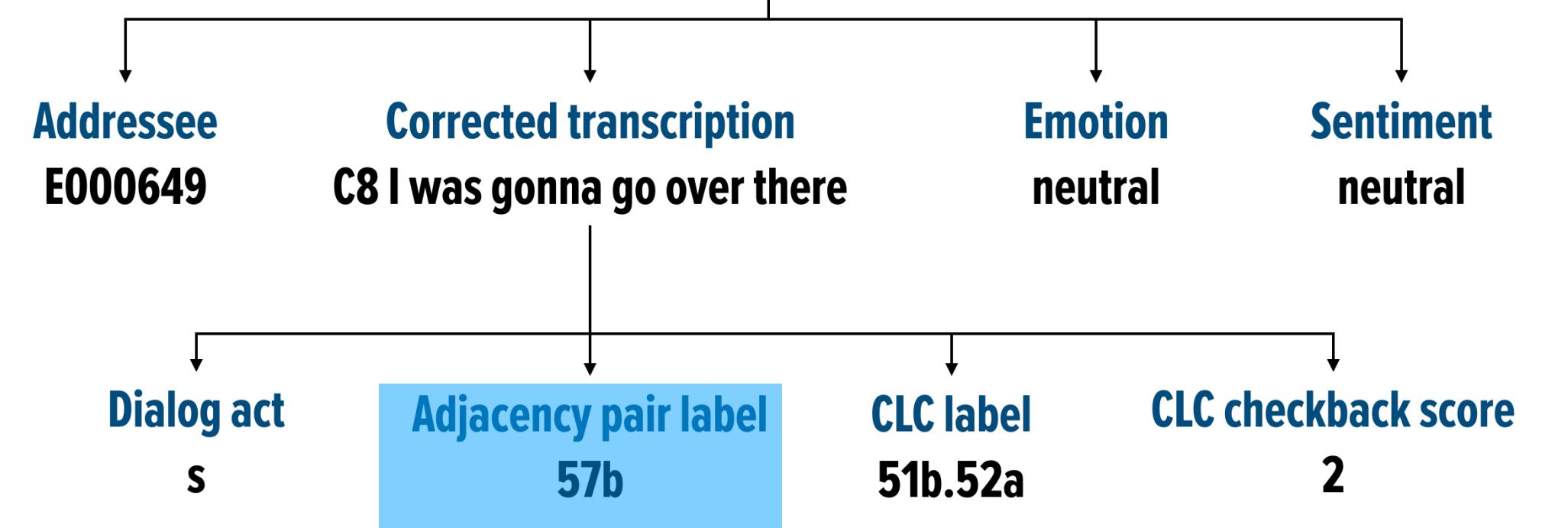
Original UUID
06e5da76-fff9-4d4f-9956-b4b4e7aeaae5

