



Understanding Privacy Risks and Perceived Benefits in Open Dataset Collection for Mobile Affective Computing

Hyunsoo Lee, Soowon Kang, Uichin Lee

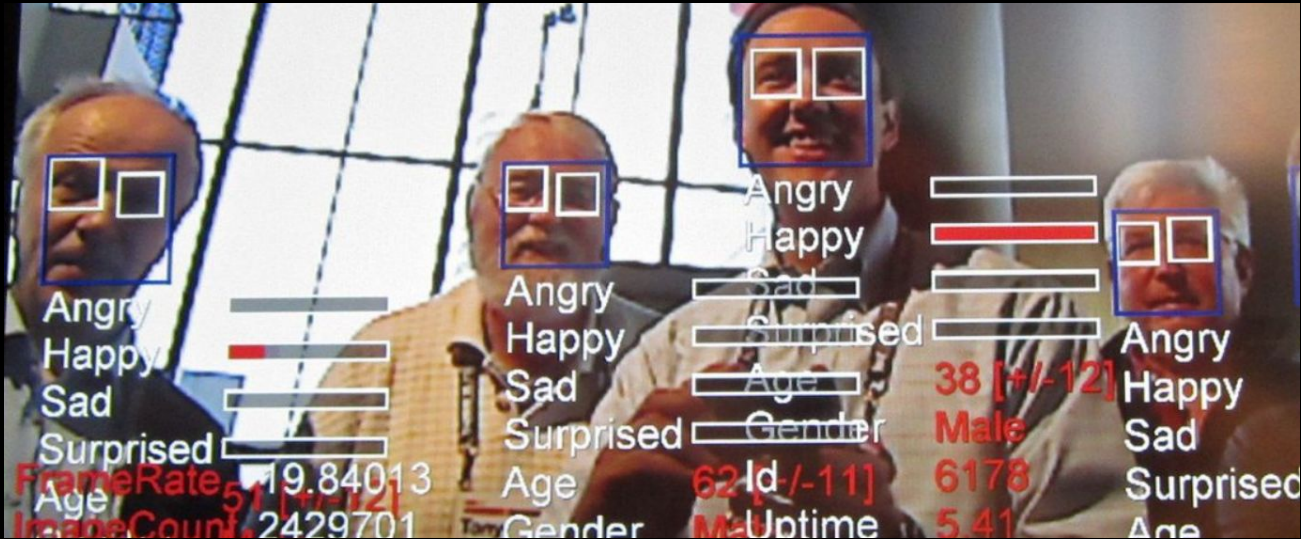
hslee90@kaist.ac.kr



2022UBICOMP

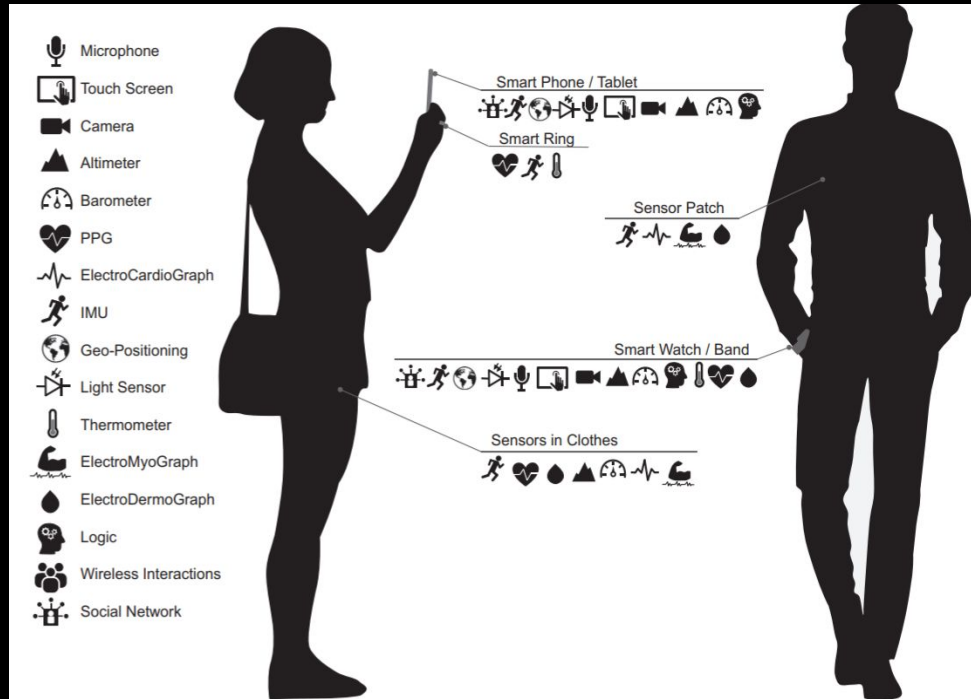
September 11 – 15
Atlanta, USA and Cambridge, UK

