2025 IEEE INTERNATIONAL CONFERENCE ON
ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP 2025)
April 06 - 11, 2025 Hyderabad, India

Towards Patronizing and Condescending Language in Chinese Videos: A Multimodal Dataset and Detector

1 Dalian University of Technology, China
2 University of Tsukuba, Japan
Hongbo Wang, Junyu Lu, Yan Han, Kai Ma, Liang Yang, Hongfei Lin

Github Repository

ID: 2608 **https://github.com/dut-laowang/PCLMM**

What is PCL?

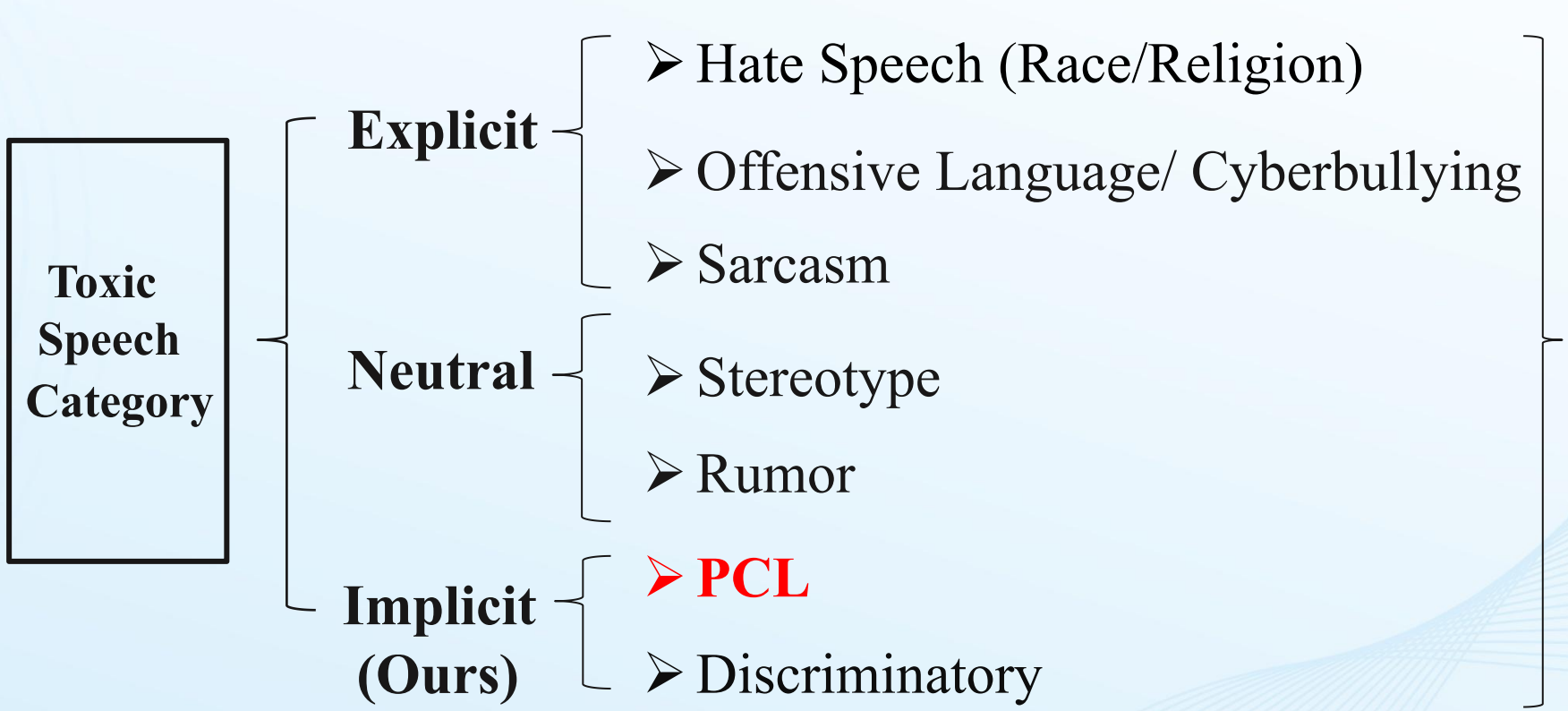
- Patronizing and Condescending Language (PCL) is a form of discriminatory toxic speech targeting vulnerable groups, such as individuals with disabilities, children, and the elderly, reflecting a superior attitude towards these communities. *(Perez-Almendros et al., 2020)*