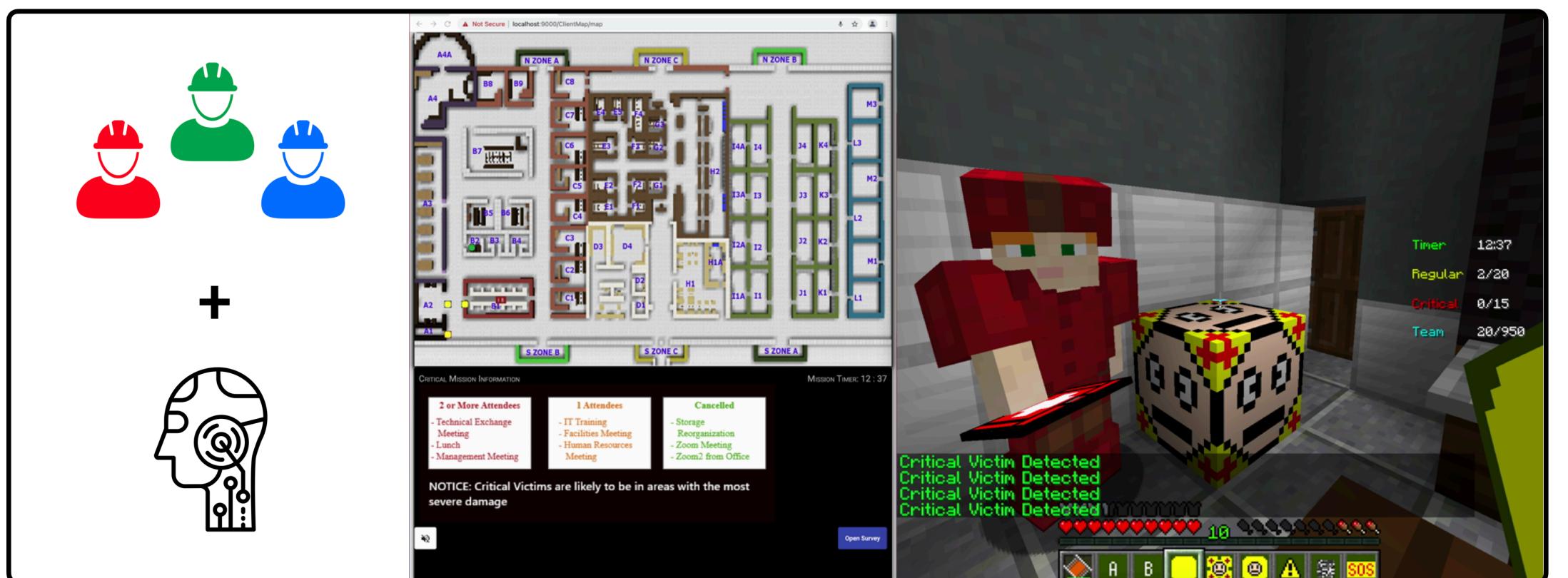


Adjacency Pairs

- An adjacency pair (AP) for a sequence of utterances is defined such that it contains two parts, each containing one or more utterances and uttered by different speakers (Levinson, 1983).

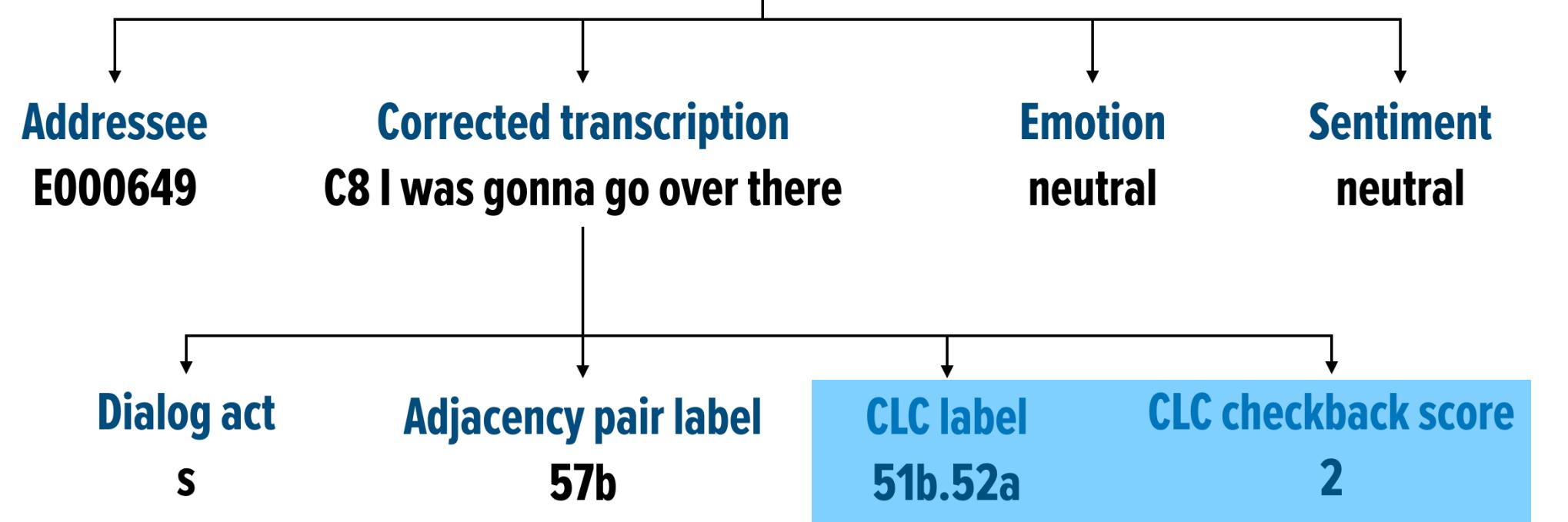
- APs capture paired utterances such as question-answer, greeting-greeting, etc.





Original UUID
06e5da76-fff9-4d4f-9956-b4b4e7aeaae5

Start time	Speaker	Trial	ASR transcription	End time
16:16	E000651	T000604	C I was gonna go over there	16:18