Sensor-Driven Research for Affective Computing



Model users' emotion/cognitive states and mental health states by using mobile and wearable sensor data collected in-the-wild

Limited Public Dataset Release



Available open datasets: StudentLife (Wang et al., 2014), The Tesseract Project (2019)

Privacy Concerns in Mobile Contexts



24/7 Passive Sensing!

Research Motivation

- Open dataset collection with mobile and wearable devices will make significant contribution to ubiquitous computing research because the large volume of behavioral and contextual data across multiple sensors can be used to develop and evaluate in-situ psychological state inference algorithms
- To increase the number of participants and make them more cooperative along the data collection projects, **reducing privacy concerns and identifying participants' potential benefits (i.e., participation motives)** are important prerequisites

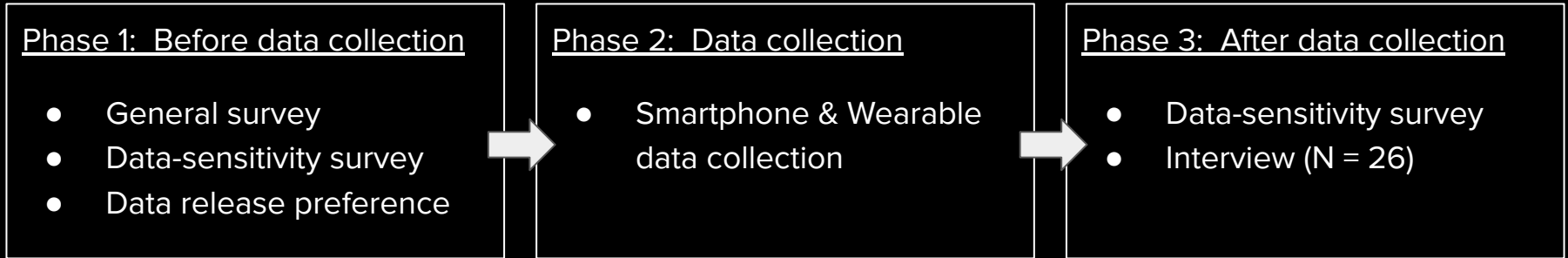
Research Goal

RQ1: What are the participants' privacy concerns and general motives regarding in-the-wild open dataset collection for affective computing research?

RQ2: What factors are associated with participants' attitudes toward the in-the-wild open dataset collection for affective computing research?

