Low-Latency Privacy-Aware Robot Behavior guided by Automatically Generated Text Datasets

Yuta Irisawa[1,2], Tomoaki Yamazaki[1], Seiya Ito[3], Shuhei Kurita[4], Ryota Akasaka[5], Masaki Onishi[2], Kouzou Ohara[1], Ken Sakurada[6]

1. Aoyama Gakuin University 2. National Institute of Advanced Industrial Science and Technology
3. National Institute of Information and Communications Technology 4. National Institute of Informatics 5. Osaka University 6. Kyoto University

Abstract

Objective : **Realize humanoid robot exhibiting privacy-aware behavior.**

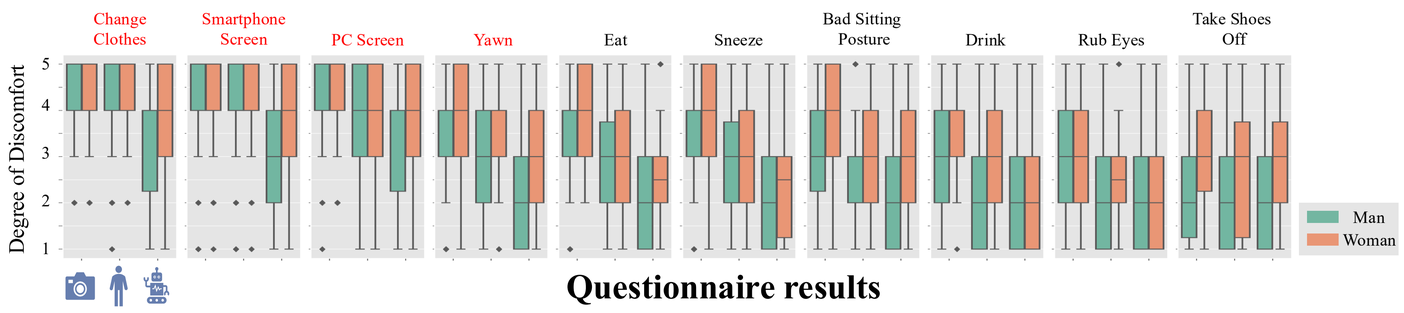
Summary : **Without collecting ethically sensitive images, distill knowledge from an LLM and achieve text-based image privacy recognition.**

Background

Humanoid robots with human-like appearance should behave naturally and consider privacy
⇒ User study with 200 participants

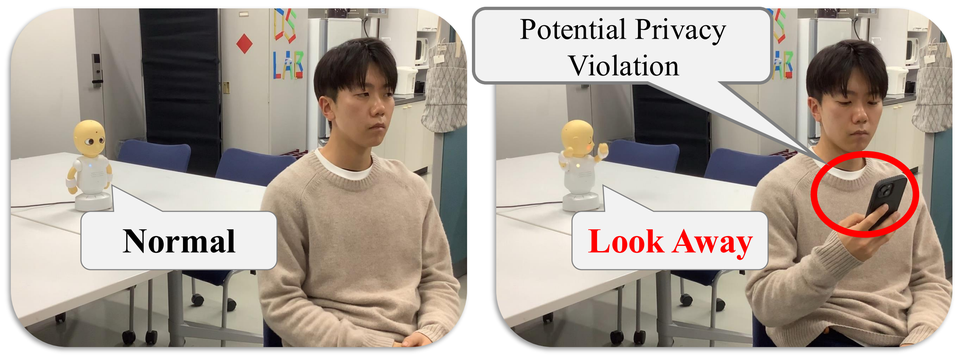
◆ Discomfort arises when a robot observes private scenes (e.g., changing clothes, smartphone/PC screen).

◆ Robot observation causes less discomfort than cameras or humans. However, discomfort remains in private contexts.



