



# A New Hope: {Contextual Privacy Policies} for Mobile Applications and An Approach Toward Automated Generation

Shidong Pan<sup>1,2\*</sup>, Zhen Tao<sup>1,2</sup>, Thong Hoang<sup>2†</sup>, Dawen Zhang<sup>1,2</sup>, Tianshi Li<sup>3</sup>, Zhenchang Xing<sup>1,2</sup>, Xiwei Xu<sup>2</sup>, Mark Staples<sup>2</sup>, Thierry Rakotoarivelo<sup>2</sup>, and David Lo<sup>4</sup>

<sup>1</sup>*School of Computing, Australian National University*

<sup>2</sup>*Software and Computational Systems Research Program, CSIRO's Data61*

<sup>3</sup>*Khoury College of Computer Sciences, Northeastern University*

<sup>4</sup>*School of Computing and Information Systems, Singapore Management University*



USENIX Security 2024, 14-16 August 2024, Philadelphia PA USA

# Privacy issues are making headlines everyday!

## If You've Got a New Car, It's a Data Privacy Nightmare

Bad news: your car is a spy. E

By Thomas Germain Published Ye



## Norway court rules against Facebook owner Meta in privacy case



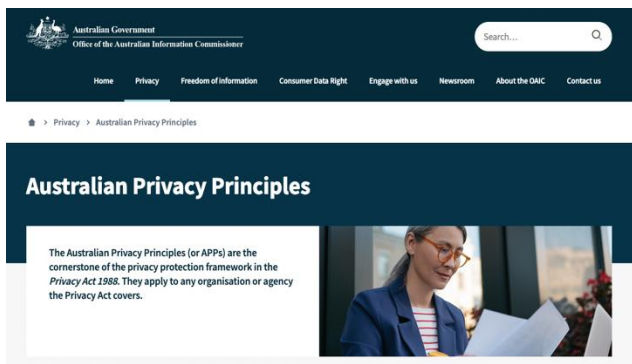
ion Sport Culture Lifestyle More v

as Asia Australia Middle East Africa Inequality Global development

### TikTok opens datacentre in Dublin in bid to combat European privacy concerns

The Chinese-owned app also announced a UK-based cybersecurity company will independently audit data controls and protections

# Privacy Policies are essential and critical



An APP entity must have a clearly expressed and up to date policy (the APP **privacy policy**) about the management of personal information by the entity [APP 1.3]

## [In re Google Assistant Privacy Litigation](#)

[457 F. Supp. 3d 797 - Dist. Court, ND California, 2020 - Google Scholar](#)

... with our **Privacy Policy** and other appropriate **confidentiality** and security ... **Privacy Policy** Litig., 58 F. Supp ... 2012) (finding no invasion of **privacy** based on Defendants' disclosure of each ...

[☆ Save](#) [🔗 Cite](#) [Cited by 64](#) [How cited](#) [All 2 versions](#)

## [In re Facebook, Inc. Internet Tracking Litigation](#)

[956 F. 3d 589 - Court of Appeals, 9th Circuit, 2020 - Google Scholar](#)

... First, the December 2010 **Privacy Policy** does not contain any agreement that Facebook would not track ... [11] Second, and more generally, the **Privacy** and Data Use **Policies** do not ...

[☆ Save](#) [🔗 Cite](#) [Cited by 267](#) [How cited](#) [All 2 versions](#)