PCL is a type of discrimination

Target at
Vulnerable groups

Various groups endured hardships

PCL - Type of Implicit Toxic



Conference

ACL

EMNLP

ICASSP

Journal

IEEE Transactions on Affective Computing

IEEE Access

Evaluation

SemEval19-24

WOAH20-21

TRAC-1/2/3/4

PCL - Hard to detect due to its implicit nature

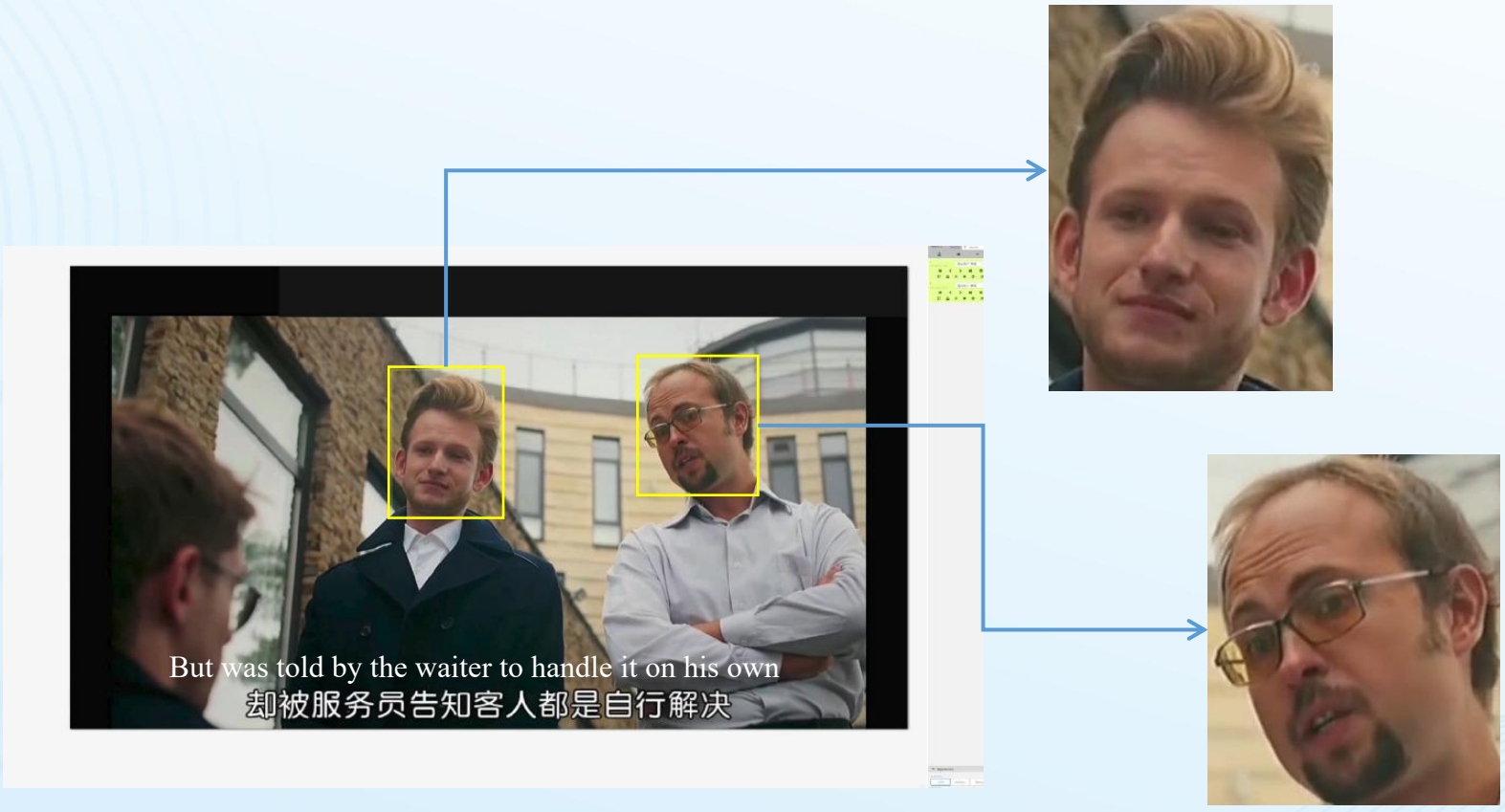
- SemEval 2022 Task 4 - PCL Detection (Only Text Modality)

