{Privacy Policy} in Practice: Challenges, Implications, and Solutions

Shidong Pan

University of Edinburgh

March 2024
Self-introduction

• Shidong Pan

• Bachelor of Advanced Computing @ ANU + Shandong University
• Master of Machine Learning & Computer Vision @ ANU

• Currently, I’m an about-to-finish PhD affiliated to:
  • School of Computing, ANU
  • Responsible AI Group, CSIRO’s Data61

• Was a visiting PhD student at Singapore Management University (SMU).
Greetings from “downstairs” : )
Outline

1. Background and Motivation

   • Content Analysis (PP Descriptions - Software Behaviours)
   • Compliance Checking (PP Descriptions - Law/Reg Requirements)
   • Transparency and Readability of PP and Privacy Notices

3. My Research Projects
   • An empirical study of Online Automated Privacy Policy Generators
   • Contextual Privacy Policy for Mobile applications
Privacy issues are making headlines everyday!
According UNCTD, 137 out of 194 countries had put in place legislation to secure the protection of data and privacy.
The Privacy Policy is 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]
Examples of privacy policies

Facebook/Meta
> 20,000 words
People do not read privacy policies!

- Privacy policies are very lengthy and detailed. The average length for popular app is about **4,000 words**.
- About **74% users don’t read** privacy policy. For those who read it, the average reading time is 73 seconds [1].

Thus, **user-centric** privacy notice and the **usable privacy** technology are pressingly needed.

Privacy Policies are commonly problematic!
Privacy Policy Research Landscape

1. Content Analysis (PP Descriptions - Software Behaviours):

   | Privacy Policies | Software Applications |

2. Compliance Checking (PP Descriptions - Law/Reg Requirements)

   | Privacy Policies | Privacy Laws & Regulations |

3. Transparency and Readability of PP and Privacy Notices:

   | Privacy Policies | Users |
Privacy Policy Research Landscape

1. Content Analysis (PP Descriptions - Software Behaviours):

- Privacy Policies
- Software Applications

2. Compliance Checking (PP Descriptions - Law/Reg Requirements)

3. Transparency and Readability of PP and Privacy Notices:
A Typical Framework of PP Content Analysis

Software Analysis

App Store → App → AppCensus → Data Flows → PoliCheck

Privacy Policy Processing & Analysis

Privacy Policy → PolicyLint → Collection Statements

Flow-to-Policy Consistency

- Consistent
  - Clear Disclosure
  - Vague Disclosure
- Inconsistent
  - Omitted Disclosure
  - Incorrect Disclosure
  - Ambiguous Disclosure

Figure 1: POLICHECK determines the consistency of a mobile application’s data flows to its privacy policy.

PolicyLint, Andow et al.,USENIX Security 2019
<table>
<thead>
<tr>
<th>Year</th>
<th>Name</th>
<th>Method</th>
<th>Authors</th>
<th>Venue</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>Polisis</td>
<td>Deep Learning (CNN + DenseNet)</td>
<td>Harkous et al.</td>
<td>Security</td>
</tr>
<tr>
<td>2019</td>
<td>PolicyLint</td>
<td>Sentence-level NLP (NER, DND, etc.)</td>
<td>Andow et al.</td>
<td>Security</td>
</tr>
<tr>
<td>2021</td>
<td>PurPliance</td>
<td>Rule-based matching (NER, Pattern Detect)</td>
<td>Bui et al.</td>
<td>CCS</td>
</tr>
<tr>
<td>2023</td>
<td>PoliGraph</td>
<td>Rule-based matching (KG)</td>
<td>Cui et al.</td>
<td>Security</td>
</tr>
</tbody>
</table>
A Taxonomy of Privacy Policy Content

Fig. 3: The privacy taxonomy of Wilson et al. [11]. The top level of the hierarchy (shaded blocks) defines high-level privacy categories. The lower level defines a set of privacy attributes, each assuming a set of values. We show examples of values for some of the attributes.