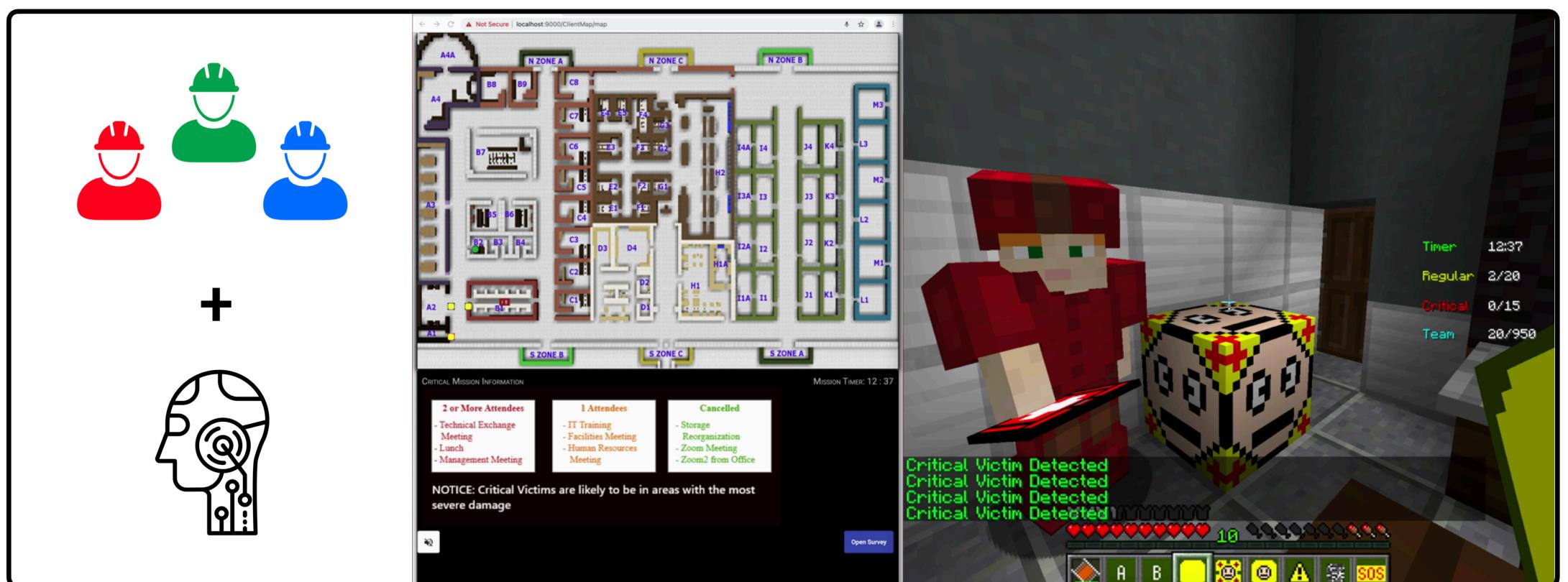


MultiCAT Annotations

Closed-Loop Communication

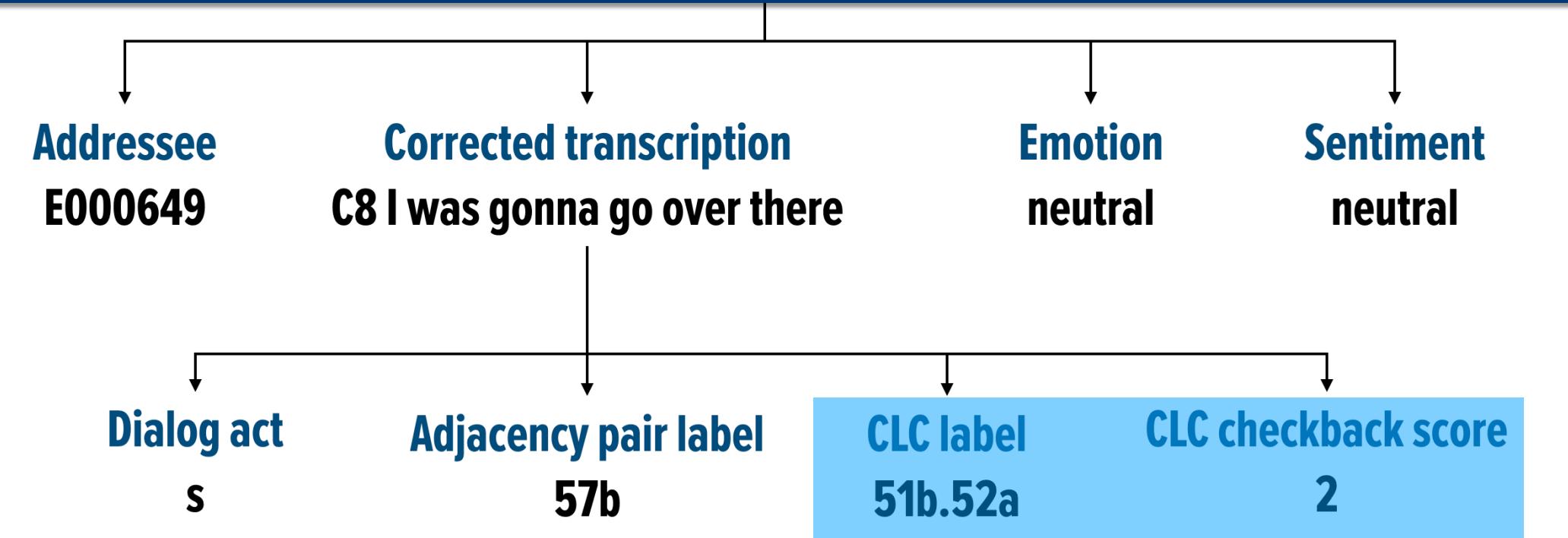
- Closed-loop communication (CLC) has been proposed in the team science literature as one of the coordinating mechanisms for effective teamwork ([Salas et al., 2005](#)).

