Aux Etats-Unis, les données de près de 200 millions d’électeurs diffusées par erreur

Ces données, disponibles sur le Web sans protection, comportaient à la fois des données publiques, comme le nom et l’adresse de tous les inscrits sur les listes électorales de chaque Etat américain, mais également des données plus sensibles.

LE MONDE | 21.06.2017 à 14h29 | Par Martin Untersinger (/journaliste/martin-untersinger/)

En quelques années, les données personnelles sont devenues le carburant des campagnes électorales, tout particulièrement aux Etats-Unis, où la législation en la matière est considérablement plus lâche qu’en France ou en Europe . Cette course aux données comporte cependant un versant plus sombre : les informations personnelles de la quasi-totalité des électeurs américains, collectées pour faire campagne, ont été, pendant au moins deux jours, accessibles sur Internet sans protection (http://gizmodo.com/gop-data-firm-accidentally-leaks-personal-details-of-ne-1796211612) .

Cette fuite sans précédent – elle concerne au moins 198 millions d’Américains, soit plus de 60 % de la population – a été repérée le 12 juin par Chris Vickery, spécialiste des fuites de données pour l’entreprise de sécurité informatique américaine UpGuard (https://www.upguard.com/breaches/the-rnc-files) . Ces données étaient stockées sur un serveur non sécurisé appartenant à Deep Root Analytics, une société spécialisée dans le ciblage des publicités politiques à la télévision et dont le Parti républicain est client de longue date.

Ces données comportaient à la fois des données publiques, comme le nom et l’adresse de tous les inscrits sur les listes électorales de chaque Etat américain, mais également des données plus sensibles. A chaque électeur, doté d’un identifiant unique interne au Parti républicain, était accolée son opinion présumée sur de nombreux sujets politiques, comme les armes à feu, les impôts , l’écologie ou la recherche sur les cellules souches.

« Des fichiers de notre système de stockage interne ont été consultés à notre insu. (…) *Cet accès a été obtenu à la suite d’un changement récent de nos paramètres. Nous ne pensons pas qu’il s’agit d’un piratage »,* a reconnu Deep Root Analytics dans un communiqué sur son site (https://www.deeprootanalytics.com/2017/06/19/data-security-statement/) . L’accès aux données a été de nouveau sécurisé le 14 juin, lorsque UpGuard a *« averti les autorités fédérales »*.

Un secteur souvent dans l’ombre

Les données exposées provenaient également de TargetPoint et DataTrust, deux autres entreprises qui gèrent, elles aussi, des données électorales pour le Parti républicain. Plus étonnant, dans l’ensemble de ces données figuraient de larges copies du contenu du forum américain Reddit , sans que l’on sache vraiment pourquoi. Si cette masse de texte publié par des internautes a pu servir à « entraîner » une intelligence artificielle , elle témoigne peut-être de la volonté de l’entreprise de récupérer des informations sur les utilisateurs très politisés de ce site, qui comportait une section pro-Trump très active et organisée.

S’il n’est guère étonnant qu’un parti politique américain utilise des données personnelles pour mener campagne, cette fuite permet cependant de lever un coin de voile sur un secteur souvent dans l’ombre et de prendre la mesure de la très grande précision et du volume faramineux de ces données.

Ces dernières ont d’ailleurs été très vraisemblablement utilisées par la campagne de Donald Trump . Dans une enquête publiée en décembre 2016, le site spécialisé AdAge révélait (http://adage.com/article/campaign-trail/trump-camp-s-inexperience-set-stage-rnc-data-win/307105/) que la campagne du milliardaire n’a travaillé que très tardivement les données personnelles des électeurs, lorsque le parti républicain a mis à son service sa machine de guerre, construite grâce aux entreprises avec lesquelles il travaillait depuis des années. Soit celles qui sont concernées aujourd’hui par cette fuite.

Une autre entreprise, Cambridge Analytica – qui se vante de pouvoir prédire la personnalité des utilisateurs des réseaux sociaux afin de mieux les convaincre – a souvent été présentée comme une des pièces maîtresses de la stratégie électorale de Donald Trump. Cette fuite montre que des outils de ciblage plus classiques étaient également utilisés, alors même qu’un des cadres de Cambridge Analytica a considérablement nuancé, après l’élection de novembre 2016, (https://www.nytimes.com/2017/03/06/us/politics/cambridge-analytica.html) les prouesses technologiques de sa firme.

Cette fuite intervient enfin dans un contexte sensible aux Etats-Unis. Les autorités ont accusé le Kremlin, en janvier, d’avoir tenté d’influencer les élections. Or, les données qui ont été rendues disponibles sont un matériau de premier choix pour qui s’intéresse au système politique américain et désirerait, le cas échéant, l’influencer. C’est sans compter que, comme toute fuite massive de données personnelles, elle va faciliter les piratages et les usurpations d’identité.