



User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection

Jim Isaak, IEEE Computer Society

Mina J. Hanna, Synopsys

With the revelation that Facebook handed over personally identifiable information of more than 87 million users to Cambridge Analytica, it is now imperative that comprehensive privacy policy laws be developed. Technologists, researchers, and innovators should meaningfully contribute to the development of these policies.

The discovery that Facebook gave unfettered and unauthorized access to personally identifiable information (PII) of more than 87 million unsuspecting Facebook users to the data firm Cambridge Analytica¹ has fueled growing interest in the debate over technology's societal impact and risks to citizens' privacy and well-being.² It is clear that national governance institutions demonstrably lack the ability to anticipate technology's future impact on the rights and duties of its citizens, much less its impact on the structure of society, ideological divides, and political schisms among its citizens and the expansion of identity politics promoted by isolated social and news media echo chambers.

The ubiquity of data gathering, storage, and analytics on our devices, systems, applications, and social media platforms—aimed at



personalizing experiences, optimizing sales, and maximizing return—have been disruptive in shaping the global economy, the flow of ideas, and access to information that resulted in the advancement of innovation around the information marketplace. This risk is further exacerbated by the fact that Internet of Things (IoT) devices are becoming more integrated into larger systems that govern every aspect of our lives, from the benign to the essential. The number of IoT devices grew from 500 million in 2003 to 8 billion in 2017 and is expected to grow to 50 billion in 2020.

These disruptive forces have a tangible influence on citizens' rights such as statutory rights—due process, equal representation before the law, the right to appeal, and trial by jury—and constitutional rights like freedom of expression, voting, and non-discrimination. Thus, it has never been more imperative to have an open discussion about the proliferation of technology in our lives and how it will affect our privacy rights and our security on both personal and national levels. It is also imperative for technologists, researchers, and innovators to take heed of the policy debate and meaningfully contribute to the development of these policies.

It is true that Facebook is currently being investigated by the Federal Trade Commission (FTC) for violating a 2011 consent decree. However, it is clear that the processes exposed by the current Cambridge Analytica controversy reflect a severe challenge to US privacy law, which is sorely deficient. In this article, we review how Cambridge Analytica was able to leverage its alliance with Facebook to access users' personal data, lay out principles for a comprehensive data privacy policy, and examine what's currently proposed on Capitol Hill and at state levels to address privacy concerns.

CAMBRIDGE ANALYTICA

In 2013, researchers at the University of Cambridge's Psychometrics Centre analyzed the results of volunteers who took a personality test on Facebook to evaluate their "OCEAN" psychological profile (openness, conscientiousness, extraversion, agreeableness, and neuroticism) and correlated it with their Facebook activity (likes and shares).³ This research drew in 350,000 US participants and established a clear relationship between Facebook activity (and other online indicators) and this five-factor personality profile. This work demonstrated that the OCEAN profile for any individual could be deduced reasonably accurately by looking at these metrics and without using a formal psychographic instrument. There is no indication, however, that

the Facebook Open API until May 2015. This is how Cambridge Analytica was able to access the Facebook data under scrutiny. Note that keeping the specific individual data was not necessary to accomplish the primary research goal, which was to establish a methodology for psychographic profiling of individuals based on social media and other indicators.

Cambridge Analytica realized they could integrate this information with a range of data from social media platforms, browsers, online purchases, voting results, and more to build "5,000+ data points on 230 million US adults." By adding OCEAN analysis to the other private and public data acquired, Cambridge Analytica developed the ability to "micro-target" individual consumers or voters with

It has never been more imperative to have an open discussion about the proliferation of technology in our lives and how it will affect our privacy rights and our security.

this research exposed participating Facebook users or their friends to any specific privacy abuse. There are indications that the university refused to share data (either individual or the resulting criteria) with what would become Cambridge Analytica.

Now that it was clear that such an analysis could be undertaken, a second research project was reportedly initiated by Global Science Research (GSR)—in cooperation with Cambridge Analytica—to identify the parameters needed to develop the OCEAN profiles using a personality quiz on Amazon's Mechanical Turk platform and Qualtrics, a survey platform. The quiz required users to grant GSR access to their Facebook profile, which granted access to users' friends' data through

messages most likely to influence their behavior.⁴ The OCEAN analysis was paired with a large number of targeted messages in "Project Alamo," which was employed for the election campaign of President Trump.⁵ Some of these messages were created for the Trump campaign, and some simply leveraged "news" available on the Internet (which might have included content funded through the Russian campaign to disrupt the US elections). As described by Cambridge Analytica's CEO, the key was to identify those who might be enticed to vote for their client or be discouraged to vote for their opponent.⁶ Every vote added or disrupted (in the intended way) tips the election results. This parallels analysis from the US 2010 elections.

