Robin Worrall

LA DOBLE PETJADA DEL *BIG DATA*

PRIVADESA I SOSTENIBILITAT, DOS GRANS REPTES D'INTERNET

Xavier Duran

El 5 de febrer del 2021 els periodistes d'investigació Charlie Warzel i Stuart A. Thompson van publicar en la web de *The New York Times* un article on explicaven que algú els havia proporcionat dades de milers de seguidors de Donald Trump que el 6 de gener havien protagonitzat una invasió violenta del Capitoli a Washington (Warzel i Thompson, 2021). I això, tot i que els números de mòbil no porten associada una identitat, havia permès traçar l'origen de moltes persones fins al lloc d'on venien i fins i tot esbrinar els seus noms, adreces i comptes de xarxes socials.

Dos anys abans, als dos autors ja els havien arribat dades de localitzacions de més de dotze milions de ciutadans nord-americans, cosa que permetia rastrejar-ne els moviments. Però en el cas del Capitoli, cada localització anava associada a un codi ID –o *mobile advertising identifier*–, que és únic per a cada usuari i està lligat al seu mòbil intel·ligent o a la seva tauleta. Només calia encreuar aquest ID amb altres bases de dades per obtenir una immensa quantitat d'informació sobre cada persona.

«En 2018 Facebook tenia dades personals per omplir, de mitjana, 400.000 documents de Word amb cada usuari»

És possible que molta gent consideri útil i fins i tot necessària la possibilitat d'utilitzar un ID per perseguir el delicte. Però, tal com expliquen els dos periodistes, aquest ID el fan servir nombroses empreses, institucions i entitats, inclosos bancs i fons d'inversió, que poden obtenir, així, una gran informació sobre qualsevol ciutadà.

La nostra activitat diària ho fa possible. Gairebé tothom navega per Internet i fa cerques i una gran majoria de ciutadans realitzen, en algun moment, una compra a distància. Però, encara que no es faci res d'això, tots estem a nombroses bases de dades, des de padrons municipals a entitats esportives o culturals, passant per serveis mèdics, entitats d'estalvi i arxius fiscals.

I molts, a més, tenim activitat a les xarxes socials. I això dona moltíssima informació. Segons explicava Mat Travizano al setembre de 2018 a *Entrepreneur* (Travizano, 2018), Facebook tenia en aquell moment dades personals per omplir, de mitjana, 400.000 documents de Word amb cada usuari. I Google en podia omplir vora tres milions per a cada internauta.

Com s'ha arribat a aquesta situació, en què aquestes empreses tenen tanta informació? I quin impacte té en la nostra privacitat i seguretat?

LES DADES, MODEL DE NEGOCI

Per comprendre-ho hem d'anar a finals dels anys noranta del segle xx. Havien sorgit multitud d'empreses tecnològiques, les anomenades *puntcom*. Les expectatives de creixement eren altes i, probablement, també massa optimistes. I el capital risc s'hi va abocar. Això va fer pujar de forma meteòrica, a borsa, unes empreses que, en la major part, no tenien ni tan sols un model de negoci. Guanyaven quota de mercat, però amb molts serveis gratuïts i, per tant, no generaven beneficis. Cap a l'any 2000, la situació de la majoria d'aquestes empreses era molt delicada, per no dir desesperada, i tenien molts més deutes i promeses de futur que no pas realitats.

I va ser aleshores quan empreses com Google van veure quin potencial tenien per monetitzar els seus serveis: les dades. Tot i que els mòbils intel·ligents no estaven gaire estesos i la capacitat dels ordinadors i dels algoritmes per processar grans quantitats de dades eren limitades, l'ús de la informació com a matèria primera va funcionar. Així, Google va ingressar 19 milions de dòlars l'any 2000, però el 2001 ja eren 86 milions; el 2002, 440 milions; el 2003, 1.500 milions, i el 2004 va ingressar 3.200 milions de dòlars. Un augment del 3.590 % en quatre anys. El 2020 va ingressar 146.920 milions només en publicitat.

La clau havia de ser la publicitat personalitzada. Amb totes les dades que havia acumulat, Google podia esbrinar els interessos dels usuaris. I si una empresa volia enviar anuncis, Google podia, sense revelar identitats, dirigir-la de manera que, en comptes de fer campanyes indiscriminades, ja es podia anar als grups més susceptibles de fer cas a aquells anuncis. Oferir equipament de golf als que s'interessen pel golf, viatges al Carib a persones que solen buscar informació sobre les platges del Carib i estades a càmpings als que s'acaben de comprar una caravana.