	TEAM	P	R	F1		TEAM	P	R	F1		TEAM	P	R	F1
1	PALI-NLP	64.6	65.6	65.1	27	ML_LTU	58.0	51.4	54.5	53	RNRE NLP	39.0	50.2	43.2
2	stce	63.3	66.9	65.0	28	ZYBank-AI	54.8	53.9	54.4	54	SATLab	34.8	55.2	42.7
3	ymf924	63.8	65.6	64.7	29	Team LRL_NC	60.7	49.2	54.4	55	J.U.S.T-DL	49.0	37.5	42.5
4	BEIKE NLP	61.2	67.2	64.1	30	CS-UM6P & ESL	55.2	53.3	54.3	56	MaChAmp	58.8	32.8	42.1
5	holdon	60.3	67.5	63.7	31	Felix&Julia	40.1	77.3	52.8	57	I2C	61.1	31.2	41.3
6	cnxup	62.7	64.7	63.7	32	Stanford ACM	40.2	76.7	52.7	58	SMAZ	36.3	47.6	41.2
7	abcxyzw	58.8	68.5	63.3	33	UtrechtUni	44.6	62.5	52.0	59	MASZ	36.3	47.6	41.2
8	nowcoder	58.2	68.5	62.9	34	CSECU-DSG	59.0	46.4	51.9	60	Amrita_CEN	32.2	52.1	39.8
9	PINGAN Omini-Sinitic	61.8	63.7	62.7	35	Sapphire	59.4	46.1	51.9	61	Anonymus	27.6	59.9	37.8
10	bigemo	57.1	69.4	62.7	36	Ablimet	61.5	44.8	51.8	62	matan-bert	35.4	40.4	37.7
11	Leo_team	60.1	64.0	62.0	37	SSN_NLP_MLRG	42.3	66.6	51.7	63	Team LEGO	24.8	56.5	34.5
12	PAI-Team	66.3	57.7	61.7	38	Team PiCkLe	46.0	58.0	51.3	64	TüSoXi	38.8	29.3	33.4
13	Anonymus	53.5	70.4	60.8	39	sua	54.0	48.6	51.2	65	RNRE NLP RFC	30.0	36.9	33.1
14	BLING	63.5	55.5	59.3	40	UCL xNSI	41.5	65.3	50.7	66	jct_meir	25.3	47.0	32.9
15	Taygete	53.6	66.3	59.2	41	MS@IW	50.2	51.1	50.6	67	isys	22.4	59.3	32.5
16	NLP-Commonsense Reasoning team	61.2	56.8	58.9	42	University of Bucharest Team	49.1	50.8	49.9	68	AliEdalat team	18.4	87.1	30.3
17	GUTS	61.3	54.9	57.9	43	**RoBERTa Baseline**	**39.4**	**65.3**	**49.1**	69	ms_pa	23.4	39.1	29.3
18	DH-FBK	64.2	52.7	57.9	44	rematchka	44.5	53.9	48.8	70	Waad	64.0	18.0	28.1
19	ULFRI	56.4	58.7	57.5	45	fengxing	63.8	39.4	48.7	71	Ryan Wang	17.0	60.9	26.6
20	TUG-CIC	60.2	54.9	57.4	46	flerynn	67.2	38.2	48.7	72	PC1	37.8	18.6	25.0
21	amsqr	54.8	59.9	57.2	47	Team YNU-HPCC	65.9	36.6	47.1	73	UTSA_NLP	14.0	35.0	20.0
22	UMass PCL	52.9	58.4	55.5	48	niksss	51.8	42.0	46.3	74	yaakov	11.2	10.1	10.6
23	LastResort	51.5	59.9	55.4	49	JustTeam	55.0	39.8	46.2	75	ilan	14.5	6.0	8.5
24	Team Double_A	47.2	66.6	55.2	50	BWQ	51.0	41.3	45.6	76	Jiaaaaaa	8.2	6.3	7.1
25	thetundramanagainstpcl	54.3	55.5	54.9	51	Tesla	36.0	57.7	44.3	77	Anonymus	29.7	3.5	6.2
26	Xu	46.2	66.9	54.6	52	ASRtrans	35.6	58.4	44.2	78	Anonymus	10.6	2.8	4.5