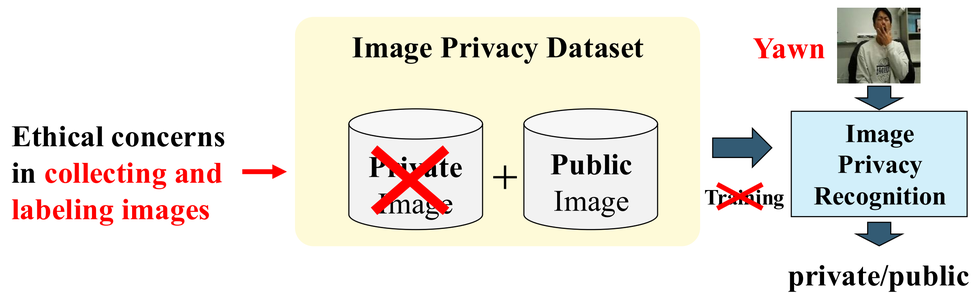
Questionnaire results

※ Discomfort rating for 10 types of actions and information were collected on a 5-point scale across 3 conditions



Problems of existing methods

■ Specialized model (Image Privacy Recognition Model) : usually requires an image privacy dataset



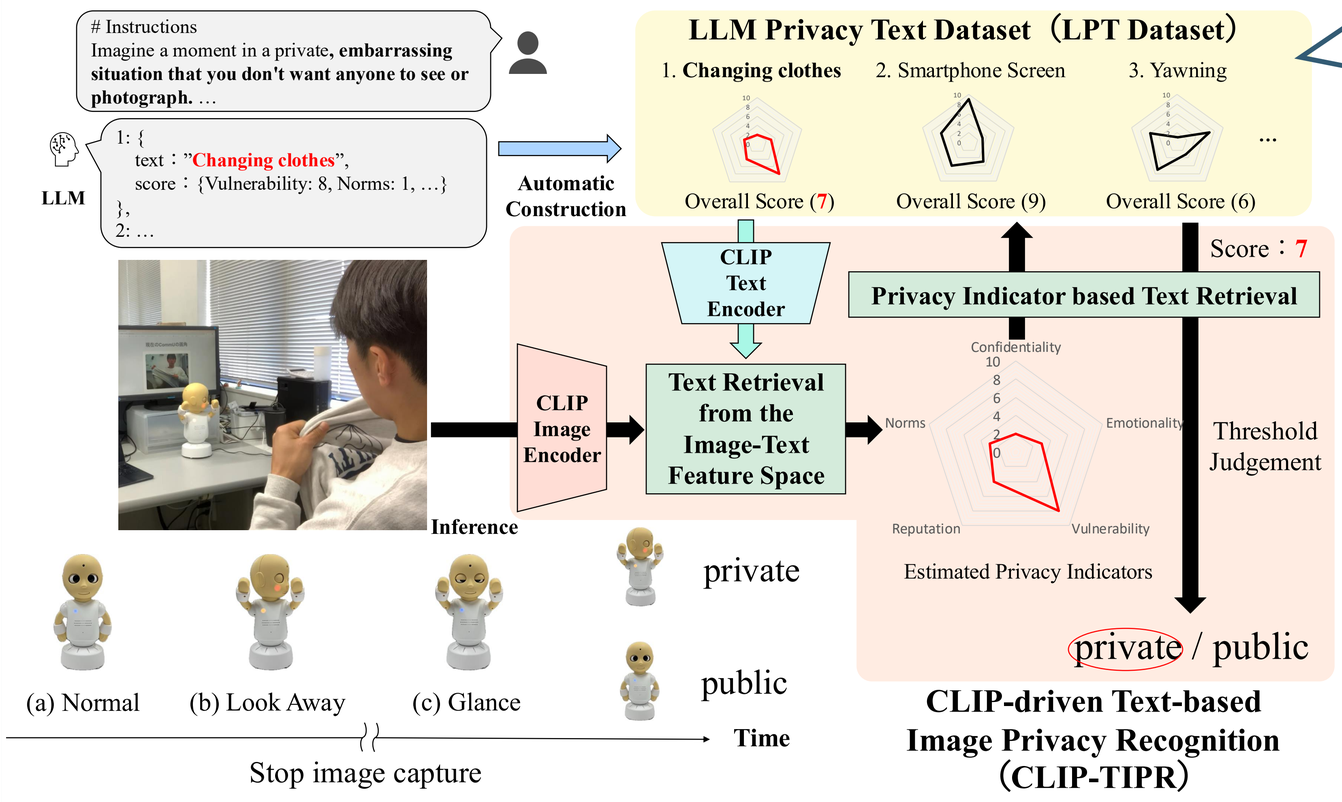
■ General-purpose model (LVLM): requires about 3 seconds for high-accuracy recognition, causing large delay

