Polisis

Yanyan Ren

Current problem

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

400 PAGES
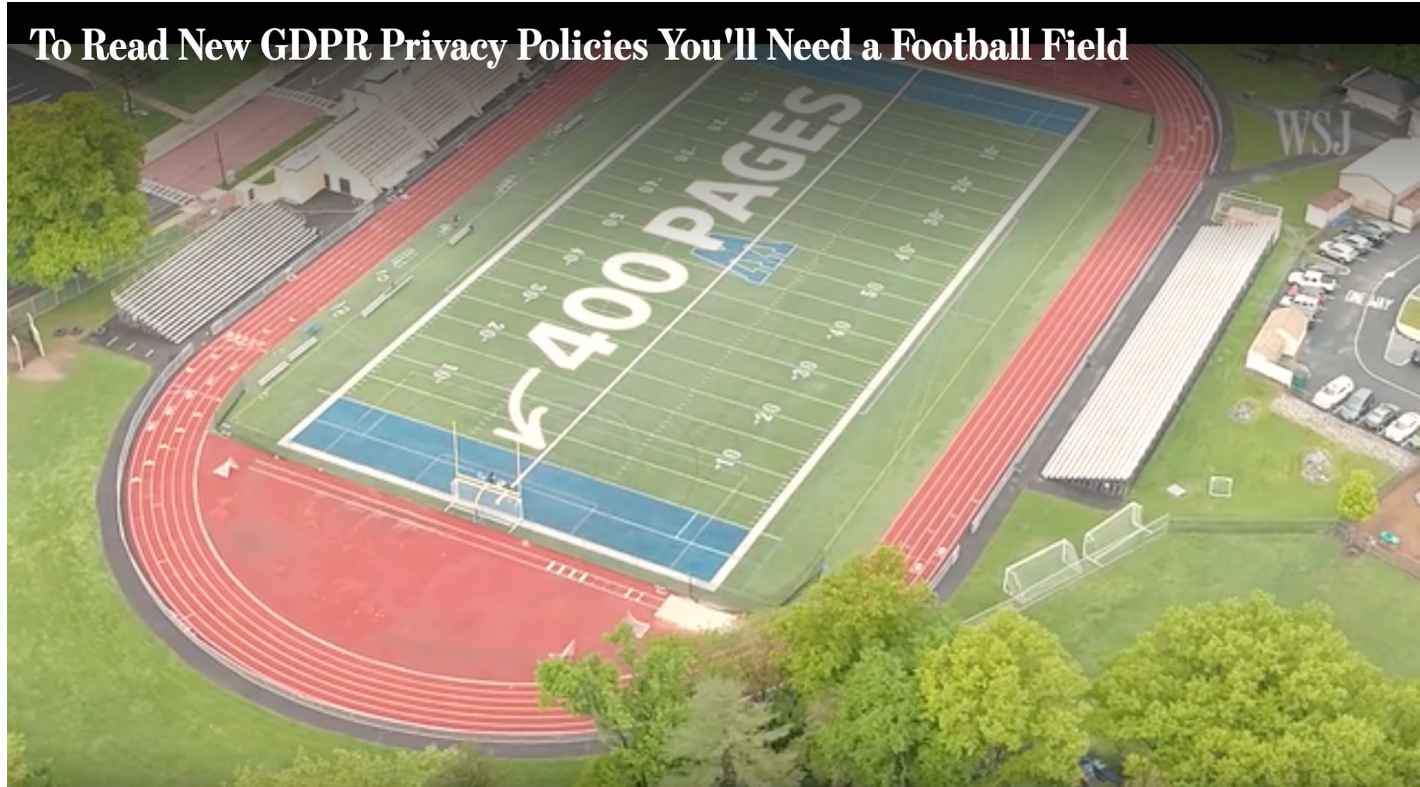
Past solutions

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

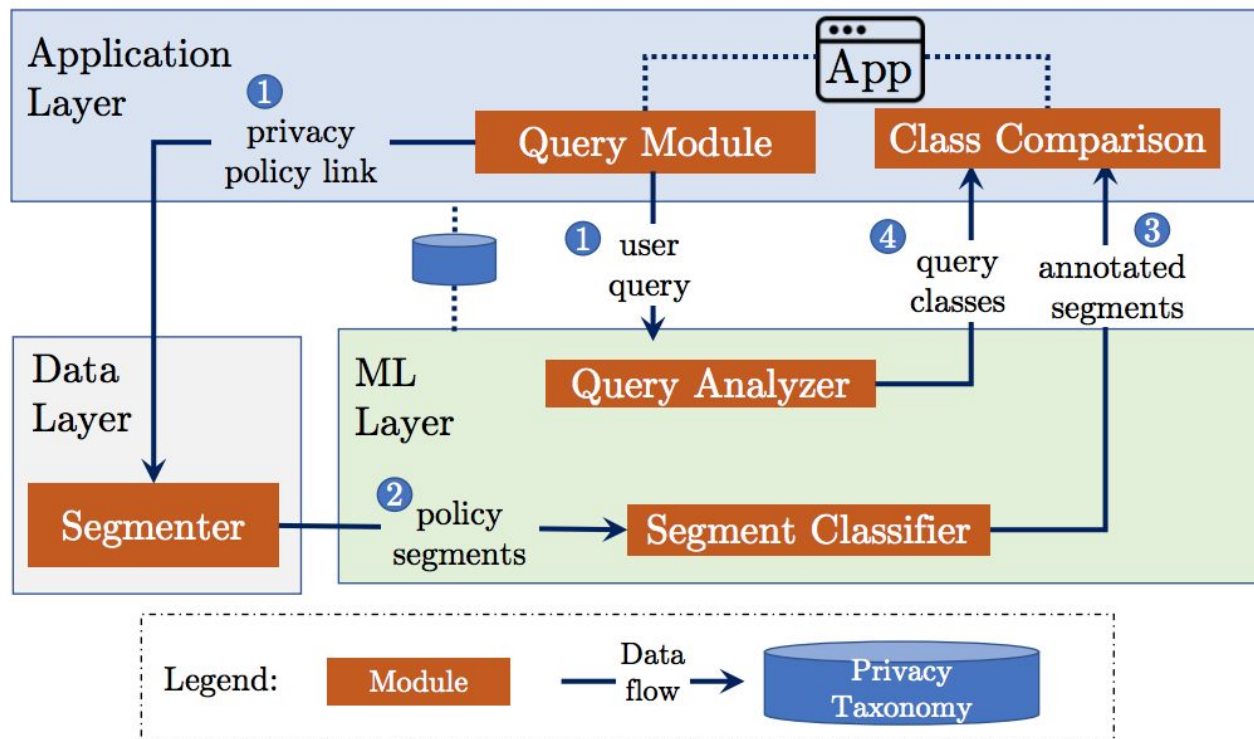What are the problems with these solutions (claimed by the author)? Do you agree?

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
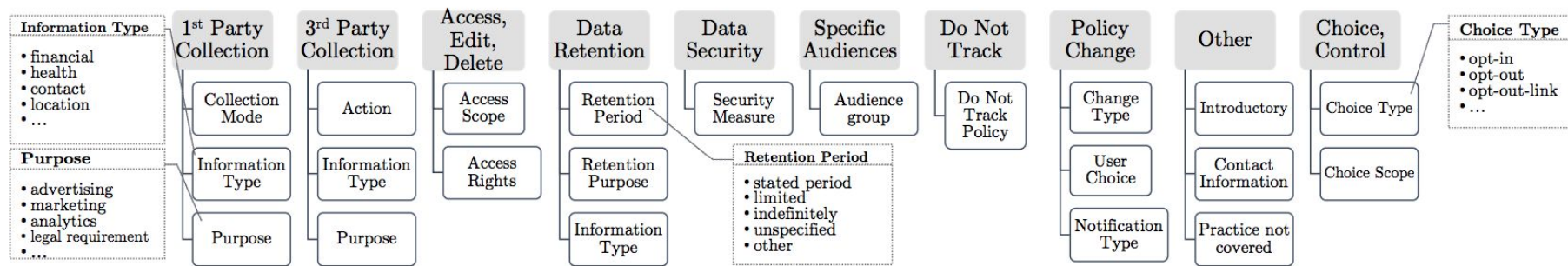- Word embeddings
- Two-level annotations

Privacy taxonomy

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

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

2. ML layer

- Two stages
- Hierarchical multi-label classification

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

3. Application layer

Supports both structured and unstructured queries

What's the consequence of getting a wrong answer?

Product critique

- Polisis: How did it decides 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

	Benefits	Limitations
Natural language	Legal value	Ambiguity, understandability, auditability
Graphical (icons)	Lay-user understandability, Measure understanding, legal design	Ambiguity, incompleteness, claim over legal compliance
Machine-readable	Automatically enforced, auditability, correctness	Human understandability, lack of adoption

Source: Three Dimensions of Privacy Policies, Morel & Pardo