Research Method: Overview

A four-week study open dataset collection project in-the-wild (N = 100, 32 females)



Research Method: Survey Items

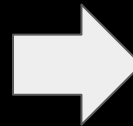
- **Survey 1: General survey (7-point Likert scale)**
 - Demographics
 - Confidence in knowledge
 - Participation motive
 - Perception on open dataset collection (Shah et al., 2019)
 - Risk-benefit assessment of open dataset collection (Oliver et al., 2012, Shah et al., 2019)
 - Perceived level of privacy concerns (Xu et al., 2008, Xu et al., 2012)
 - Level of trust (Liu et al., 2005)

Research Method: Survey Items

- **Survey 2: Data sensitivity survey (7-point Likert scale)**
 - Perceived sensitivity of each collected data
 - Conducted twice (Pre vs. Post)



Smartphones
Polar H10
Fitbit Inspire HR



Total 23 types of data
from smartphones and
wearables

Collected Data

Device	Category	Data Type
Smartphone	Location	GPS
	Network	Wi-Fi/Bluetooth/Cellular Nearby wireless signals (e.g., SSID), Data traffic
	Device status	Power status, Ringer mode
	Battery	Battery level, Charging status
	Calls/Texts	Phone call history, Text message history
	Keyboard	Keyboard type, Input key type, Keyboard distance
	Media	Camera use, Screenshot
	App	App usage stats, Notification history
	Activities	Activity types
Wearables	Biosignals	Heart rate
	Activities	Calorie, Steps, Stairs, Distance, Sleep

Research Method: Survey Items

- **Survey 3: Data release preference**

- Complete release (N = 85)
- Selective release (N = 15) → Given an option to select 'non-release data'

Participants	Non-Release Data	Incentive Offer
P1	Type of input key, Distance b/w consecutive input keys, Camera use events, App usage statistics, Notification history	
P2	Distance b/w consecutive input keys, App usage statistics, Notification history	
P3	GPS, Phone call history, Text message history, Type of input key, Camera use events, App usage statistics, Notification history	
P4	Notification history	Yes
P5	Distance b/w consecutive input keys, App usage statistics	
P6	GPS, Phone call history, Text message history, Type of input key, Camera use events, App usage statistics, Notification history, Heartrate, PSS, PHQ-9, PPC, RSE, Class satisfaction	
P7	GPS, Phone call history, Text message history, App usage statistics, Notification history	
P8	Notification history	
P9	GPS	
P10	GPS	
P11	Notification history	
P12	GPS, Phone call history, Notification history	Yes
P13	GPS	
P14	GPS	
P15	GPS	Yes

Research Goal

RQ1: What are the participants' privacy concerns and general motives regarding in-the-wild open dataset collection for affective computing research?

RQ2: What factors are associated with participants' attitudes toward the in-the-wild open dataset collection for affective computing research?

RQ1-1. Perceived level of privacy concerns and general motives

General survey results

(1: Highly Disagree ~ 7: Highly Agree, and N/A or don't know)

Category	Survey Item	Mean	SD
	Financial compensation	5.67	1.20
<ul style="list-style-type: none">● Financial compensation was the most influential participation motive for open dataset collection● Participants showed contradictory responses in terms of evaluating risks● Participants generally showed moderate level of privacy concerns			
Privacy Concerns	Perceived surveillance	4.01	0.04
	Perceived intrusion	4.29	1.52
	Secondary use of personal data	4.50	0.04

RQ1-2. Data sensitivity and privacy concerns

Pre-post sensitivity of each sensor data

(1: Highly Negative ~ 7: Highly Positive, and N/A or don't know)

Device	Category	Data Type	Pre-Sensitivity Mean (SD)	Post-Sensitivity Mean (SD)
<ul style="list-style-type: none">● Participants responded sensitive to certain data types from smartphone: <u>text logs, call logs, app usage, app notifications, camera and GPS</u>● Participants' perceived sensitivity on data types did not change after data collection ($p = .56$)				
	Media	Camera use event	3.89 (1.54)	3.72 (1.60)
	Location	GPS	3.89 (1.54)	3.72 (1.60)

Research Goal

RQ1: What are the participants' privacy concerns and general motives regarding in-the-wild open dataset collection for affective computing research?

RQ2: What factors are associated with participants' attitudes toward the in-the-wild open dataset collection for affective computing research?

