Polisis

Yanyan Ren
Current problem

1. What’s your experience with privacy policy?
2. Do you read it? Which parts?

To Read New GDPR Privacy Policies You'll Need a Football Field
Past solutions

- Machine-readable formats
- Nutrition labels and privacy icons
- Crowdsourcing short notices extraction

What are the problems with these solutions (claimed by the author)? Do you agree?

Polisis

- Scalable, Dynamic, Multi-dimensional
- Two parts: structured and unstructured queries
- Useful for users, researchers and regulator

1. How did you like Polisis?
2. Would you use it?
Key concepts

- Segments
- Word embeddings
- Two-level annotations
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.
Main framework

1. Data layer
   - Scrapped top 200 websites
   - Semantically coherent segments

2. ML layer
   - Two stages
   - Hierarchical multi-label classification

3. Application layer
   - Supports both structured and unstructured queries

Questions:

1. Does the policy need to be in a certain format? (length, language, etc.)
2. How well would the model adapt to policy changes? New privacy icons?

1. Why did they use CNN (usually used for image classification)?
2. How did they evaluate the model? What's the performance?

What's the consequence of getting a wrong answer?
Product critique

- Polisis: How did it decide which part of text is important?
- Pribot: What kind of questions got handled well? What didn’t? Why?
- What can you do as an evil company?
# Future directions

1. Do you believe Polisis is the right way to solve the problem?
2. What's the future direction for Polisis research?

<table>
<thead>
<tr>
<th></th>
<th>Benefits</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Natural language</strong></td>
<td>Legal value</td>
<td>Ambiguity, understandability, auditability</td>
</tr>
<tr>
<td><strong>Graphical (icons)</strong></td>
<td>Lay-user understandability, Measure understanding, legal design</td>
<td>Ambiguity, incompleteness, claim over legal compliance</td>
</tr>
<tr>
<td><strong>Machine-readable</strong></td>
<td>Automatically enforced, auditability, correctness</td>
<td>Human understandability, lack of adoption</td>
</tr>
</tbody>
</table>

Source: Three Dimensions of Privacy Policies, Morel & Pardo