Table 3: Classification Precision/Recall/F1 (respectively abbreviated as P/R/F) for every single category, and their Macro Average of OPP-115 by Polnics [3], LR, SVM and HMM [6].

| Label                  | Polnics P | Polnics R | Polnics F | LR P | LR R | LR F | SVM P | SVM R | SVM F | HMM P | HMM R | HMM F |
|------------------------|-----------|-----------|-----------|------|------|------|-------|-------|-------|-------|-------|-------|-------|
| 1st Party Collection   | 0.79      | 0.70      | 0.73      | 0.70 | 0.70 | 0.70 | 0.75  | 0.70  | 0.70  | 0.69  | 0.70  | 0.70  |
| 3rd Party Sharing      | 0.73      | 0.70      | 0.73      | 0.63 | 0.67 | 0.70 | 0.53  | 0.61  | 0.61  | 0.53  | 0.61  | 0.61  |
| User Choice/Control    | 0.74      | 0.74      | 0.74      | 0.45 | 0.52 | 0.52 | 0.35  | 0.47  | 0.35  | 0.47  | 0.35  | 0.47  |
| Access, Edit, Deletion | 0.89      | 0.80      | 0.80      | 0.47 | 0.71 | 0.47 | 0.71  | 0.61  | 0.47  | 0.71  | 0.61  | 0.47  |
| Data Retention         | 0.83      | 0.66      | 0.73      | 0.35 | 0.16 | 0.12 | 0.12  | 0.08  | 0.12  | 0.12  | 0.08  | 0.12  |
| Data Security          | 0.88      | 0.83      | 0.85      | 0.59 | 0.66 | 0.67 | 0.67  | 0.67  | 0.67  | 0.67  | 0.67  | 0.67  |
| Policy Change          | 0.95      | 0.84      | 0.85      | 0.59 | 0.66 | 0.75 | 0.52  | 0.68  | 0.52  | 0.68  | 0.52  | 0.68  |
| Do Not Track           | 0.94      | 0.97      | 0.96      | 0.45 | 1.0  | 1.0  | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   |
| Specific Audiences     | 0.90      | 0.94      | 0.95      | 0.49 | 0.69 | 0.57 | 0.70  | 0.67  | 0.60  | 0.66  | 0.62  | 0.66  |
| Macro Average          | 0.85      | 0.79      | 0.84      | 0.49 | 0.69 | 0.56 | 0.66  | 0.66  | 0.50  | 0.50  | 0.50  | 0.50  |

Table 1: Classification Precision/Recall/F1 (respectively abbreviated as P/R/F) for every single category, their Macro Average, and the total Accuracy of OPP-115 by ChatGPT, GPT4 and Claude2.

<table>
<thead>
<tr>
<th>Label</th>
<th>ChatGPT P</th>
<th>ChatGPT R</th>
<th>ChatGPT F</th>
<th>GPT4 P</th>
<th>GPT4 R</th>
<th>GPT4 F</th>
<th>Claude2 P</th>
<th>Claude2 R</th>
<th>Claude2 F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Party Collection</td>
<td>0.89</td>
<td>0.90</td>
<td>0.92</td>
<td>0.89</td>
<td>0.87</td>
<td>0.90</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>3rd Party Sharing</td>
<td>0.92</td>
<td>0.90</td>
<td>0.91</td>
<td>0.89</td>
<td>0.87</td>
<td>0.89</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>User Choice/Control</td>
<td>0.92</td>
<td>0.90</td>
<td>0.91</td>
<td>0.89</td>
<td>0.87</td>
<td>0.89</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>Access, Edit, Deletion</td>
<td>0.89</td>
<td>0.94</td>
<td>0.91</td>
<td>0.89</td>
<td>0.87</td>
<td>0.89</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>Data Retention</td>
<td>0.82</td>
<td>0.68</td>
<td>0.75</td>
<td>0.82</td>
<td>0.68</td>
<td>0.75</td>
<td>0.82</td>
<td>0.68</td>
<td>0.75</td>
</tr>
<tr>
<td>Data Security</td>
<td>0.79</td>
<td>0.68</td>
<td>0.75</td>
<td>0.82</td>
<td>0.68</td>
<td>0.75</td>
<td>0.82</td>
<td>0.68</td>
<td>0.75</td>
</tr>
<tr>
<td>Policy Change</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>Do Not Track</td>
<td>0.91</td>
<td>1.00</td>
<td>1.00</td>
<td>0.91</td>
<td>1.00</td>
<td>1.00</td>
<td>0.91</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Specific Audiences</td>
<td>1.00</td>
<td>0.92</td>
<td>1.00</td>
<td>1.00</td>
<td>0.92</td>
<td>1.00</td>
<td>1.00</td>
<td>0.92</td>
<td>1.00</td>
</tr>
<tr>
<td>Macro Average</td>
<td>0.82</td>
<td>0.95</td>
<td>0.93</td>
<td>0.95</td>
<td>0.92</td>
<td>0.95</td>
<td>0.95</td>
<td>0.92</td>
<td>0.95</td>
</tr>
<tr>
<td>Year</td>
<td>Name</td>
<td>Method/Intro</td>
<td>Authors</td>
<td>Venue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----------------</td>
<td>--------------------------------------------------</td>
<td>---------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td>PoliCheck</td>
<td>Android apps entity-sensitive policy and data-flow Analysis</td>
<td>Andow et al.</td>
<td>Security</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2021</td>
<td>PrivacyFlash Pro</td>
<td>iOS apps data-flow to disclosure</td>
<td>Zimmeck et al</td>
<td>NDSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2023</td>
<td>Lalaine</td>
<td>iOS apps data-flow to privacy-label</td>
<td>Xiao et al.</td>
<td>Security</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2024</td>
<td>Matcha</td>
<td>Android app IDE In-IDE data/permission usage to privacy-label</td>
<td>Li et al. 2024</td>
<td>IMWUT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Content Analysis tools