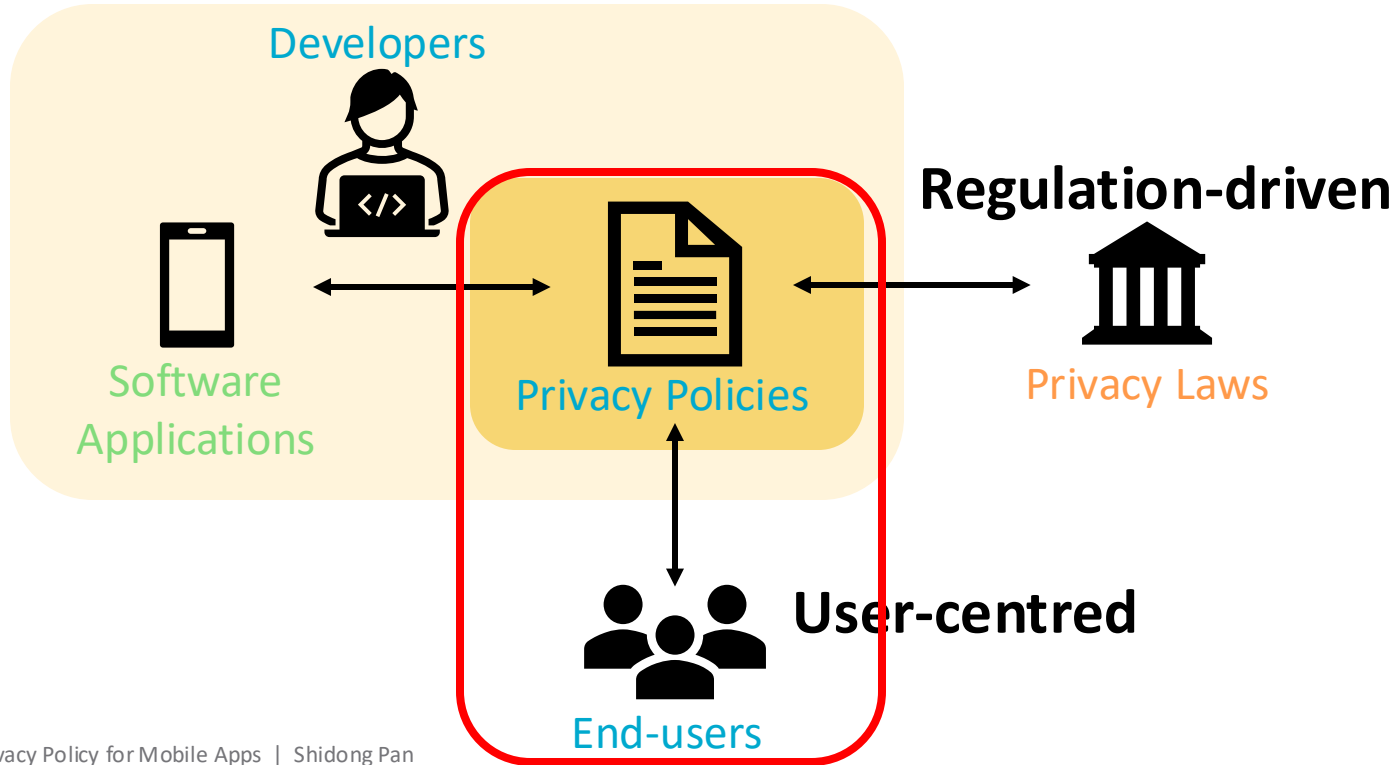
## [Kauders v. Uber Technologies, Inc.](#)

[486 Mass. 557, 159 NE 3d 1033 - Mass: Supreme Judicial Court, 2021 - Google Scholar](#)

... general principles of state contract law as **rules** of decision ... the link to the terms and conditions and the **privacy policy** ... question then becomes whether this type of **notice** was reasonable ...

[☆ Save](#) [🔗 Cite](#) [Cited by 54](#) [How cited](#) [All 2 versions](#)

# The privacy policy is the **pivot** of current software privacy ecosystem



# People do not read privacy policies!

- About **74% users don't read** privacy policies. For those who read it, the average reading time is 73 seconds [1].
- Thus, **user-centric** privacy policies and the usable privacy technology are pressingly needed.



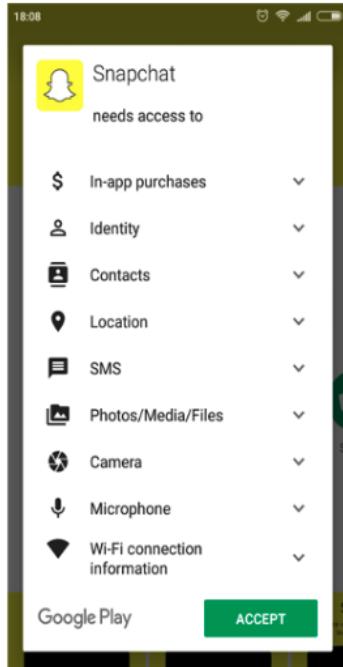
The Hobbit: An Unexpected Journey (2012)

[1] Jonathan A. Obar. 2016. The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services. <https://doi.org/10.2139/ssrn.2757465>

# Why do people not read PPs?

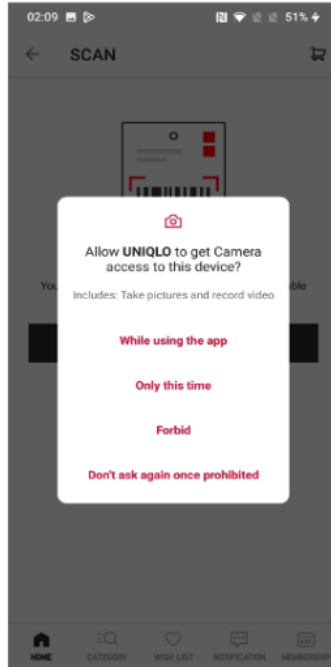
- **[Disconnection]** between privacy knowledge and context
  - [When reading the PP] What does it mean in the app?
  - [When using the app] What did they say in PP about my personal information?
- **[Poor readability]** of PPs
  - Privacy policies are very lengthy and detailed. The average length for popular app is about **4,000 words**.
  - The average readability score is 46/100 for 12 leading apps [2], readers should have at least a **college-level education background** to easily read the PPs.