RQ2. Factors Associated with People's Attitudes on Open Dataset Collection

- Semi-structured interview on **26** participants (P1 - P15: selective release, P16 - P26: complete release)

Theme	Sub-Theme
Incentive	Participation compensation / Privacy-utility trade-offs
Knowledge & Experience	Research interest / Learning personal life patterns / Wearable device experience
Scientific contribution	Data volume / AI service and research / Institutional scientific contribution
Privacy risks	Routine identification / Judgment & Categorization / Surveillance / Data misuse & leakage
Lack of justification	Research purpose / Sensor data usage
Autonomy	Higher autonomy / Lower autonomy
Trust	Institutional trust / Data handling trust

Revisiting...

- **Data release preference**

- Complete release (N = 85)
- Selective release (N = 15) → Given an option to select 'non-release data

Participants	Non-Release Data	Incentive Offer
P1	Type of input key, Distance b/w consecutive input keys, Camera use events, App usage statistics, Notification history	
P2	Distance b/w consecutive input keys, App usage statistics, Notification history	
P3	GPS, Phone call history, Text message history, Type of input key, Camera use events, App usage statistics, Notification history	
P4	Notification history	Yes
P5	Distance b/w consecutive input keys, App usage statistics	
P6	GPS, Phone call history, Text message history, Type of input key, Camera use events, App usage statistics, Notification history, Heartrate, PSS, PHQ-9, PPC, RSE, Class satisfaction	
P7	GPS, Phone call history, Text message history, App usage statistics, Notification history	
P8	Notification history	
P9	GPS	
P10	GPS	
P11	Notification history	
P12	GPS, Phone call history, Notification history	Yes
P13	GPS	
P14	GPS	
P15	GPS	Yes

RQ2. Factors Associated with People's Attitudes on Open Dataset Collection

- Incentives

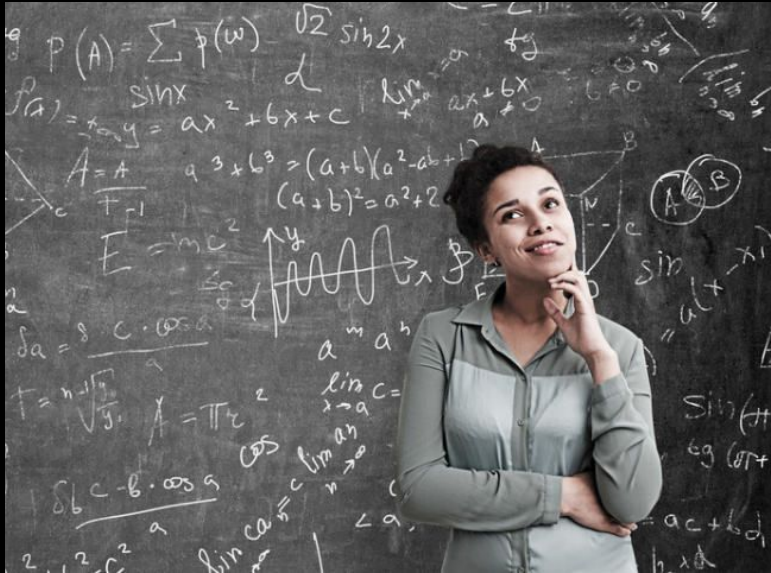


“Of course I did it for the money! (laughs) ... but later on, I also came to think about the scientific contribution that I can make from agreeing to this open dataset collection.” - P3

“Only 10,000 KRW [approximately 10 USD] for complete release? Come on, you guys are being cheapskates! I’d reconsider if the offer was higher.” - P6

