ALONI COHEN

Photo by Matthew Henry via Unsplash

Rethinking Probabilistic Identifiers for the California Privacy Law

The California Consumer Privacy Act should focus on the potential power of aggregated data, not simply its form.

EXECUTIVE SUMMARY

The California Consumer Privacy Act (CCPA) will go into effect in 2020. It governs the ways companies can use the personal information of consumers. Understanding the meaning of *probabilistic identifiers* (PI), a type of personal information, is critical to understanding the scope of CCPA, but unfortunately the legislative language for this term is vague and inconsistent.

The California Attorney General (AG) has the ability to clarify this meaning through regulation. To ensure the closing of potential loopholes and to regulate cross-device tracking, this project recommends the AG clarify that probabilistic identifiers can take many forms, including data that is not "personal information," and acknowledge that it is the power of the data, not simply its form, that matters for consumer privacy.

To learn more about this project, please visit aspentechpolicyhub.org.

⬥ THE ASPEN INSTITUTE

THE PROBLEM

In the digital advertising industry, probabilistic identification usually refers to technologies that enable *cross–device tracking.* Cross-device tracking links consumer behavior across the web, collecting information about people and devices in order to infer profiles of those consumers or devices. Under the current language of the CCPA, it is clear that inferred profiles fall within the Act's scope. However, the same is not true for raw data underlying these inferences, leaving open a loophole that could undermine the CCPA's intended scope.

> *To regulate cross–device tracking, probabilistic identifiers must include data that can be used to perform identification, not just data that has already been used to do so.*

THE SOLUTION

This project argues the meaning of probabalistic identifiers should be expanded to include data that creates inferred profiles. It recommends the AG clarify that PIs take many forms; that the potential for inference from data when aggregated is relevant; and that PIs may consist of multiple pieces of information that do not *separately* constitute personal information.

These points are crucial due to the way the digital ad industry monetizes information — using data to build device networks/inferred profiles to target ads. The existing legislative language leaves ambiguous whether data that is not personal information (e.g.: URLs, device types, etc), but when aggregated can be used to probabilistically identify, falls within the scope of CCPA. Many would agree that this is precisely what CCPA intends to address.