# Development of “Just-in-time” Privacy Policies



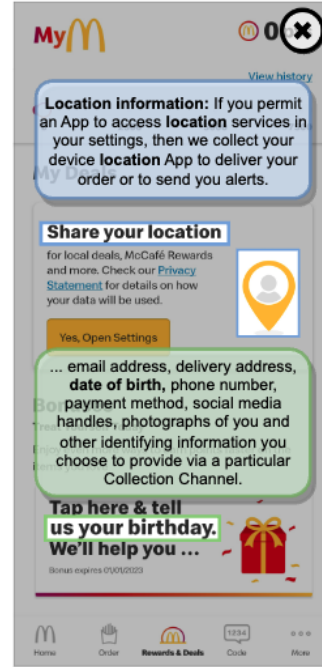
(a) Install-time

Before Android 6.0



(b) Invoke-time

Before Android 6.0



(c) Context-aware

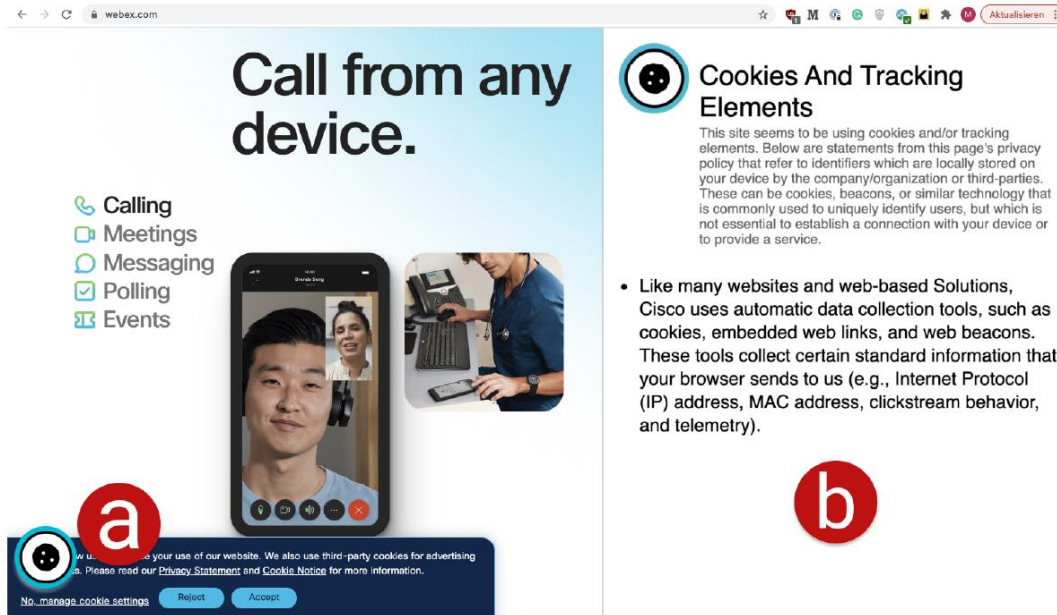
Ours

**Disconnection -> Connection**

**[When reading the PP]** What does it mean in the app?

**[When using the app]** What did they say in PP about my personal information?

# Readability Improvement by Contextualise PPs



## Context Privacy Policy for Websites

Windl, Maximiliane, et al. "Automating contextual privacy policies: Design and evaluation of a production tool for digital consumer privacy awareness." *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022.



# Contextualise Privacy Policy for Mobile Apps



Privacy Policies



Contextualise Lengthy PPs for mobile apps

# Contextual Privacy Policy for Mobile Apps

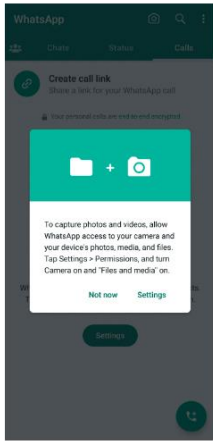


## Example 1: WhatsApp

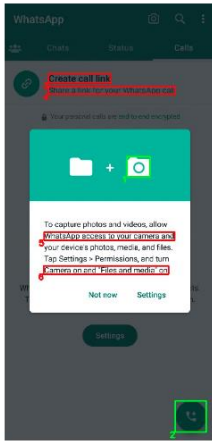
There are two inputs of SeePrivacy to generate CPP for a mobile app:

- 1: WhatsApp's [privacy policy](#) in HTML format;
2. The current GUI screenshot that you would like to generate CPP (the leftmost).

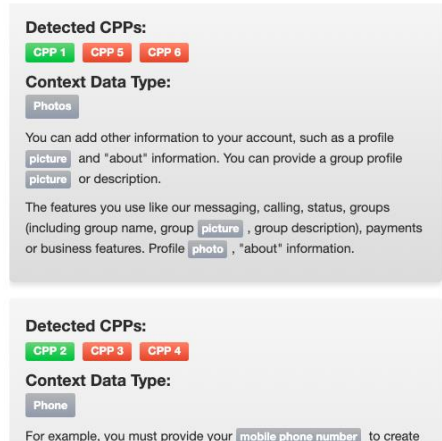
GUI screenshot 1.