Note that not having a Facebook account did not provide protection—the litany of available data sources is not limited to Facebook, and the analysis can easily apply to other points of personal preference. In addition, every website with the Facebook logo is linked to Facebook, allowing for tracking of non-members as well as members who might not have opted in for the service. There are many similar sources of online tracking—for instance, web beacons—most of which are tied to “cookies” that can be used across websites, and access can be sold to interested buyers. Also, by combining real news with misinformation or unconstrained Internet content, target voters will find reinforcing messages on many sites without realizing they are some of the few people in the world getting those messages, nor are they given any warning that these are political campaign messages.

With real-time monitoring of ad responses on targeted individuals, including real-time substitution to find “click bait” that worked, the ad campaign was able to both maximize its impact and detect trends not visible at the macro scale. Tipping the scale in a few states—with as few as 100,000 voters—using individualized, high-impact messages is sufficient to impact election results. This might not be the only reason for the specific 2016 US election outcome, but there is every indication that it was a useful if not a critical contribution.^{7,8}

The idea that psychographic analysis can have a significant impact on behavior has been questioned. However, a recent paper by Stanford professor Michal Kosinski (who was part of the 2013 Cambridge University research team) and colleagues confirms that it can have a significant impact with a sample base of 3.5 million users.⁹

With a broad base of personal information readily available, micro-targeting of individuals can be easily deployed. Targeted messaging can be applied to affect their behavior,

bypassing existing regulations on disclosure, informed consent, or even foreign intervention. The cost of applying these methods are meager. These factors suggest that changes in policies at both corporate and legislative levels are needed to ensure that consumers and voters’ personal data is protected, that they are notified of the affiliation of those seeking to influence them, and that they have the best opportunity to participate as informed citizens and consumers.

THE CORE PRINCIPLES OF PRIVACY AND DATA PROTECTION

In our view, any privacy and data protection legislation should include the following principles, based on the forthcoming “Personal Privacy, Awareness and Control” position statement from IEEE-USA.¹⁰

Public transparency:

- › The public must be able to learn the types of data being collected by any website or other electronic means, what data is retained, how it is used, and what is shared with third parties (directly or indirectly). The same information must be available from those third parties.
- › All data collection mechanisms must be disclosed to users, including web beacons or other mechanisms for tracking user activity or data. This information must be sufficient for users to be able to identify and pursue disclosure and controls related to these data collectors.
- › Each website and application must disclose any ongoing content placed on the user’s device, as well as the uses of that content.

Disclosure for users:

- › For each website and application, users must be able to obtain complete disclosure of the information that is retained about

them by the site or application or by any third parties accessing that information, directly or indirectly.

Control:

- › User “do not track” requests must be respected, blocking disclosure by third-party cookies and retention of non-relationship-critical data between sessions. Users must explicitly opt-in to each specific data component to be retained in this situation. This requirement extends to all “partner” third-party sites, cloud services, and collection devices.
- › Users must easily be able to delete personally identifiable data from any site, cloud service, or collection device.
- › Users must easily be able to identify, terminate, delete, and uninstall any content or applications placed on their devices or cloud service.
- › Disputes related to the purging of user data or applications must not default to licenses and arbitration processes that restrict legal response options.
- › Consent by users for a website to collect data about themselves must not be interpreted to extend to information about their “friends” or “contacts.”
- › Minors must be protected by a legally mandated age of consent to release their private information.

Notification:

- › Users must be directly and promptly informed of the loss or misuse of their private information by any organization collecting or storing that information.
- › Where and when possible, users shall have the right to know the source of violations and the responsible parties who violate their privacy.
- › Paid advertising and content

must be accompanied by clear information notifying the recipient that this is paid content, with a clear link to the source of the material and the intended beneficiary of the desired consumer action.

- › For online content, metadata should lead to the sponsoring site(s), allowing the user to understand and pursue the transparency, disclosure, and control actions indicated above.


CURRENT PROPOSED LEGISLATION

At the federal level, following the grueling and lengthy hearings before the House Judiciary and Senate Judiciary and Commerce Committees where Mark Zuckerberg was asked to testify on Facebook's privacy and data policy, a few senators put forward bills attempting to govern public data privacy. The most overarching bill comes from the offices of Senators Richard Blumenthal (D-CT) and Ed Markey (D-MA). The bill, titled the CONSENT Act (S.2639) or "Customer Online Notification for Stopping Edge-Provider Network Transgressions" (www.congress.gov/bill/115th-congress/senate-bill/2639/text?q=%7B%22search%22%3A%5B%22privacy%22%5D%7D&r=15), requires the FTC to establish privacy protections for customers of online edge providers.