D'entrada, això no sembla dolent. Molta gent pensava, i encara pensa, que si rep publicitat personalitzada li fan un favor, perquè li trien la que li pot interessar. I si tot i així no la vol, no hi perd res. Algunes molèsties, a tot estirar.

El problema apareix quan la informació ja no serveix només per a enviar publicitat personalitzada, sinó per a decidir si es permet a aquell usuari contractar una assegurança mèdica o de vida, si és prou solvent per concedir-li un crèdit o una hipoteca o fins i tot si val la pena contractar-lo per a una feina.

Tenir dades és tan valuós que moltes empreses ja ho situen com un objectiu, al marge de si fabriquen neveres o cotxes o si lloguen vehicles o apartaments. El novembre del 2018, el CEO de Ford va anunciar l'objectiu de monetitzar i vendre les dades que recollia dels cent milions de persones que conduïen els seus vehicles (Sadowski, 2020, p. 31). A part dels seus hàbits i rutes, potser podria esbrinar si conduïen de forma prou prudent o per carreteres prou segures, cosa que interessa molt a companyies d'assegurances.

Sobre això ja n'hem parlat a MÈTODE en una altra ocasió (Duran, 2018), però les possibilitats de saber coses de nosaltres creixen de forma desmesurada. A més, Google, Facebook i Amazon, que es reparteixen la major part del mercat publicitari digital mundial –es calcula que el 2018 entre Google i Facebook controlaven el 84 % del mercat mundial exceptuant la Xina–, saben qui som, perquè per utilitzar els seus serveis ens hi hem de d'inscriure i donar com a mínim el nostre correu electrònic.

Amb dades del 2018, Amazon va tenir 6.720 milions de dòlars d'ingressos en publicitat. Sembla poc si ho comparem amb els 38.370 milions de Facebook i els 83.680 de Google. Però pensem que, en teoria, la major part dels seus ingressos han de ser per vendre productes, no per anuncis. Però la publicitat li dona més marge que les vendes.

DEDUCCIONS PERILLOSES I IDENTIFICACIÓ AMB UNA FOTO

I ja no parlem, tot i la seva importància, de la filtració o venda directa de dades. A primers d'abril de 2021, es van filtrar dades personals de més de 553 milions d'usuaris de Facebook. Més enllà d'això, amb milions de dades i algoritmes cada vegada més sofisticats es poden fer moltes més deduccions –que poden ser encertades o no– i incidir en aspectes més íntims. És el que li va passar a Ángel Cuevas, professor de la Universitat Carlos III de Madrid. Estava en una reunió de feina a Barcelona i va rebre a través de Facebook publicitat que el convidava a connectar amb la comunitat gai i reservar un apartament «amb gent com tu».

Cuevas va pensar com interpretava Facebook si era o no homosexual si ell no havia donat mai informació sobre la seva orientació –se suposa que Facebook ho feia, encertadament o no, a partir d'altres dades–. Però, a més, es va preguntar per què l'empresa permetia que alguns anunciants li enviessin publicitat basada en la seva hipotètica orientació sexual.

A partir d'aquí, ell i el seu equip van investigar i van descobrir que a l'Aràbia Saudita, per exemple, unes 500.000 persones tenien l'etiqueta d'homosexual. A països on això pot significar condemna de presó o fins i tot a mort això és molt perillós, perquè, encara que Facebook no reveli identitats, en comunitats relativament petites identificar individus concrets no és tan difícil (García et al., 2018). De fet, una pràctica per a estudiants avançats de tractament de dades pot ser prendre de forma aleatòria una persona qualsevol d'arreu del món i mirar d'esbrinar el màxim de coses sobre ella (Véliz, 2020, p. 65).

És cert que el Reglament Europeu de Protecció de Dades, en ple vigor des de maig del 2018, proporciona protecció als ciutadans, i que per acceptar que des d'una

Tyler Merbler

Anete Lusina

Per a bona part de la població, utilitzar les xarxes socials o fer una compra en línia ja forma part de la rutina. Totes aquestes activitats que realitzem a Internet proporcionen molta informació a empreses que han convertit les dades en matèria primera amb què aconseguir beneficis econòmics milionaris.