Proposed Framework

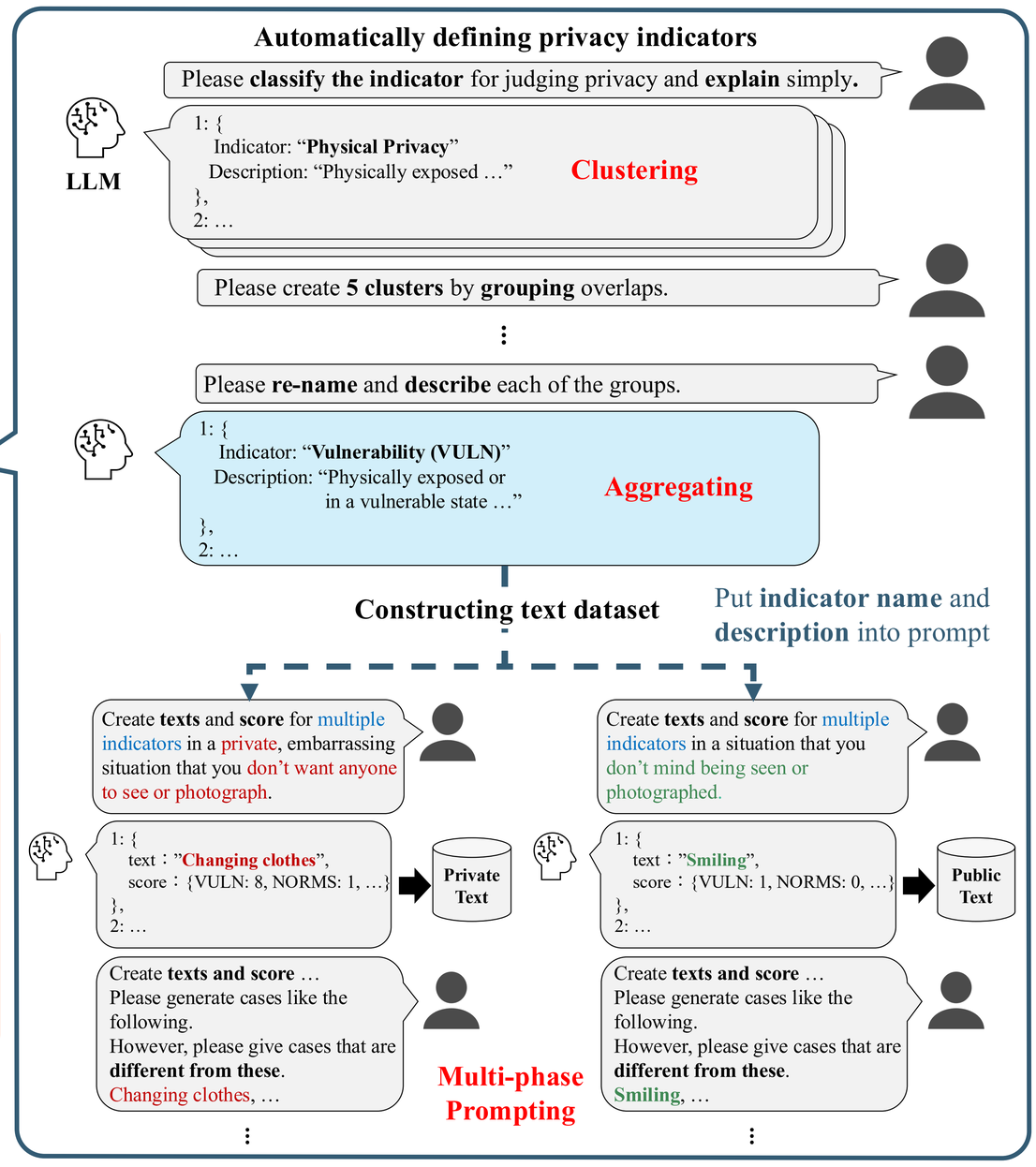
Contributions

i. Enable humanoid robots to **recognize privacy-sensitive situations** and **respond appropriately**.

ii. Propose a **low-latency and robust** recognition method with **fewer ethical concerns** by utilizing a **text dataset automatically constructed by an LLM**.