Table 2: SemEval Task 4: Ranking by teams for Subtask 1: Binary Classification. The table reports Precision (%), Recall (%) and F1-Score (%) for the positive class.

The ambiguous toxicity of PCL limits the effectiveness of pure text-based detectors (e.g., BERT)

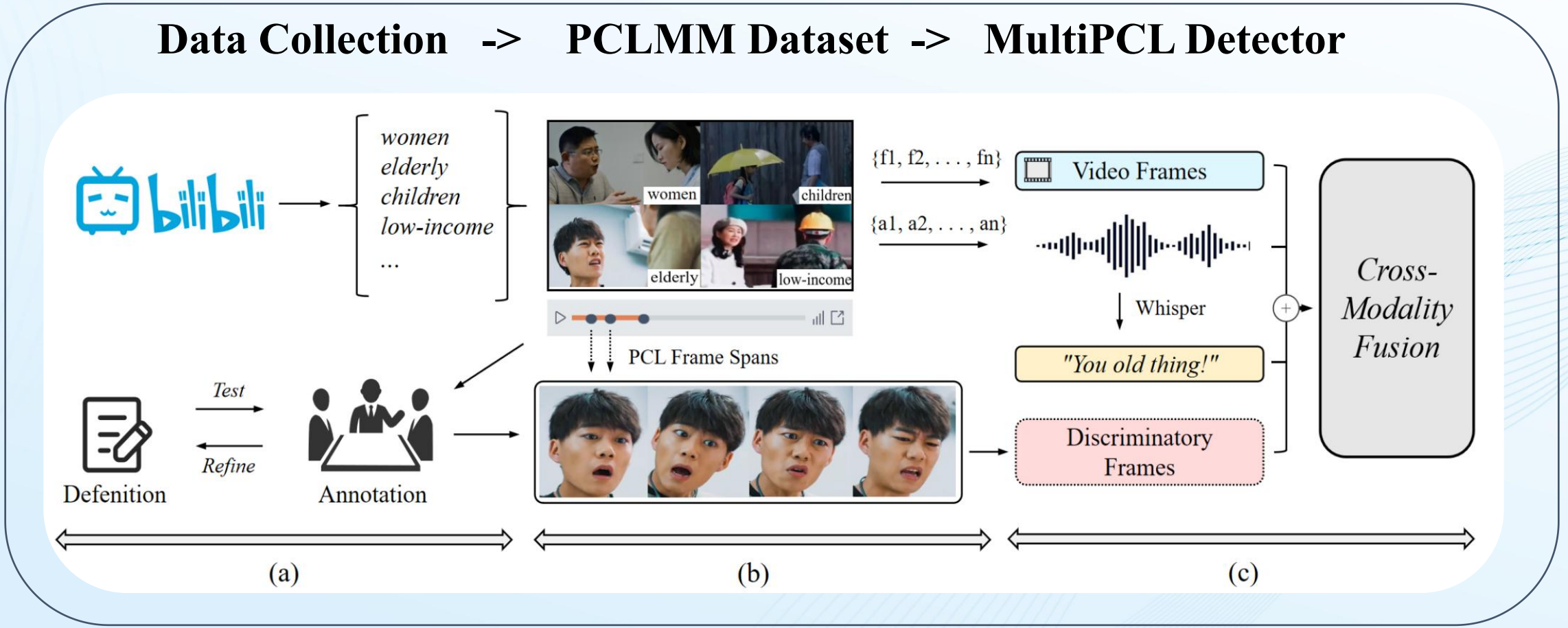
PCL - Hard to detect due to its implicit nature