- Implemented in military contexts to reduce frequency of communication breakdowns ([Burke et al., 2004](#)), and is being explored in the context of healthcare as well ([Parush et al., 2011](#)).



Original UUID
06e5da76-fff9-4d4f-9956-b4b4e7aeaae5

Start time	Speaker	Trial	ASR transcription	End time
16:16	E000651	T000604	C I was gonna go over there	16:18

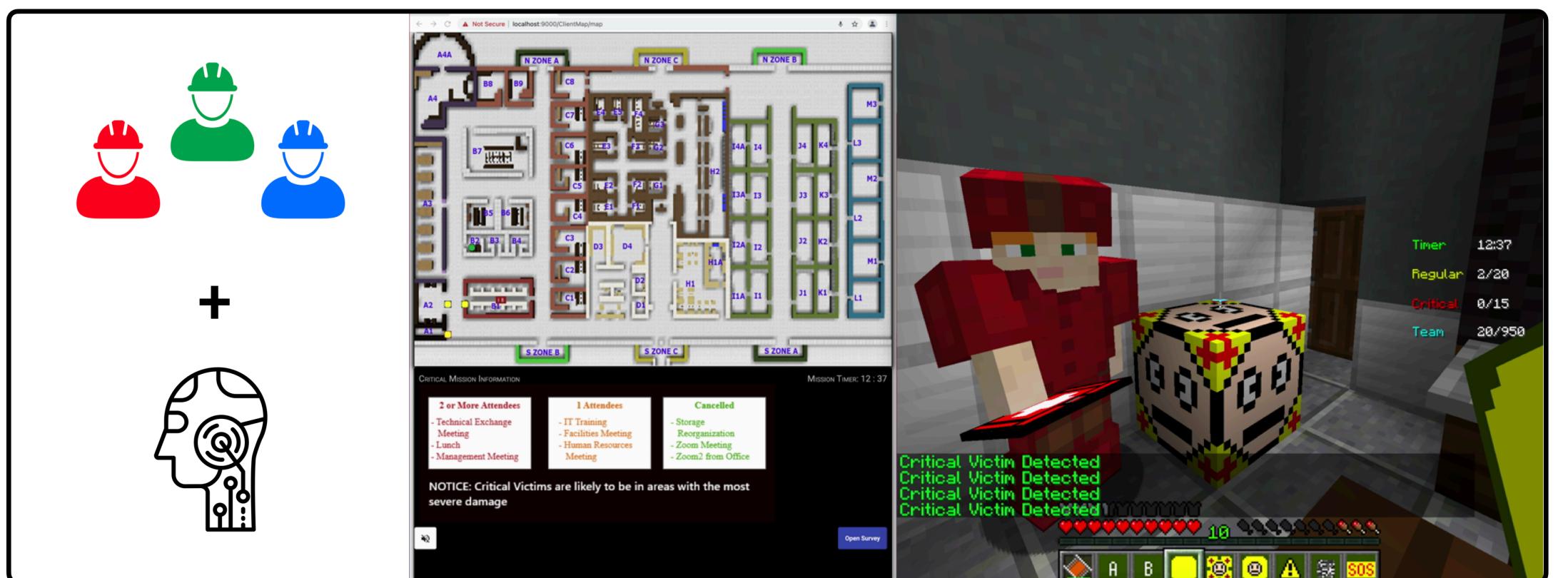


MultiCAT Annotations

Closed-Loop Communication

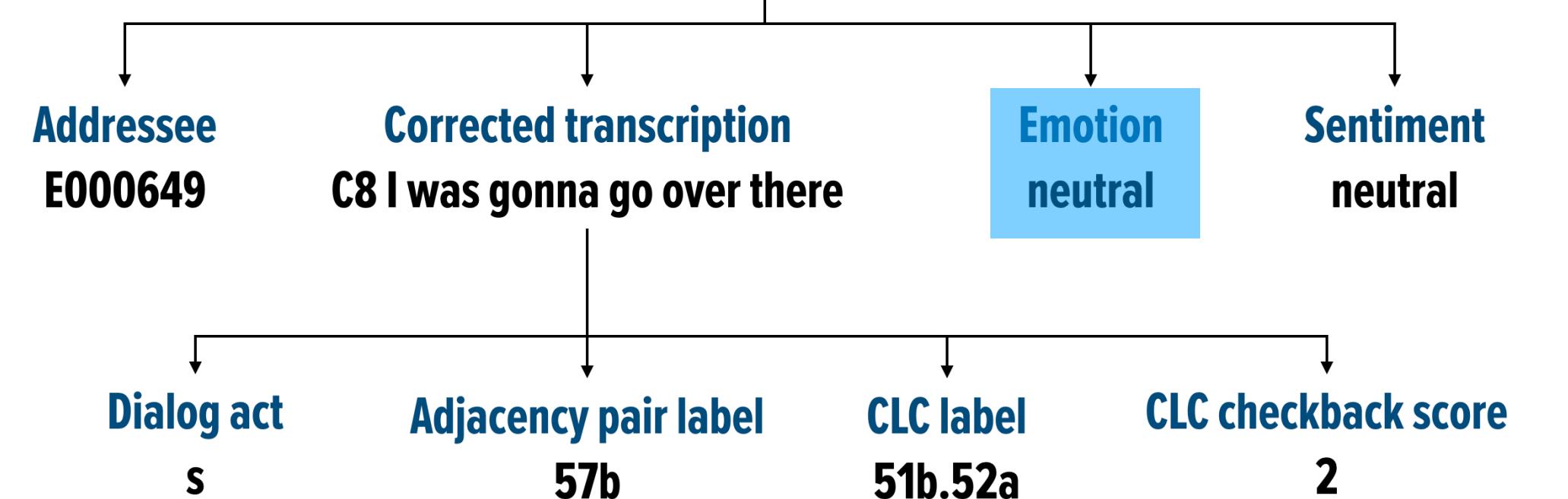
- Most definitions of what we refer to as a CLC ‘event’ include the following three sub-events occurring in sequence:

- (1) *Call-out*: Interlocutor 1 (I_1) shares information with/gives an instruction to interlocutor 2 (I_2),
- (2) *Check-back*: I_2 confirms their understanding of the information/instruction by repeating it back to I_1 ,
- (3) *Closing*: I_1 confirms that I_2 has received and understood the information or performed the desired action.



Original UUID
06e5da76-fff9-4d4f-9956-b4b4e7aeaae5

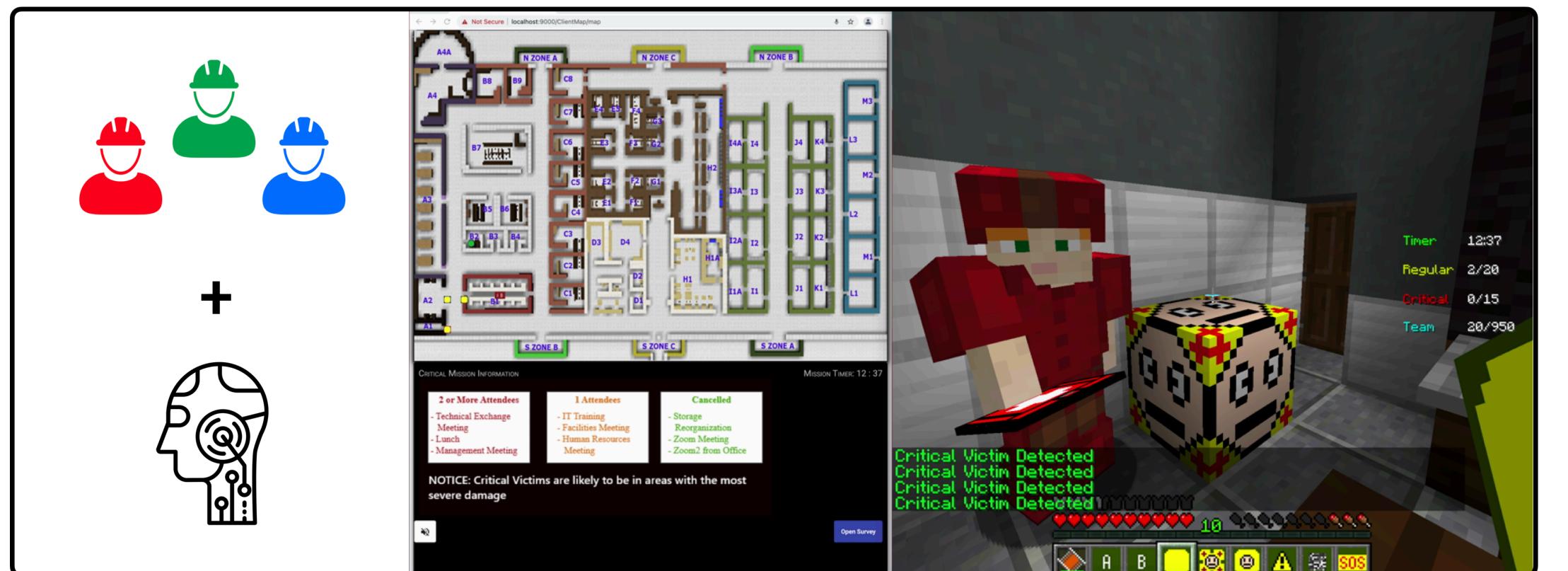
Start time	Speaker	Trial	ASR transcription	End time
16:16	E000651	T000604	C I was gonna go over there	16:18