«Tenir dades és tan valuós que moltes empreses ja ho situen com un objectiu, al marge de si fabriquen neveres o cotxes o si lloguen vehicles o apartaments»

Gràcies al *mobile advertising identifier*, un codi d'identificació únic per a cada usuari vinculat al seu dispositiu, les dades de milers dels seguidors de Donald Trump que van participar a l'assalt del Capitoli en gener de 2021 van acabar a les mans de dos periodistes d'investigació de *The New York Times*. Només encreuant aquest codi amb altres bases de dades es va obtenir una immensa quantitat d'informació sobre cada persona.

web tinguin accés a les nostres dades n'hem de donar el consentiment explícit. Però també ho és que les condicions dels famosos «accepto» són sovint molt embolicades i que si volem accedir a un servei o a una web tendim a dir que sí a tot el que ens proposen. Fins i tot s'han sancionat entitats bancàries per faltes greus, com no facilitar al client la possibilitat d'utilitzar els serveis sense accedir a proporcionar dades. A més, empreses no europees potser no sempre respecten aquest reglament. L'abril de 2021, al Regne Unit es va anunciar una demanda contra TikTok, una *app* xinesa per compartir vídeos curts, per recopilar, suposadament de forma il·legal, dades personals de milions d'infants, com números de telèfon, ubicació de la connexió i, fins i tot, dades biomètriques. El 2019, la firma xinesa va rebre una multa de 5,7 milions de dòlars de la Comissió Federal del Comerç d'Estats Units per un mal ús de les dades.

A més, quan entrem en una web pot ser que, en realitat, estiguem cedint dades a nombroses webs amb què la primera té acords. Són les famoses *cookies* de tercers, que permeten seguir el nostre rastre i fer servir les dades. Això i molts altres problemes amb la privacitat estan exposats en el reportatge «Tot el que sabem de tu», emès el novembre de 2020 en el programa *30 minuts* de TV3 (Duran et al., 2020).

En el mateix reportatge s'explica el cas de l'artista i investigadora Joana Moll. Quan desenvolupava el seu projecte *The dating brokers* ("Els marxants de cites") va descobrir, de manera casual, que a Internet es venien dades de clients de pàgines de contactes. Moll va comprar un milió de perfils d'usuaris de tot el món per 136 euros. Hi incloïa uns 600.000 perfils d'homes i uns 300-400.000 perfils de dones, amb direccions de correu electrònic, noms d'usuari, dates de naixement, orientació sexual, descripcions molt detallades del físic i la personalitat, si tenien fills, si fumaven, si prenien drogues...

Però, a més, va descobrir que aquestes webs pertanyien a empreses que, al seu torn, formaven part de grans grups. Com que es donava el consentiment per cedir dades a totes les empreses del grup, algú que hagués entrat en una web de contactes podria haver acabat proporcionant dades sensibles a més de 700 empreses.

Les possibilitats de l'ús de dades personals i els riscos que signifiquen són nombrosos (Sadowski, 2020; Véliz, 2020). I probablement no hi ha lloc on amagar-se, com demostra el cas de Clearview, també exposat en el reportatge anterior. Es tracta d'una aplicació creada als Estats Units que, a partir de la imatge d'una persona qualsevol i gràcies al reconeixement facial, proporciona altres fotografies d'aquesta mateixa persona, fins i tot d'anys enrere, que estan a Internet, i les adreces de les webs on es troben. A partir d'aquí, l'usuari pot anar a les webs i elaborar un perfil sobre qualsevol individu.

La base de dades de Clearview té uns tres mil milions d'imatges capturades sense coneixement dels que hi surten. Ja ha rebut diverses demandes, però si bé el que diguin els jutges és important, el que volem destacar és que cada vegada hi ha més tecnologies que permeten trobar piles d'informació de qualsevol persona, per discreta que sigui, perquè tots podem estar etiquetats en fotografies, sovint sense saber-ho.

EL NEGOCI I LES EMISSIONS DELS SERVIDORS

Abans hem explicat que Amazon obté una bona part dels seus ingressos per publicitat. De fet, dels 280.000 milions que va ingressar el 2019 –amb 11.500 milions de beneficis–, el 50,37 % provenien de les vendes *online*. Això significa poc més de la meitat; l'altra no ve d'aquestes vendes. Aproximadament un 6 % provenia de botigues físiques, però més important era el 19,17 % de vendes de tercers. Amazon els proporciona tota la infraestructura digital i es queda un percentatge important de les vendes. De fet, en les vendes pròpies Amazon ajusta molt el preu i estableix marges petits, però té una estratègia: cobra ràpidament dels seus clients i paga a termini més llarg als seus proveïdors. Així genera molta liquiditat.

El 2019 també, un 12,48 % dels ingressos provenien d'Amazon Web Services (AWS). Es tracta d'un servei perquè les empreses puguin tenir els seus arxius i programes en el núvol i no hagin de gastar en servidors o bases de dades propis.