- Discriminatory facial expression frames targeting vulnerable groups are widely present (e.g., in videos)
- Cognitively, PCL can be effectively distinguished using Multi-Modality

Towards Patronizing and Condescending Language in Chinese Videos

Ⅰ Introduction



Towards Patronizing and Condescending Language in Chinese Videos

II PCLMM (Definition)

To ensure rigor, we have refined the definition of Chinese PCL:

Chinese PCL refers to discriminatory, falsely sympathetic, and hypocritical remarks directed at six vulnerable groups within the Chinese community: ***disabled individuals***, ***women***, ***the elderly***, ***children***, ***single-parent families***, **and** ***low-income groups***. A key feature of PCL is the speaker's condescending attitude, making statements that do not improve the group's situation. PCL expressions are often accompanied by contemptuous and discriminatory facial expressions.

The following situations will not be classified as PCL:

1. When vulnerable groups describe their own experiences of unfair treatment.
2. Objective news reports on discriminatory incidents.
3. Public service announcements containing discriminatory content but without discriminatory intent.

Towards Patronizing and Condescending Language in Chinese Videos

II PCLMM (Data Collection & Annotation)

Category(6)	* Keyword(10)	* Query(30)
Elderly	Nursing home Elderly individuals Senior population Retirement community Elderly aged Empty-nest seniors	Discrimination Disdain Snub Dislike Despise Condescending Loathe Hate .....
disabled		
women		
children		
single-par.		
low-income		

1800 search lists → bilibili

2654 pre-selected videos

715 high-quality videos

	Non-PCL	*PCL*	*PCL Frame Spans*	*Total*
Total num	519	196	330	715
Total len (hrs)	15.1	6.5	2.3	21.6
Total frame (M)	1.6	0.7	0.2	2.3
μ Video len (min)	1.7	1.9	0.4	1.8
μ Text len (char)	455	536	158	477

Criteria for Judgment:
Samples containing one or more discriminatory segments will be classified as PCL samples

PCL Frame Spans:
Manually annotated PCL discriminatory segments

Towards Patronizing and Condescending Language in Chinese Videos

II PCLMM (Sentiment Analysis)

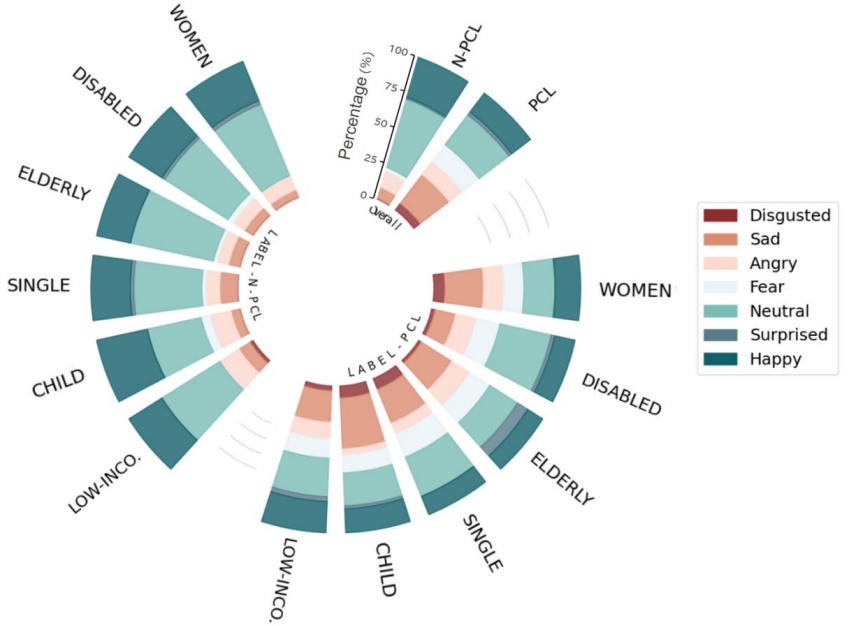


Fig. 2: Sentiment analysis for the six vulnerable groups in PCLMM.