PrivacyFlash Pro: Automating Privacy Policy Generation for Mobile Apps, Zimmeck et al., NDSS 2021
1. With the development of deep learning-based NLP (e.g., Transformers, GPTs), more researchers choose to embrace rule-based/pattern matching NLP methods.

2. Accurate mobile app behaviour extraction (Static Analysis + Dynamic Analysis) is challenging!

3. Current studies commonly focus on data entities alignment, neglecting the fine-grained data purposes.
Privacy Policy Research Landscape

1. Content Analysis (PP Descriptions - Software Behaviours):

   - Privacy Policies
   - Software Applications

2. Compliance Checking (PP Descriptions - Law/Reg Requirements)

   - Privacy Policies
   - Privacy Laws & Regulations

3. Transparency and Readability of PP and Privacy Notices:
### CCPA Requirement

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Indentra</th>
<th>Jun 21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disclosure of right to request how personal information is collected, used, sold, disclosed for a business purpose, and shared [CCPA §1798.105(a)(5)(A), 1798.110(a), 1798.115(a), Regs §999.306(c)(1)(a)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Disclosure of right to request deletion of personal information [CCPA §1798.105(b), 1798.130(a)(5)(A), Regs §999.306(c)(2)(a)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Disclosure of whether personal information is sold and right to opt-out of sale [Regs §999.306(c)(1)(b), 999.305(c)(b), 999.306(c)(2)(b)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Disclosure of right to not be discriminated against when requesting any rights [CCPA §1798.130(a)(5)(A), 1798.125(a), Regs §999.306(c)(3)(c)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Instructions for submitting requests and link to online form or portal if offered [Regs §999.306(c)(1)(b), 999.305(c)(b), 999.306(c)(2)(b)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Instructions for authorized agents to make requests [Regs §999.306(c)(3)(a)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Description of the process used to verify requests [Regs §999.306(c)(3)(c)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>List of categories of personal information collected in preceding 12 months [CCPA §1798.130(a)(5)(B), 1798.115(b), Regs §999.306(c)(5)(c)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>List of categories of personal information sold in preceding 12 months [CCPA §1798.130(a)(5)(C), 1798.115(c)(2), Regs §999.306(c)(3)(c)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>List of categories of personal information disclosed for business purpose in preceding 12 months [CCPA §1798.130(a)(5)(C), 1798.115(c)(2), Regs §999.306(c)(3)(c)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>For each personal information category, categories of third parties to whom information was disclosed or sold [Regs §999.306(c)(1)(c)(2)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Categories of sources from which personal information is collected [Regs §999.306(c)(1)(b)(3)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Business or commercial purpose for collecting or selling personal information [Regs §999.306(c)(1)(b)(3)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Whether the business has actual knowledge that it sells personal information of minors under 16 years of age and special classes of minors [Regs §999.306(c)(1)(b)(3), 999.305(c)(1)(b)(3)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Contact information for questions or concerns [Regs §999.306(c)(6)(a)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Date policy was last updated [Regs §999.306(c)(7)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Special requirements for businesses buying, receiving, selling, or sharing personal information of 10,000,000 or more consumers in a calendar year [Regs §999.306(c)(8), 999.317(a)(2)(d)]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>For online notices, follow generally recognized industry standards, such as the W3C Web Content Accessibility Guidelines, version 2.1 of June 5, 2018 [Regs §999.306(a)(2)(d)]</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### Table II: CCPA privacy policy requirements and generators’ compliance