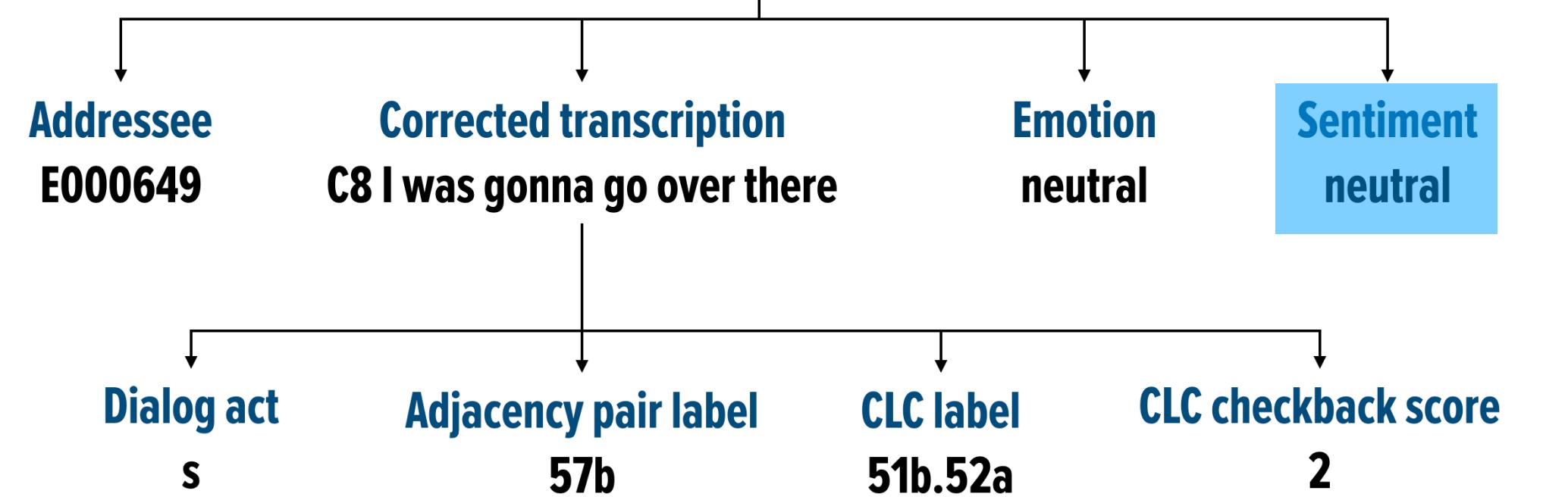
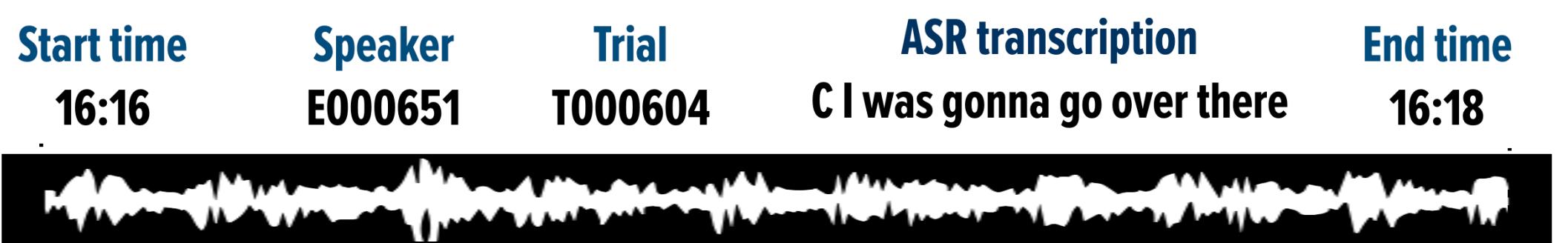
MultiCAT Annotations

Emotion

- We annotated for Ekman's universal emotions, along with an additional 'neutral' label:
 - *Anger*
 - *Disgust*
 - *Fear*
 - *Joy*
 - *Neutral*
 - *Sadness*
 - *Surprise*