Aquest sembla un dels grans negocis de futur. Es calcula que el 2019 al núvol hi havia 45 zettabytes de dades –45.000 trilions de bytes, equivalent a més de 7.000 milions d'anys de vídeo d'alta definició– i que el 2025 n'hi haurà 175 zettabytes. Això significarà un mercat de vora 600.000 milions d'euros. I actualment AWS controla una mica més del 40 % el mercat. Azure, de Microsoft, en tenia, el 2019, el 29,4 % i Google Cloud el 3 %.

Els gran servidors d'Amazon mostren la gran visió del seu propietari i director executiu, Jeff Bezos. Tot i que parlar de núvol fa pensar a molta gent que les dades es passegen per l'atmosfera fins que algú les captura en el seu ordinador, aquest núvol el formen estructures físiques, que guarden programes i dades, i milers de quilòmetres de fibra òptica per on viatgen. Potser enviem un correu electrònic al veí del costat i quan li arriba ha passat per un servidor que està vora el pol Nord. Els servidors –ordinadors on hi funcionen programes accessibles des de diferents punts de la xarxa– busquen en cada moment els camins més adients per la xarxa de fibres i quan ens baixem una cançó o fem una compra *online* no sabem ben bé per quines parts del planeta han passat els bits que ho han permès.

Però això significa un gran consum d'energia i moltes emissions de CO_2, encara que no siguin tan evidents com els fums que surten per les xemeneies. Quan teclegem a l'ordinador, a la tauleta o al mòbil, el tenim engegat i gastant. I també està gastant el dispositiu on algú rebrà el missatge i el servidor de la web que estem consultant. I estan gastant els servidors, engegats les 24 hores del dia, perquè Internet mai no dorm. No en som conscients, però si Internet està sempre disponible i en qualsevol moment podem moure'ns per milions de webs o tenir activitat a qualsevol xarxa social és perquè aquests grans servidors estan actius.

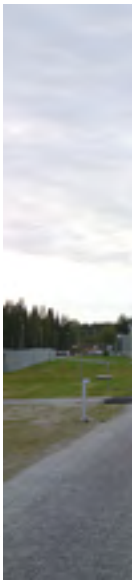
Donar el servei és un bon negoci perquè moltes empreses, fins i tot grans, ja no gasten en infraestructures pròpies sinó que les lloguen –Netflix mateix és client d'AWS.

Però al marge de qui faci diners, el negoci és nefast per al medi ambient. Els processadors han guanyat en eficiència, però la quantitat d'informació creix de forma exponencial. Alguns pronòstics apunten que de cara al 2030 el conjunt de les tecnologies de la informació i la comunicació passaria a consumir el 2 % de l'electricitat a escala global (Stern, 2020). I vora un 40 % del consum energètic dels centres de dades és degut a la seva refrigeració. Per eficients que siguin, els servidors, amb milers de processadors, s'escalfen i aquesta calor s'ha de dissipar.

Una opció és instal·lar els servidors a zones molt fredes. Facebook en té un a Luleå, al nord-est de Suècia, on a més pot disposar de grans quantitats d'energia hidroelèctrica i estalviar costos i emissions de CO_2.

Tot i així, el consum és molt elevat. I si parlem d'emissions equivalents de CO_2, en serem conscients de l'impacte. Alguns estudis calculen unes emissions anuals d'Internet de mil milions de tones, equivalents a un 2,8 % de les emissions totals –més que el sector de l'aviació, responsable d'un 2 %.

«Un estudi de la Universitat de Bristol estimava que el 2016 el visionat de vídeos de YouTube va produir 11,13 milions de tones de CO_2»



Cristopher Down

QUANT CO_2 EMET UN CORREU ELECTRÒNIC?

Segons un estudi de la companyia energètica britànica OVO, els britànics envien diàriament 64 milions de correus electrònics innecessaris, d'aquests que només diuen «hola» o «gràcies» o equivalents (Tweedale, 2021). Cada correu fa emetre un gram de CO_2. Per tant, els correus que es podrien estalviar són responsables de 23.475 tones de

Cottonbro

Tot i que el Reglament Europeu de Protecció de Dades ofereix protecció a la ciutadania, les empreses de fora d'Europa no sempre respecten aquest marc regulatiu. A l'abril de 2021, el Regne Unit va anunciar una demanda contra TikTok, una *app* xinesa de vídeos curts molt popular entre els més joves, per recopilar dades personals de milions de nens, com números de telèfon, ubicació de la connexió i, fins i tot, dades biomètriques.