Detected CPPs by SeePrivacy.

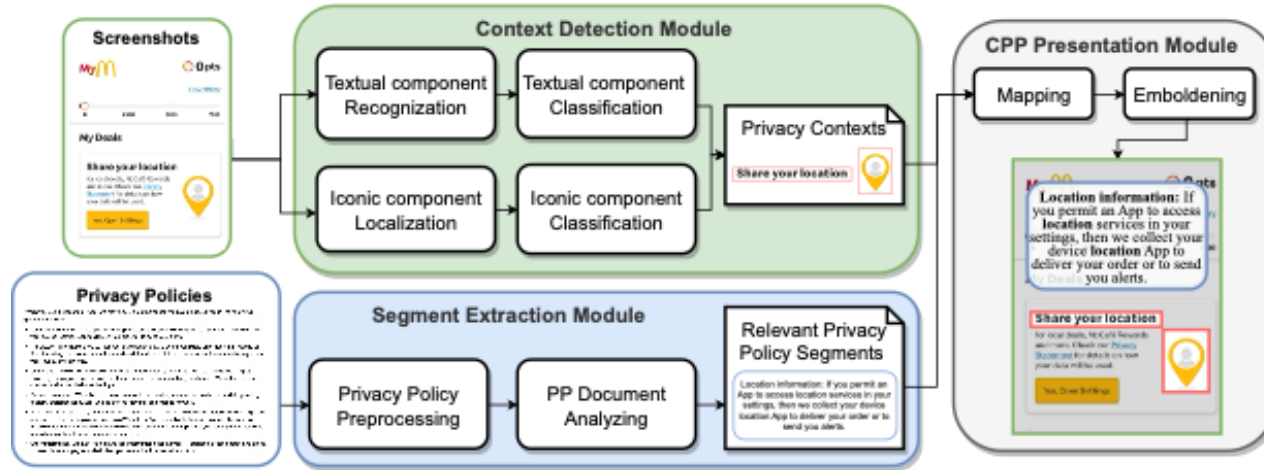


Retrieved privacy policy segments by SeePrivacy.



The aim of **Contextual Privacy Policy** is to fragment **privacy policies** into concise **snippets**, displaying them only within the corresponding contexts within the application's **graphical user interfaces** (GUIs).

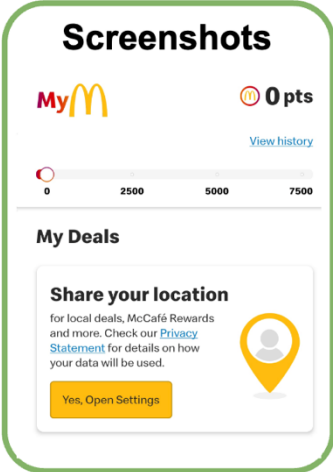
# Contextual Privacy Policy for Mobile Applications



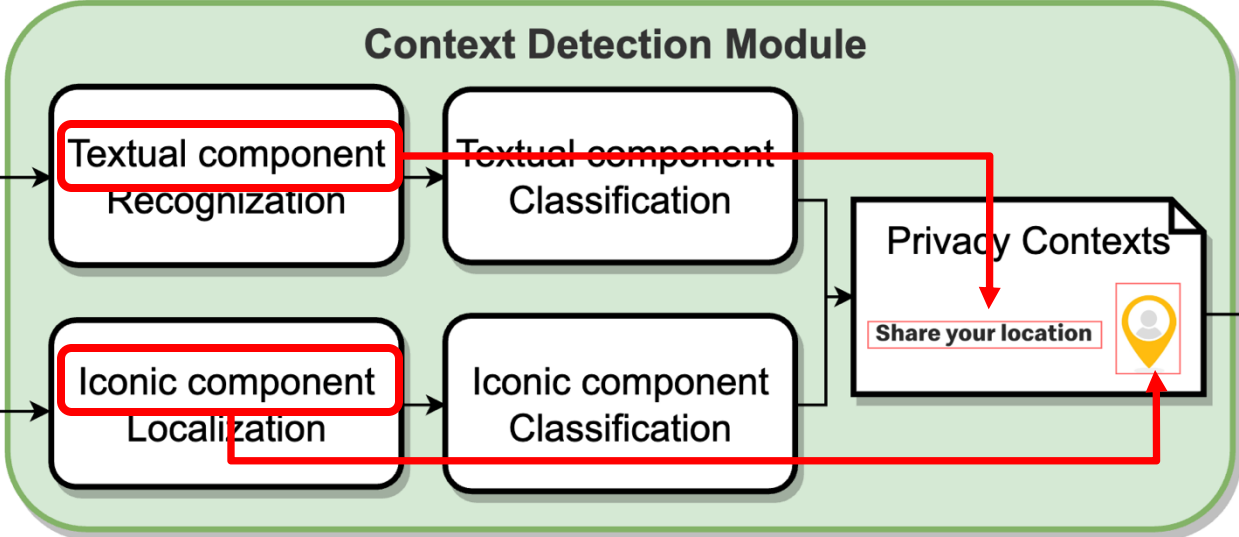
Our multi-modal framework synergistically combines Computer Vision (CV) techniques, Natural Language Processing (NLP) techniques, and pre-trained Large Language Model (LLMs).