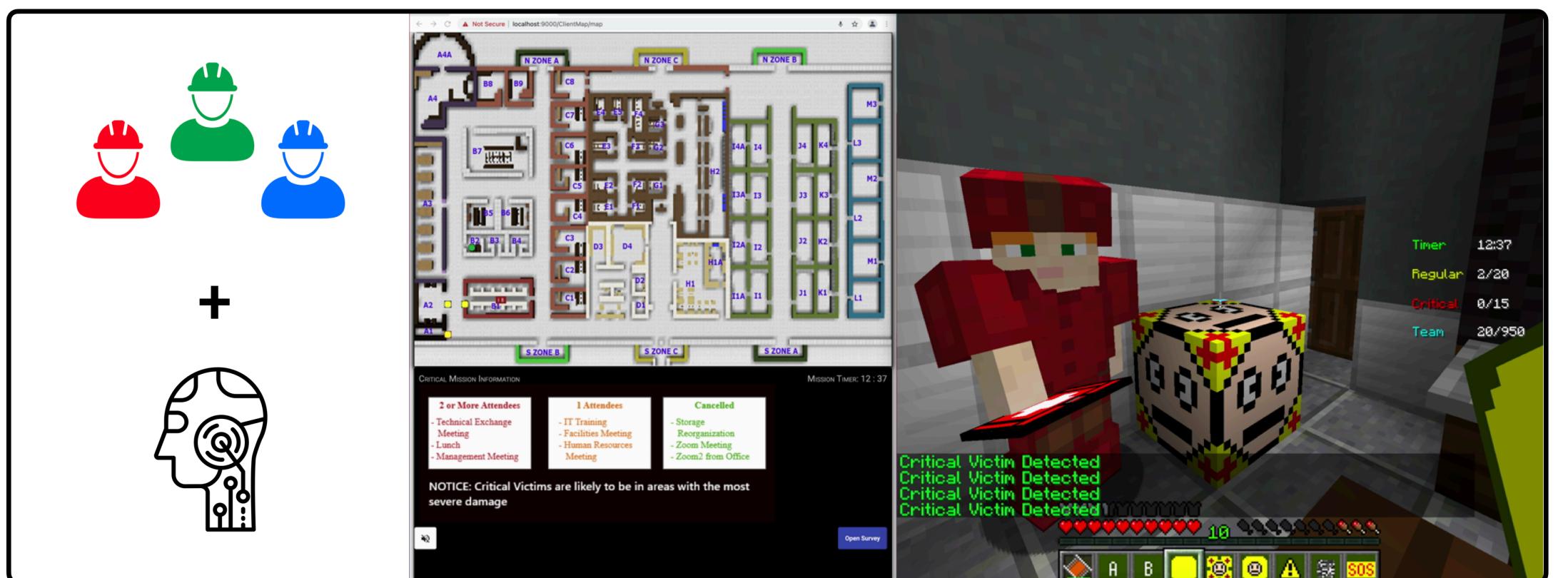
Original UUID
06e5da76-fff9-4d4f-9956-b4b4e7aeaae5