CLIP-driven Text-based Image Privacy Recognition (CLIP-TIPR)



Experiments

Details of LPT Dataset

- Model: GPT-4o
- Number of Indicators: 5
- Generated Texts: 1,000 × 2 (private/public)

Evaluation Metrics

macro F1-score

Comparison Methods

#	Methods	Text Dataset
I	ZSVCIP	LPT Dataset (binary)
II	CLIP-TIPR (Text-based)	Single Abstract Indicator ※ Construct with only overall score
III	CLIP-TIPR (Indicator-based)	LPT Dataset
IV	LVLM-simple	-
V	LVLM-cot	-

Examples of LPT Dataset text and scores

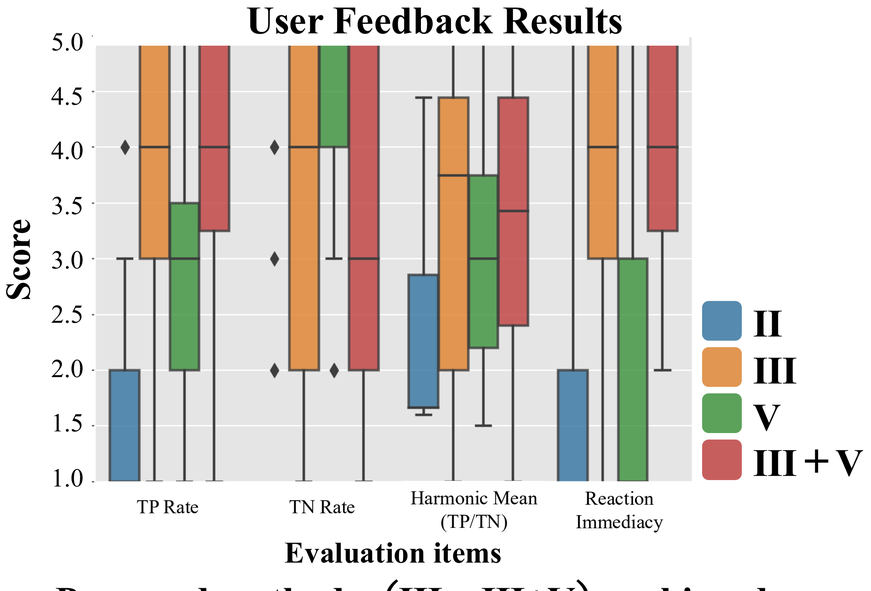
Text	CONF	EMO	VULN	REP	NORMS	OA
Having your pants accidentally torn in a very public place.	0	**6**	**8**	**7**	**9**	**10**
Having your phone sent a personal message to a group chat by mistake.	5	**7**	0	**6**	**8**	**9**
Photographed yawning widely in a serious meeting.	0	2	1	3	**6**	**6**

Quantitative Evaluation Results

#	macro F1-score ↑			Processing Time (s)↓
	CPI Dataset	PrivacyAlert	VISPR	
I	0.500	0.543	**0.525**	0.059
II	0.440	0.577	0.306	**0.058**
III	**0.680**	0.582	0.461	0.074
IV	0.417	0.249	0.437	2.040
V	0.644	**0.706**	0.415	3.198

CLIP-TIPR (III) achieved accurate and robust performance

User Feedback Results



Proposed methods (III・III+V) achieved accurate and low-latency recognition

Future Work

- Improve recognition for **fine-grained actions and objects.**
- Address differences in **privacy perceptions across cultures and individuals.**