RQ2. Factors Associated with People's Attitudes on Open Dataset Collection

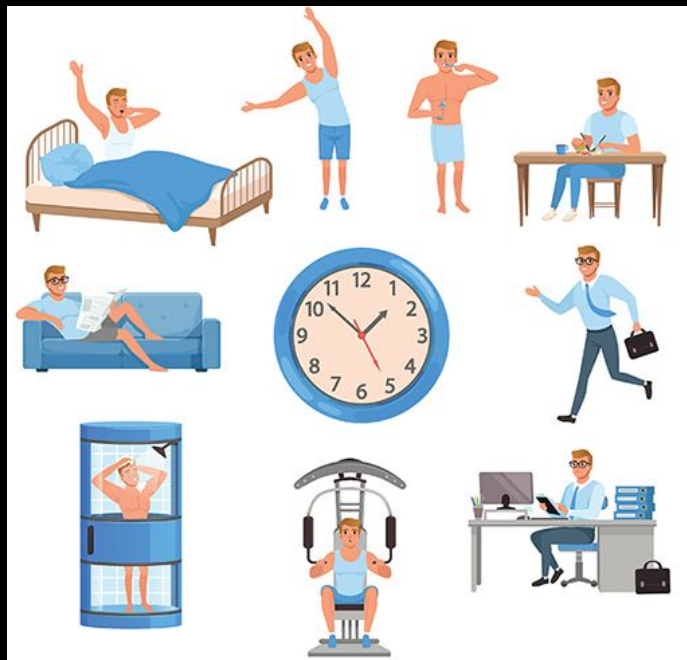
- **Scientific Contribution**



*“Umm, I think this experiment could be a bit dangerous as it collects and aims to release an extensive dataset of an individual. **However, I do this because this institution is dedicated to scientific research and I’m also a member of the community.** I wouldn’t have made the same decision if it was for another institution.” - P7*

RQ2. Factors Associated with People's Attitudes on Open Dataset Collection

- Privacy Risks



Routine Identification

“I visit my girlfriend’s house once or twice a week. Then someone might take a look at my data and think, ‘Oh, this guy always drops by this spot at this time of the week? What is this place?’”

“And if you combine this location data with my heart rate or sleep patterns around that time and if my heart rate skyrockets, you may wonder what I’m doing!” - P9

RQ2. Factors Associated with People's Attitudes on Open Dataset Collection

- Privacy Risks



Judgment & Categorization

“I didn’t want my call/text logs to be released ironically because there’s nothing. . . I only have a few friends and only a handful of contacts. . . So people would naturally assume that I’m a loner and anti-social. Even though you told me that the data are anonymized, I just hate it.” - P7

RQ2. Factors Associated with People's Attitudes on Open Dataset Collection

- Privacy Risks



Surveillance

“You see, I feel being monitored. I was upset with the key distance data. If I’m right, key distance equates with text message contents if you do the math. Then, don’t you think it’s easy to infer my personal life?” - P11

RQ2. Factors Associated with People's Attitudes on Open Dataset Collection

- Privacy Risks



Data Misuse & Leakage

“I’ve done similar research and once a data of a participant was leaked by mistake. Thank god it wasn’t very personal data, but still, things like that can happen!” - P5

RQ2. Factors Associated with People's Attitudes on Open Dataset Collection

- Lack of Justification



Research Purpose & Sensor Data Usage

“Why do you need all these data for emotional intelligence research? ... it doesn't tell us why you need that specific data and how you're going to use it for the analysis. It's kind of hard to associate some data with emotion-related research.” - P10

Summary and Discussion

Research Objectives

1. Understanding general motives (i.e., benefits) and level of privacy concerns in multimodal open dataset collection for affective computing
2. Exploring factors that affect participants' attitudes towards open dataset collection

Results

1. Financial motive was an important driver that facilitated the open dataset collection, but altruistic motives (i.e., scientific contribution) were also observed
2. Although participants generally outweighed benefits over risks, participants reported concerns regarding specific sensor data types centering around smartphone

Design Implications

Context-Aware Privacy Support for Mobile Open Dataset Collection (e.g., Dynamic consent)



Understanding Privacy Risks and Perceived Benefits in Open Dataset Collection for Mobile Affective Computing

Hyunsoo Lee, Soowon Kang, Uichin Lee

hslee90@kaist.ac.kr



2022UBICOMP

September 11 – 15
Atlanta, USA and Cambridge, UK