- Contextual Detection Module (CV+LLM): Detecting privacy-related Textual and Iconic GUI components as contexts
- Segment Extraction Module (NLP): Extracting corresponding privacy-related policy segments

# Contextual Detection Module

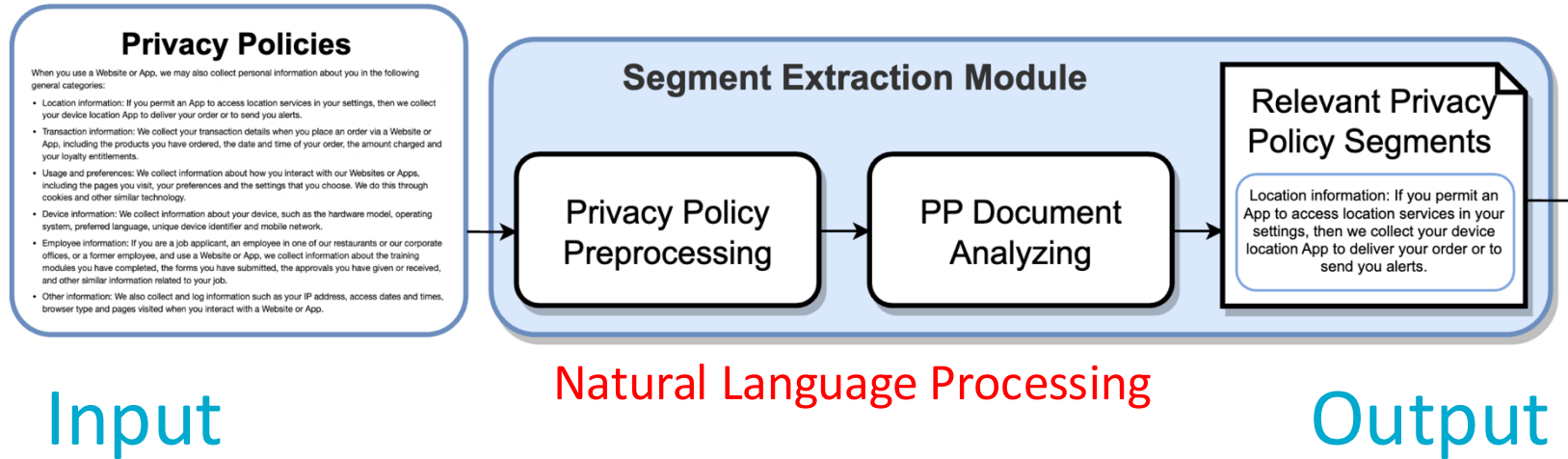


Input

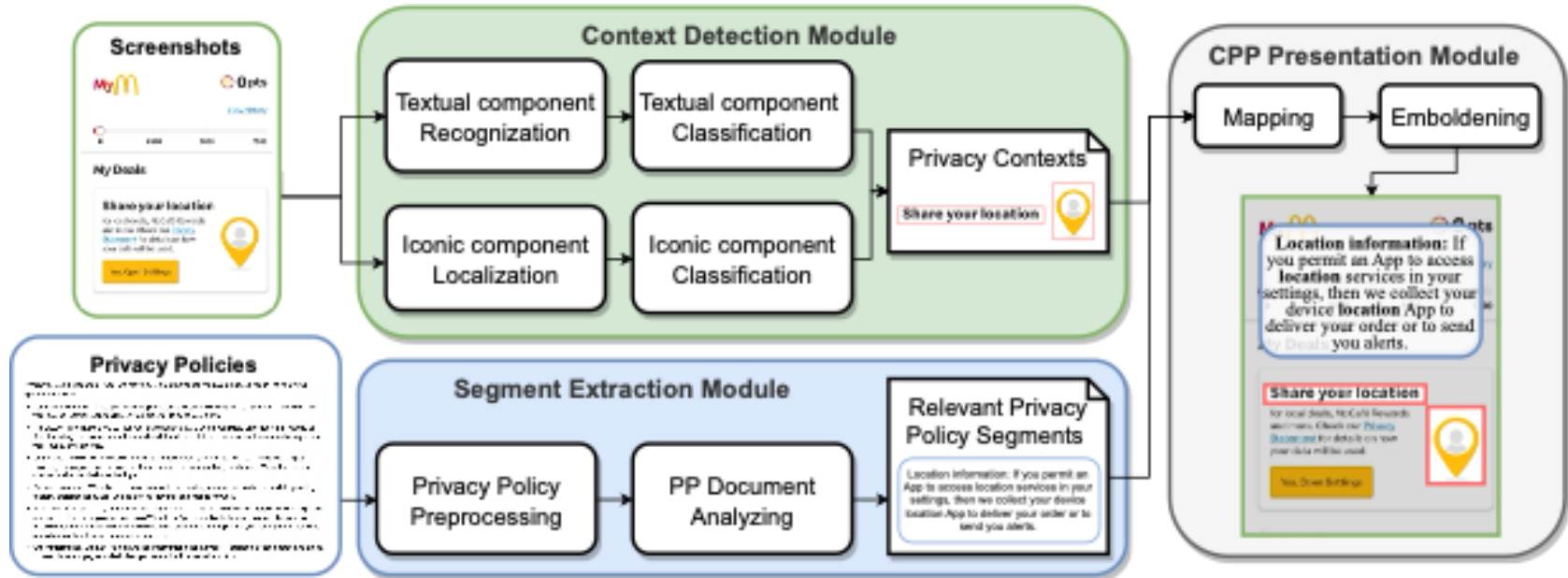


Output

# Contextual Privacy Policy for Mobile Applications



# Contextual Privacy Policy for Mobile Applications



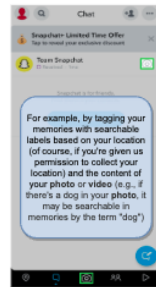
# Evaluation of the Framework: Dataset

1. Systematically gathering specific pages from each selected application.
2. Labelling all contexts on screenshots.
3. Annotating privacy policy segments for each context.

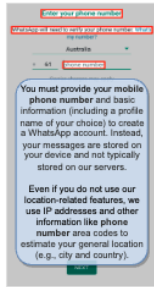
In total, our effort resulted in the collection of 402 screenshots from 50 diverse mobile apps, accompanied by **1,217** labelled privacy-related contexts, each correlating to specific segments within privacy policies.

Table 3: Basic statistic of CPP4APP.

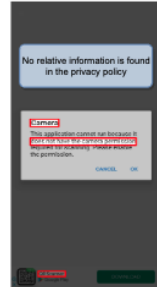
No. Mobile applications & Privacy policies	50
No. Screenshots	402
