

# Researchers create an AI to help us make sense of privacy policies

March 1 2018

---



A new AI tool created to help identify certain kinds of substance abuse based on a homeless youth's Facebook posts could provide homeless shelters with vital information to incorporate into each individual's case management plan. Credit: CC0 Public Domain

If you're anything like the average Internet user, you probably didn't

spend the estimated 244 hours it would take to read every privacy policy for every website you visited last year. That's exactly why a team led by Carnegie Mellon University just launched an interactive website aimed at helping users make sense of their privacy on the web.

"We've combined crowdsourcing, machine learning, and [natural language processing](#) techniques to extract annotations from [privacy policies](#) that help answer key questions that users often care about," says Norman Sadeh, the lead principal investigator on the Usable Privacy Policy Project, a School of Computer Science professor in Carnegie Mellon's Institute for Software Research, and a faculty member in the CyLab Security and Privacy Institute.

The team used artificial intelligence (AI) algorithms to crawl 7,000 of the most popular websites' privacy policies and identify those that contain language about data collection and use, third-party sharing, data retention, and user choice, among other [privacy issues](#). The project website enables people to navigate machine-annotated privacy policies and jump directly to statements of interest to them, including those often buried deep in the text of privacy policies.

The researchers' AI also evaluated each privacy policy for readability. For example, ABC News topped the rankings with a privacy policy written at a "College Graduate" reading level (Grade 26). Google's privacy policy was found to be written at a college reading level (Grade 14), the same as those of YouTube, Reddit and Amazon. Facebook's privacy policy was found to be a tad friendlier, written at a Grade 12 reading level.

"We found that the text of the policies is often vague and ambiguous, and people tend to struggle to interpret and determine what personal information is collected, how it's used, and what other entities it's shared with," Sadeh says. "From a legal standpoint, this is problematic."

To "train" their AI, the team asked a group of law students to manually annotate 115 privacy policies. The AI learned from those annotations and then crawled the policies from over 7,000 of the most popular sites on the web.

"While not perfect, our techniques are capable of automatically extracting a large number of privacy statements from the text of privacy policies," says Sadeh. "Eventually, the goal is to make this information available to users via a simple and intuitive browser plug-in that would provide users with personalized summaries highlighting those issues they are most likely to care about."

**More information:** [explore.usableprivacy.org/](https://explore.usableprivacy.org/)

Provided by Carnegie Mellon University

Citation: Researchers create an AI to help us make sense of privacy policies (2018, March 1) retrieved 11 May 2024 from <https://phys.org/news/2018-03-carnegie-mellon-ai-privacy-policies.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.