The bill will require explicit opt-in consent from users of Facebook and other online platforms before these online platforms use, share, or sell any of their users' PII, as well as explicit notification any time data is gathered, shared, or sold to a third party, in addition to adding new reporting requirements in case of a data breach involving sensitive customer proprietary information. The bill describes violations of this act similar to unfair or deceptive acts prescribed under section 18(a)(1)(B) of the Federal Trade Commission Act (15 U.S.C. 57a(a)(1)(B)), thus giving the FTC jurisdiction to prosecute violators.

The Social Media Privacy Protection and Consumer Rights Act of 2018 (S.2728; www.congress.gov/bill/115th-congress/senate-bill/2728/text?q=%7B%22search%22%3A%5B%22privacy%22%5D%7D&r=2), introduced by Senator Amy Klobuchar (D-MN), draws similar constraints to the CONSENT Act regarding disclosure of privacy policy and obtaining initial consent and privacy preferences, but adds restrictions on modifications to privacy terms, provisions regarding withdrawal of consent, and procedures when a violation of privacy has occurred (for example, notification, data erasure, and ceasing to collect any further data).

California is taking the lead in the US by advancing a privacy bill to the State Legislature that would grant its citizens data privacy rights.¹¹ The bill adds limits to selling data on users younger than 16 years of age and prevents businesses from denying service to users should they choose to exercise their rights under the bill.

The privacy debate on Capitol Hill might have lost some momentum from when it started in April after the Facebook/Cambridge Analytica data misuse was revealed. At present, Congress might be more occupied with passing authorization and appropriation bills for 2019. Yet the debate is far from settled, and it remains to be seen if the Blumenthal-Markey or the Klobuchar bills advance to the floor for a vote and if the California legislative measure will set a precedent for the rest of the states, which could follow suit. 

REFERENCES

1. H. Davies, "Ted Cruz Using Firm That Harvested Data on Millions of Unwitting Facebook Users," *The Guardian*, 11 Dec. 2015; www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data.
2. M. Kosinski et al., "Manifestations of User Personality in Website Choice and Behavior on Online Social Networks," *Machine Learning*, vol. 95,

no. 3, 2014, pp. 357–380.

3. "Cambridge Analytica—The Power of Big Data and Psychographics," presentation, Cambridge Analytica, June 2016; www.youtube.com/watch?v=n8Dd5aVXLCc&t=86s.
4. B. Anderson and B. Horvath, "The Rise of the Weaponized AI Propaganda Machine," *Scout*, 9 Feb. 2017; <https://scout.ai/story/the-rise-of-the-weaponized-ai-propaganda-machine>.
5. R. Lindholm, "Project Alamo's Data-driven Ads on Facebook won Trump the Election," *Semantiko*, 19 Feb. 2017; <https://semantiko.com/project-alamo-trump-facebook-ads>.
6. "Alexander Nix, CEO, Cambridge Analytica—Online Marketing Rockstars Keynote," presentation, Cambridge Analytica, 10 Mar. 2017; www.youtube.com/watch?v=6bG5ps5KdDo.
7. R.M. Bond et al., "A 61-Million-Person Experiment in Social Influence and Political Mobilization," *Nature*, vol. 489, 2012, pp. 295–298.
8. "Cambridge Analytica CEO interviewed by John Humphrys," BBC Radio 4, 1 Oct. 2016; www.youtube.com/watch?v=hAW1kTLuAIs.
9. S.C. Matz et al., "Psychological Targeting as an Effective Approach to Digital Mass Persuasion," *PNAS*, vol. 114, no. 48, 2017, pp. 12714–12719.
10. "Personal Privacy, Awareness and Control," *IEEE-USA*, to be published in 2018; www.globalpolicy.ieee.org.
11. I. Lapowsky, "Bill Could Give Californians Unprecedented Control Over Data," *Wired*, 22 June 2018; www.wired.com/story/new-privacy-bill-could-give-californians-unprecedented-control-over-data.

JIM ISAAK is a past president of the IEEE Computer Society. Contact him at jiminh@gmail.com.

MINA J. HANNA is chair of the IEEE-USA AI and AS Policy Committees. Contact him at minajeane.stanford@gmail.com.