Screenshots per Mobile application	8.04
No. Privacy-related contexts	1,217
Contexts per Mobile application	24.34
No. Words in Privacy policies	297,010
Words per Privacy policy	5,940



(a) Screenshot 6-1



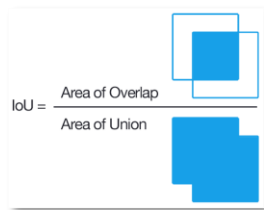
(b) Screenshot 11-9



(c) Screenshot 16-1

Figure 4: Three examples appeared in the human evaluation.

# Quantitatively Evaluation: SeePrivacy Effectiveness



$$segment\_sim(s_{ret}, s_{gt}) = \frac{1}{\text{MIN}(n, m)} \sum_{i=1}^n \sum_{j=1}^m \overline{lcs}(s_i^{ret}, s_j^{gt})$$

$$\overline{lcs}(s_i^{ret}, s_j^{gt}) = \frac{lcs(s_i^{ret}, s_j^{gt})}{\text{MIN}(\text{LEN}(s_i^{ret}), \text{LEN}(s_j^{gt}))}$$

## (d) Overall Context Identification

(a) CDM - Textual GUI components

Category	Accuracy	Precision	Recall
Name	0.98	0.98	1.00
Birthday	1.00	1.00	1.00
Address	0.38	0.41	0.86
Phone	0.87	0.90	0.96
Email	0.61	0.97	0.62
Profile	0.72	0.77	0.92
Contacts	0.98	1.00	0.98
Location	0.92	0.96	0.96
Photos	0.87	1.00	0.87
Voices	0.91	0.91	1.00
Financial info	0.93	0.93	0.93
Social media	0.67	0.75	1.00
<b>Average</b>	<b>0.82</b>	<b>0.87</b>	<b>0.93</b>