We used the advanced open-source model **DeepFace** to analyze facial expressions in the PCLMM dataset (separately counting positive and negative samples)

Observation:

- The difference in emotional polarity between N-PCL and PCL samples is not significant
- A large amount of **false praise and sympathy** exists in PCL, leading to an increase in Positive Emotion.

Towards Patronizing and Condescending Language in Chinese Videos

II PCLMM (Toxicity Analysis)

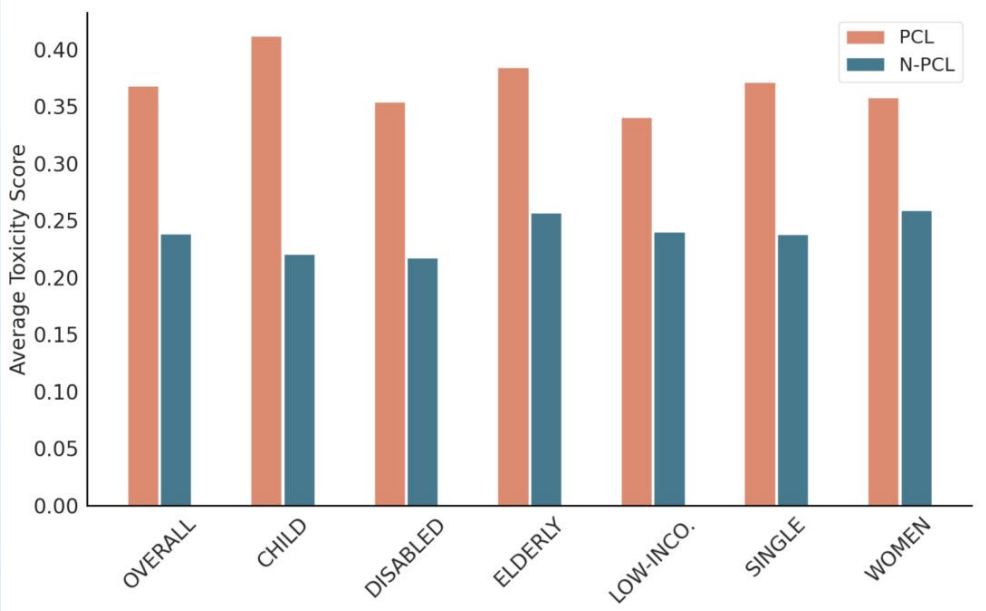


Fig. 3: Average toxicity scores in PCLMM.

We scored our transcribed texts using the **Perspective API** for six groups

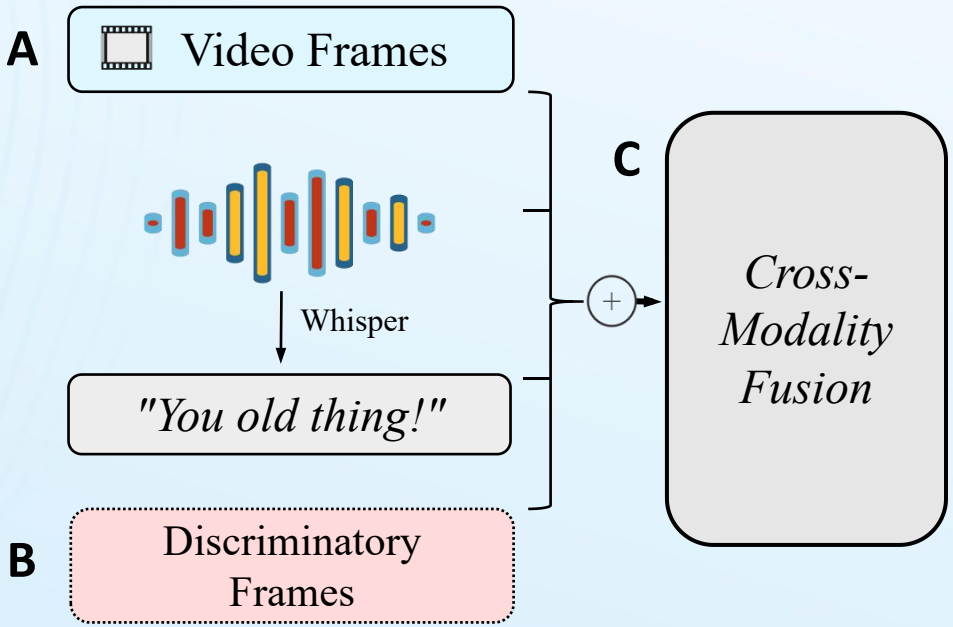
Observation:

- Although the toxicity of PCL samples is slightly higher than that of N-PCL, overall, the toxicity score is far lower than that of hate speech (mean: 0.7)

Towards Patronizing and Condescending Language in Chinese Videos

III Method

$$X : X(F; F_v; A; T) \rightarrow y, \text{ where } y \in \{0, 1\}$$



A *Video Features*

$$\mathbf{z}_i = \mathrm{ViT}(f_i), \quad \mathbf{z}_i \in \mathbb{R}^{d_v}, \quad i = 1, 2, \ldots, n$$

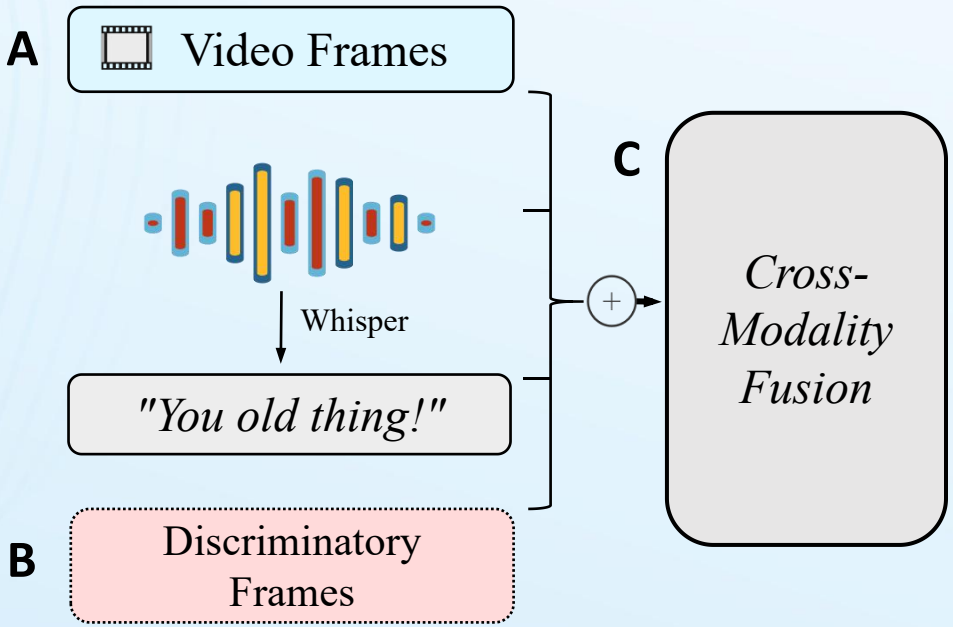
B *Facial Expression Features*

$$\mathbf{z}_i^v = \begin{cases} \text{FER-VT}(f_i v), & \text{if MTCNN detects a face in } f_i \\ f_i v = 0, & \text{if no face is detected} \end{cases}$$

Towards Patronizing and Condescending Language in Chinese Videos

III Method

$$X : X(F; F_v; A; T) \rightarrow y, \text{ where } y \in \{0, 1\}$$



C *Modality Fusion*

$$\text{MHCA}(\mathbf{Q}_i, \mathbf{K}_j, \mathbf{V}_j) = \text{Softmax}\left(\frac{\mathbf{Q}_i \mathbf{K}_j^\top}{\sqrt{d_k}}\right)\mathbf{V}_j$$

$$A_{i,j} = \text{MHCA}(\mathbf{Q}_i, \mathbf{K}_j, \mathbf{V}_j), \quad i, j \in \{z, z^v, z^a, z^t\} \quad \mathbf{Z} = \sum_{i,j} A_{i,j}$$