1. Collect Personal Information: Collect data subjects’ information which can identify their personal IDs. [GDPR Art. 13.1]  
2. Data Retention Period: Retention period of personal information. [GDPR Art. 13.2(a)]  
3. Data Processing Purposes: The purposes of processing personal data. [GDPR Art. 13.1(c)(2)]  
4. Contact Details: The contact details of the controller or the Data Protection Officer. [GDPR Art. 13.1(a)(b)]  
5. Right to Access: The right of the data subject to request from the controller to access their personal information. [GDPR Art. 13.2(b)]  
6. Right to Rectify or Erase: The right of the data subject to request from the controller to rectify or erase of their personal information. [GDPR Art. 13.2(b)]  
7. Right to Restrict Processing: The right of the data subject to request from the controller to restrict processing concerning the data subject. [GDPR Art. 13.2(b)]  
8. Right to Object to Processing: The right of the data subject to object to processing. [GDPR Art. 13.2(b)]  
9. Right to Data Portability: The right of the data subject to receive and transmit his/her personal data to another controller. [GDPR Art. 13.2(b)]  
10. Right to Lodge a Complaint: The right of the data subject to lodge a complaint with a supervisory authority. [GDPR Art. 13.2(b)]  

PrivacyFlash Pro: Automating Privacy Policy Generation for Mobile Apps, Zimmeck et al., NDSS 2021

Have You been Properly Notified? Automatic Compliance Analysis of Privacy Policy Text with GDPR Article 13, Liu et al., WWW 2021
Privacy Policy Research Landscape

1. Content Analysis (PP Descriptions - Software Behaviours):

   Privacy Policies <-> Software Applications

2. Compliance Checking (PP Descriptions - Law/Reg Requirements)

   Privacy Policies <-> Privacy Laws & Regulations

3. Transparency and Readability of PP and Privacy Notices:

   Privacy Policies <-> Users
Motivation: transparent and readability of PPs

1. Requirements in laws/regulations
   1. EU General Data Protection Regulations (GDPR)
   2. California Consumer Privacy Act of 2018 (CCPA)
   3. Australian Privacy Principles (APP): have *a clearly expressed* and up-to-date APP Privacy Policy about how the entity manages personal information.

2. Users’ and consumers’ practical need
   According to a survey conducted by The Washington Times, 36% of interviewees *never* read privacy policies, and 38% of interviewees sometimes read privacy policies.
Development History

1. Privacy Policy

2. Privacy Icons

3. Platform for Privacy Preference

4. Privacy Labels

5. Contextual Privacy Policy
Privacy Policy Research Landscape: My research

1. Content Analysis (PP Descriptions - Software Behaviours):
   - Privacy Policies
   - Software Applications

2. Compliance Checking (PP Descriptions - Law/Reg Requirements)
   - Privacy Policies
   - Privacy Laws & Regulations

3. Transparency and Readability of PP and Privacy Notices:
   - Privacy Policies
   - Users

---

An Empirical Study of Automated {Privacy Policy Generators}
[Topic]: Content Analysis + Compliance Analysis
[Venue]: USNEIX Security 2024

A NEW HOPE: Contextual Privacy Policies for Mobile Applications and An Approach Toward Automated Generation

---

Users
How do those problematic privacy policies be crafted?

To develop privacy policies, developers may **copy-paste-modify** existing privacy policies, ad-hoc.

Developers → Mobile applications → Privacy policies

Most (citizen) developers do not have legal support!
Online Automated Privacy Policy Generators can provide more automated and systematic solutions for developers, rather than through ad-hoc copy-paste-modify.