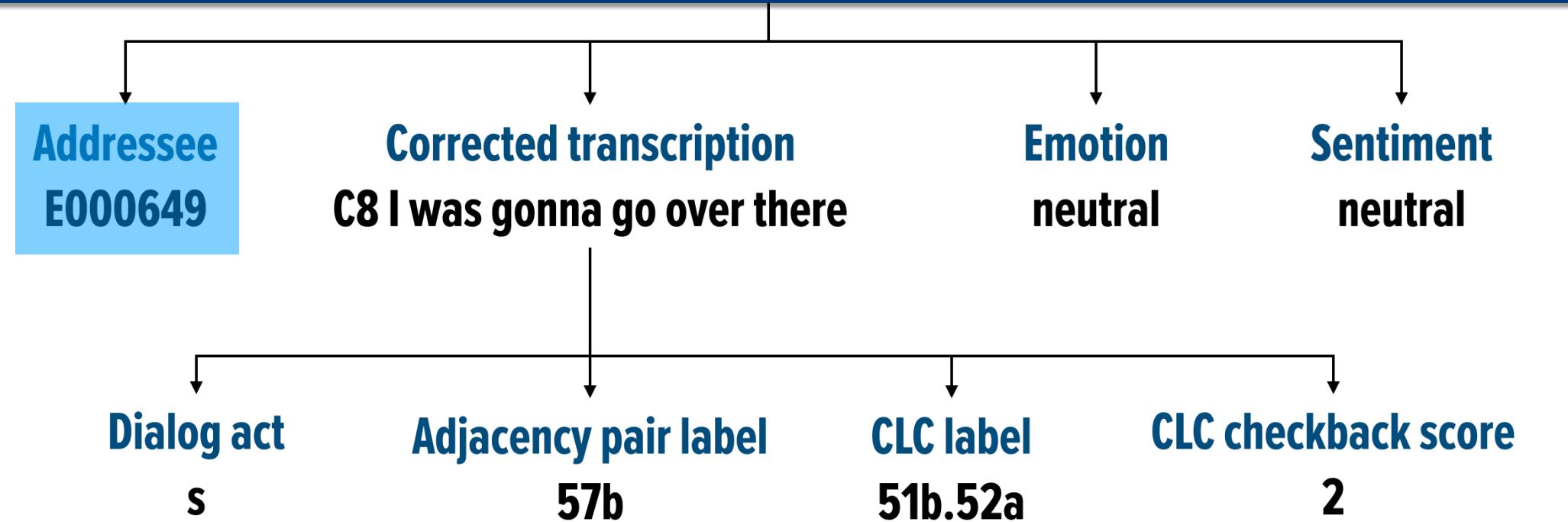
MultiCAT Annotations

Sentiment

- We annotated utterances with the following sentiment labels:
 - *Positive*
 - *Negative*
 - *Neutral*



Original UUID
06e5da76-fff9-4d4f-9956-b4b4e7aeaae5



MultiCAT Annotations

Entrainment

- We labeled utterances with the addressee (who the utterance was directed to) - these labels can be used for assessing *entrainment* — the unconscious modification of verbal, gestural, and linguistic actions by conversation partners to more closely resemble one another (Borrie and Liss, 2014), important for turn taking and building rapport.

Dataset: Selected Statistics

49	Trials
25	Teams
73	Speakers
11024	Utterances
151	Mean utterances/speaker
225	Mean utterances/trial
10	Mean word tokens/utterance

Notable firsts

- **First** publicly-available dataset for closed-loop communication detection.
- **First** dataset for unsupervised multi-party (i.e., > 2 interlocutors) vocal entrainment detection with remote participants.

Dialog Act Classification: Baseline results

Model	Fine-grained		Coarse-grained	
	Macro F_1 (SEM)	Accuracy (%)	Macro F_1 (SEM)	Accuracy
He et al.'s (2021)	30.75	63.24	42.15	93.92
LLaMA-3	34.76 (0.48)	66.47 (0.15)	44.55 (0.90)	94.66 (0.07)

Table 3: Macro F_1 and accuracy for DA classification on fine-grained and coarse-grained classes.

We found that the LLaMA-3 baseline outperforms He et al.'s (2021) by a considerable margin.

CLC Detection: Baseline

Stage	Accuracy	F_1
Call-out detection	.77	.79
Check-back detection	.76	.43
Complete CLC event detection	.51	.45

Sent/Emo classification

Sentiment	Support	Models		
		Strat.	Multi.	LLaMA-3
Negative	370	15.0	43.5	52.1 (0.0)
Neutral	1310	51.5	62.7	68.8 (0.0)
Positive	611	28.0	49.8	54.0 (0.0)
All	2291	31.5	52.0	58.4 (0.24)

Table 5: Results for sentiment prediction.

Emotion	Support	Models		
		Strat.	Multi.	LLaMA-3
Anger	18	5.4	0.0	3.9 (0.03)
Disgust	25	0.0	9.3	15.8 (0.0)
Fear	70	3.2	16.2	27.2 (0.02)
Joy	154	4.2	20.1	19.6 (0.01)
Neutral	1799	77.5	76.5	87.7 (0.0)
Sadness	145	5.6	30.5	36.7 (0.01)
Surprise	80	3.7	29.2	31.6 (0.02)
All	2291	14.2	26.0	31.8 (0.92)

Table 6: Results for emotion prediction.

Entrainment

- Attempted to validate existing neural entrainment distance (NED) measure proposed by Nasir et al. (2022), found that NED measure output by model trained on dyadic data does **not** pass validation checks on MultiCAT data.
- This suggests that a more general approach is required for modeling multi-party entrainment.

Relationship between annotation types

	DA	AP	CLC	Sentiment	Emotion
DA	—	N/A	✓	✓	N/A
AP		—	✓	✓	✓
CLC			—	✓	✓
Sentiment				—	✓
Emotion					—

Table 7: Combinations of label types for which the p -values computed using a χ^2 test of independence is less than 0.000125 (indicated by ✓). The matrix is symmetric, hence we omit the entries below the diagonal. We enter ‘N/A’ in the cells corresponding to the DA/AP and DA/Emotion combinations, as they do not satisfy the rules-of-thumb for χ^2 tests of independence discussed by Kroonenberg and Verbeek (2018).

Annotations and team performance

	Mission 1	Mission 2	Combined
# of trials	17	16	33
Proficiency	130 (26)	104 (19)	118 (17)
AP	126 (17)	100 (13)	118 (12)
CLC	125 (13)	99 (11)	116 (9)
DA	125 (11)	99 (9)	117 (8)
Sent	125 (10)	100 (8)	116 (7)
Emo	123 (9)	98 (7)	115 (6)
Multicat	123 (8)	97 (7)	115 (6)
All	122 (8)	97 (6)	115 (5)

Table 8: MAE (with SEM in parentheses) over all folds for our score prediction models.