(b) CDM - Iconic GUI components

R-i Class	Category	Accuracy	Precision	Recall
Call	Phone	0.96	1.00	0.96
Email	Email	0.96	0.95	1.00
Avatar	Profile	0.92	0.96	0.96
Follow	Contacts	0.89	0.94	0.94
Group	Contacts	0.73	0.73	1.00
Location	Location	0.93	0.98	0.95
Crosshair	Location	0.91	0.94	0.97
Photo	Photos	0.92	0.92	1.00
Wallpaper	Photos	0.94	0.94	1.00
Videocam	Photos	0.83	1.00	0.83
Microphone	Voices	1.00	1.00	1.00
Cart	Financial Info	0.91	0.95	0.95
Facebook	Social media	0.96	1.00	1.00
Twitter	Social media	1.00	1.00	1.00
<b>Average</b>		<b>0.92</b>	<b>0.95</b>	<b>0.96</b>

(c) Segments Extraction Module

Category	Accuracy	Precision	Recall
Name	1.00	1.00	1.00
Birthday	0.96	1.00	0.95
Address	0.52	0.82	0.50
Phone	1.00	1.00	1.00
Email	0.99	0.99	1.00
Profile	0.99	0.99	1.00
Contacts	0.79	1.00	0.78
Location	0.94	0.95	0.98
Photos	0.95	0.97	0.97
Voices	0.86	0.76	1.00
Financial info	1.00	1.00	1.00
Social media	1.00	1.00	1.00
<b>Average</b>	<b>0.94</b>	<b>0.98</b>	<b>0.96</b>

Category	Accuracy	Precision	Recall
Name	0.98	0.98	1.00
Birthday	1.00	1.00	1.00
Address	0.38	0.41	0.86
Phone	0.88	0.91	0.97
Email	0.66	0.96	0.68
Profile	0.70	0.74	0.93
Contacts	0.75	0.94	0.78
Location	0.94	0.97	0.96
Photos	0.86	0.98	0.88
Voices	0.91	0.91	1.00
Financial info	0.81	0.93	0.87
Social media	0.80	0.89	0.89
<b>Average</b>	<b>0.81</b>	<b>0.88</b>	<b>0.90</b>

## Performance



# Human Evaluation

**[Connection]** between privacy knowledge and context

The data type matches with the detected context, privacy policy segment. A human evaluation shows that **77%** of the extracted privacy policy segments were perceived as well-aligned with the detected contexts.

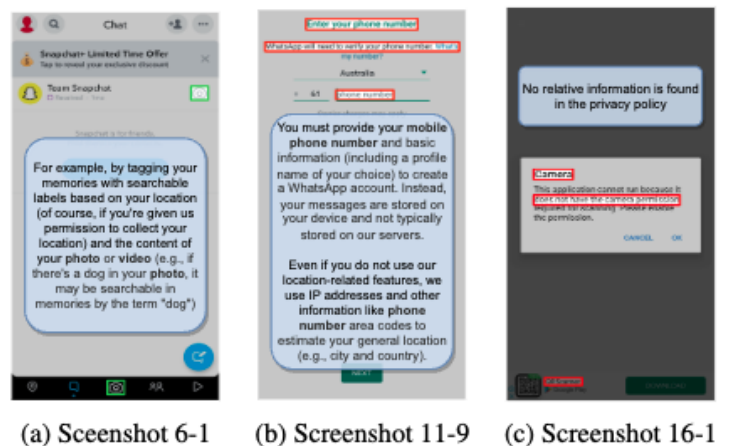


Figure 4: Three examples appeared in the human evaluation.

Topic	Mean	Median	SD
Data type & Context	4.22	5	1.14
Data type & Policy segment	3.95	4	1.28
Policy segment & Context	3.75	4	1.40

# Human Evaluation

## [Readability] of Privacy Policies

1. 93% participants evaluated it as either "Very useful" or "Useful."
2. A substantial increase in the willingness of participants to engage with CPPs, as reflected by a mean score that surged from 2 to 4.1.

No.	Pre- and post-questions	Scale (5-Likert)	Mean	Median	SD
Q <sub>1</sub>	How concerned are you about your privacy information while using mobile apps?	5 for very concerned, 1 for very unconcerned	4.13	4	1.09
Q <sub>2</sub>	Do you read mobile app's privacy policies when you encountered?	5 for always read, 1 for never read	2.00	2	1.03
Q <sub>3</sub>	What do you think the usefulness of this tool in terms of providing privacy information for mobile apps?	5 for very useful, 1 for very useless	4.53	5	0.62
Q <sub>4</sub>	Will you read contextual privacy policies when you encountered in future?	5 for always read, 1 for never read	4.07	4	0.85

# Limitations

## 1. Efficiency:

- A complete evaluation of its efficiency in practice is needed

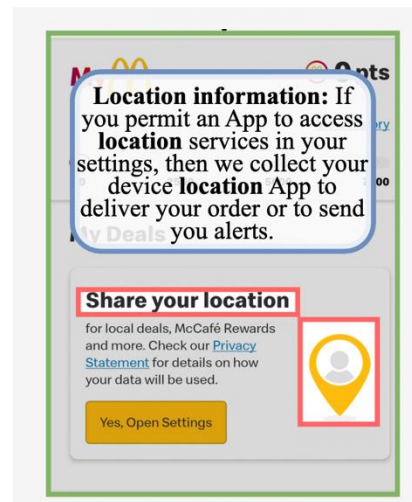
## 2. Evaluation:

- Limited user study size
- A more sophisticated user study design is expected to better demonstrate the usefulness of SeePrivacy

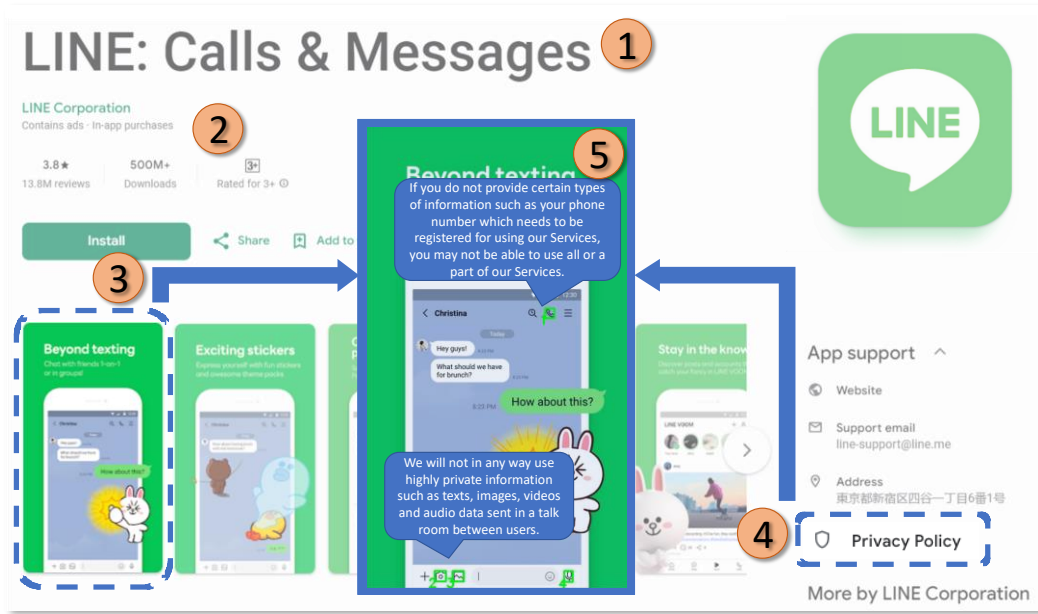
## 3. UX/UI:

- Privacy-utility Trade-off

Model	Acc.	Prec.	Rec.	Size(MB)	Time(s)
kNN [38]	0.75	0.61	0.52	0	6.78
Liu et al. [73]	0.90	<b>0.90</b>	0.79	293	5.02
MobileNetV2 [88]	0.86	0.73	0.73	9	0.56
MobileNetV3 [55]	0.79	0.62	0.66	6	0.52
Ours (ViT-based)	<b>0.97</b>	<b>0.90</b>	<b>0.90</b>	328	11.28
Ours (ResNet-based)	0.91	0.80	0.81	43	<b>0.37</b>



# Adoption Scenario: CPP in Market



A plug-and-play application is displaying CPP in the app market.

Users can gain contextualized insights into potential data practices and pertinent policy statements of the mobile application **prior to installation**, enabling them to make informative decisions and establish better expectations **before downloading the app**.

# Implications and Broader Impacts


SeePrivacy About Examples Demo PET Contact

## What is SeePrivacy?





Privacy policies have become the most critical approach to safeguarding individuals' privacy and digital security. To enhance their presentation and readability, the concept of **Contextual Privacy Policies (CPPs)** was gradually developed, aiming to fragment policies into shorter snippets and display them only in corresponding contexts.

We are the first to propose a novel multi-modal framework, namely **SeePrivacy**, designed to automatically generate contextual privacy policies for mobile apps.

Our framework does **not** require the access to apps' source code or Android APIs; hence, the framework can be easily deployed with lower security concerns.



## What can SeePrivacy bring to you?

			
<b>Privacy notice</b> SeePrivacy aims to protect personal	<b>Just-in-time reminder</b> Privacy notices are closely and timely	<b>Readability</b> SeePrivacy enhances the presentation	<b>Comprehensibility</b> SeePrivacy effectively assists users in

These findings suggest that our framework could serve as a significant tool for **bolstering user interaction with, and understanding of, privacy policies**. Furthermore, our solution has the potential to make privacy notices more **accessible** and **inclusive**, thus appealing to a broader demographic.

Showcase website: <https://cpp4app.github.io/>

Live demo: <https://huggingface.co/Cpp4App/>

# Shidong Pan

## I'm on the faculty job market.



Scan for my CV

Two(2) papers USENIX Security 2024

Mobile Privacy

Usable Privacy & Security + Software Engineering + RAI

International Student Program Coordination Experience