However, their quality and other characteristics can vary and are not yet deeply understood.
The Prevalence of APPGs in Market

The market occupancy ratio of 10 examined APPGs is around 20.1%!

Table 4: Summary of market use of different APPGs.

<table>
<thead>
<tr>
<th>Method</th>
<th>Market Occupancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fingerprint Keyword Searching</td>
<td>6.6% (3,066)</td>
</tr>
<tr>
<td>Document Similarity Comparison</td>
<td>18.1% (8,425)</td>
</tr>
<tr>
<td>Intersection</td>
<td>4.4% (2,042)</td>
</tr>
<tr>
<td>Union (Total)</td>
<td><strong>20.1% (9,332)</strong></td>
</tr>
</tbody>
</table>
The Compliance of APPGs against Regulations

The #2 App Privacy Policy Generator is the most popular one, boasting a 72.7% adoption rate. Users tend to select easy-to-use APPGs even though at the cost of a potentially-higher risk to breach privacy regulations.
Implications and Findings to Stakeholders

- **App developers/APPG users:** While app developers may benefit from using APPGs to create privacy policies more efficiently, they should be aware of APPGs' latent limitations.

- **APPG providers:** Our analysis suggests APPG providers should work on improving recognised data use, since the majority of APPGs on the market only provide a very limited scope of personal information and device permissions.

- **Privacy regulators:** Regulators should recognize the importance of this issue and be engage with this emerging market trend.
Privacy Policy Research Landscape: My research

1. Content Analysis (PP Descriptions - Software Behaviours):
   - Privacy Policies
   - Software Applications

2. Compliance Checking (PP Descriptions - Law/Reg Requirements)
   - Privacy Policies
   - Privacy Laws & Regulations

3. Transparency and Readability of PP and Privacy Notices:
   - Privacy Policies
   - Users

---

An Empirical Study of Automated {Privacy Policy Generators}

[Topic]: Content Analysis + Compliance Analysis

[Venue]: USNEIX Security 2024

Is It a Trap? A Large-scale Empirical Study And Comprehensive Assessment of Online Automated Privacy Policy Generators for Mobile Apps

Shisheng Tan, Dawei Zhang, Zhou Tao, Qi Li, Xuedi Huang, Xiaolei Sun, and David Lo

3. Transparency and Readability of Mobile Apps

[Topic]: Transparency and Readability of Privacy Policies

[Venue]: USNEIX Security 2024

A NEW HOPE: Contextual Privacy Policies for Mobile Applications and An Approach Toward Automated Generation

Shisheng Tan1, Zhou Tao2, Qi Li3, Xuedi Huang3, Xiaolei Sun3, Mark Staples4, Thierry Jelonek5, and David Lo6

1School of Computing, Australian National University
2Software and Computational Systems Research Program, CSIRO’s Data61
3Rutgers College of Computer Sciences, Northeastern University
4School of Computing and Information Systems, Singapore Management University
5School of Data Science, Curtin University
6School of Computing, Curtin University
Development of “Just-in-time” Privacy Notices

(a) Install-time
(b) Invoke-time
(c) Context-aware

Android 6.0
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).
Our multi-modal framework synergistically combines Computer Vision (CV) techniques, pre-trained Large Language Model (LLMs), and Natural Language Processing (NLP) techniques.
An Adoption Scenario: CPP in Market

We will not in any way use highly private information such as texts, images, videos and audio data sent in a talk room between users.

If you do not provide certain types of information such as your phone number which needs to be registered for using our Services, you may not be able to use all or a part of our Services.
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/
User-Centric Privacy Enhancing Toolkit (U-PET)

Users

Privacy Information

✅

❌

Transparent and Engaging Privacy Information

U-PET

Users
1. Automated Privacy Policy Generators (APPGs)

2. Privacy Nutrition Labels From Privacy Policies (Policy2Label)

3. Contextual Privacy Policy for Mobile Apps (Cpp4App)

4. Data Rights Extraction and Automation in Mobile Apps (AutoYourRight)
{Privacy Policy} in Practice: Challenges, Implications, and Solutions

Hope you enjoy the content : )

Shidong.Pan@anu.edu.au
University of Edinburgh
March 2024