Towards Patronizing and Condescending Language in Chinese Videos

Ⅳ Experiment

M	Model	P_p	R_p	$F1_p$	$F1_m$	Acc
A	MC	35.81	56.89	45.21	54.28	64.14
T	RC	54.84	50.00	52.31	69.14	78.32
	BP	58.06	52.94	55.38	71.13	79.72
	GPT4	65.52	55.88	60.32	74.55	82.52
F	FT	65.52	47.50	55.07	70.46	78.47
V	VM	61.76	52.50	56.76	70.90	77.78
	VT	65.62	52.50	58.33	72.22	79.17

In the unimodal setting, the text (T) modality performs best, while the audio-only (A) modality performs poorly

$F1_m$ denotes the macro-averaged F1 score. Abbreviations: MC (MFCC), RC (RoBERTa-Chinese) BP (BERT-PCL), FT (FER-VT), VM (VideoMAE), VT (ViT).

Towards Patronizing and Condescending Language in Chinese Videos

Ⅳ Experiment

M	Model	P_p	R_p	$F1_p$	$F1_m$	Acc
A+F	MC+FT	39.13	45.00	41.86	58.55	65.28
A+T	MC+BP	58.82	50.00	54.05	69.08	76.39
T+F	BP+FT	62.89	55.00	58.67	72.06	78.47
A+V	MC+VT	58.00	72.50	64.44	74.14	77.78
V+F	VT+FT	62.79	67.50	65.06	75.46	79.86
V+T	VT+BP	63.04	72.50	67.44	76.79	80.56

In the multimodal scenario, the video/facial expression modality plays a crucial auxiliary role in feature understanding, significantly improving the performance of V+T/V+F

$F1_m$ denotes the macro-averaged F1 score. Abbreviations: MC (MFCC), RC (RoBERTa-Chinese)
BP (BERT-PCL), FT (FER-VT), VM (VideoMAE), VT (ViT).

Towards Patronizing and Condescending Language in Chinese Videos

Ⅳ Experiment

M	Model	P_p	R_p	$F1_p$	$F1_m$	Acc
A+T+F	MC+BP+FT	61.90	65.00	63.41	74.43	79.17
V+T+F	VT+BP+FT	64.44	72.50	68.24	77.47	81.25
V+T+A	VT+BP+MC	65.91	72.50	69.05	78.15	81.94
V+A+F	VT+MC+FT	67.44	72.50	69.88	78.84	82.64
V+A+T+F	**MultiPCL**	**68.09**	**80.00**	**73.56**	**81.06**	**84.03**

Our proposed MultiPCL framework integrates four modalities: text, audio, video, and facial expressions, significantly outperforming existing baselines. The performance improvements for single- (**6.51%**), dual- (**4.27%**), and tri-modal(**2.22%**) , respectively.

$F1_m$ denotes the macro-averaged F1 score. Abbreviations: MC (MFCC), RC (RoBERTa-Chinese)
BP (BERT-PCL), FT (FER-VT), VM (VideoMAE), VT (ViT).

Towards Patronizing and Condescending Language in Chinese Videos

Ⅴ Conclusion

Our Contributions:

- ***PCLMM Dataset*** — The first multimodal PCL video dataset for discriminatory speech detection, featuring 715 annotated videos totaling over 21 hours
- ***MultiPCL Detector*** — The first approach integrating video and discriminatory facial expression features for multimodal PCL detection

Future Work:

- ***Expand Research*** — Expanding PCL studies to microaggressions, sarcasm, and stereotypes, exploring synergies with other abusive speech tasks
- ***LLM Evaluation*** — Serving as a benchmark to assess multimodal large language models, particularly video-integrated models, for microaggression detection

ACKNOWLEDGMENTS

We also thank the Natural Science Foundation of China (*No. 62376051,62076046,62076051,61702080*),
the National Language Commission Key Program (*No. ZDI145-80*),
the Liaoning Province Applied Basic Research Program (*No. 2022JH2/101300270*),
the Liaoning Provincial Natural Science Foundation Joint Fund Program (*2023-MSBA-003*),
and the Fundamental Research Funds for the Central Universities award number (*DUT24LAB123*).

We would like to thank all reviewers for their constructive comments!

For more information **https://github.com/dut-laowang/PCLMM**

Github Repository