Els servidors que permeten a Internet funcionar de forma permanent són grans estructures físiques que tenen un gran consum d'energia i moltes emissions de CO_2. Vora un 40 % del consum energètic d'aquests centres és degut a la necessitat de refrigerar els servidors funcionant a ple rendiment. Una solució que algunes empreses han trobat és instal·lar els servidors a zones molt fredes; en la imatge, el centre de dades de Facebook a Luleå (Suècia).

diòxid de carboni; en altres llocs es parla de 16.433 tones. Semblen moltes emissions i equivalen a 22 vols d'anada i tornada entre Londres i Nova York. Però com que les emissions anuals totals britàniques van ser 435,2 milions de tones el 2019, deixar d'enviar correus innecessaris només les reduiria en un 0,0037 % (si bé tot suma, és clar).

Però hem de tenir en compte moltes pràctiques més. Un estudi fet per la Universitat de Bristol estimava que el 2016 el visionat de vídeos de Youtube va produir 11,13 milions de tones de CO_2 (Preist et al., 2019). Comparat amb les emissions mundials, uns 35.000 milions de tones, torna a semblar poc. Però equivalen a les produïdes per una ciutat com Frankfurt o com Glasgow o per països com Luxemburg o Zimbabwe el mateix any. I a YouTube hi hem d'afegir tota la música descarregada i totes les sèries, pel·lícules i documentals vistos a plataformes. I tots els jocs dins l'anomenat *gaming*. O bé fer una cerca en un buscador, traduir un text, enviar fotografies o presentacions... Com més complex el material, més bits i més emissions.

El problema no són només les emissions actuals, sinó les perspectives de creixement. Segons càlculs de l'Agència Internacional de l'Energia, els bitcoins provoquen tantes emissions com Nigèria o Uruguai. Per a la plataforma Digiconomist encara en són més i superen les de Colòmbia i Bangladesh. Segons un article publicat a principis d'abril a *Nature Communications* (Jiang et al., 2021), al ritme actual, a la Xina, el 2024, tot el procés que envolta les transaccions i validacions en bitcoins produirà tantes emissions de gasos d'hivernacle com Itàlia o Txèquia. I en l'àmbit intern, les emissions se situarien en un dels deu primers llocs de 182 ciutats i 42 sectors industrials de la Xina.

Això és perquè fer del bitcoin una moneda virtual segura requereix una sèrie de càlculs complexos per garantir-ne la fiabilitat i mantenir, al mateix temps, la privacitat, basant-se en l'anomenat *blockchain* ("cadena de blocs"). Però si pensem que el bitcoin és una més de les monedes virtuals i que ara només representa el 0,4% dels diners en circulació, podrem suposar l'impacte que pot tenir d'aquí uns anys.

Hi ha solucions que no passin per utilitzar menys les eines digitals? D'entrada, potser haurem d'aprendre a no malbaratar recursos i fer-ne un ús més racional. Al mateix temps, podem esperar que els tecnòlegs trobin solucions més sostenibles i que l'energia vingui cada vegada més de fonts renovables. Si busquem a Google «green Internet» veurem que el tema desperta interès: dona 5.360.000.000 de resultats. La cerca sola ja ha produït més emissions i visitar unes quantes d'aquestes webs encara en generarà més. Si en alguna hi ha solucions viables i eficients, podem donar les emissions per bones.

REFERÈNCIES

Duran, X. (2018). Tot allò que saben de nosaltres: Es pot navegar amb privacitat per l'oceà del 'big data'? *Mètode*, *99*, 4–9. https://metode.cat/revistes-metode/article/tot-allo-que-saben-de-nosaltres.html

Duran, X., Bonet, X. (autors), & Solà, C. (director). (2020, 8 de novembre). Tot el que sabem de tu [episodi de programa de televisió]. En C. Fernández (Productor) *30 minuts*. Espanya: Corporació Catalana de Mitjans Audiovisuals. https://www.ccma.cat/tv3/alacarta/30-minuts/tot-el-que-sabem-de-tu/video/6067763/

García, D., Mitike Kassa, Y., Cuevas, A., Cebrián, M., Moro, E., Rahwan, I., & Cuevas, R. (2018). Analyzing gender inequality through large-scale Facebook advertising data, *PNAS*, *115*(27), 6958–6963. https://doi.org/10.1073/pnas.1717781115

Jiang, S., Li, Y., Lu, Q., Hong, Y., Guan, D., Xiong, Y., & Wang, S. (2021). Policy assessments for the carbon emission flows and sustainability of Bitcoin blockchain operation in China. *Nature Communications*, *12*, 1938. https://doi.org/10.1038/s41467-021-22256-3

Preist, C., Schien, D., & Shabajee, P. (2019). Evaluating sustainable interaction design of digital services: The case of YouTube. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3290605.3300627

Sadowski, J. (2020). *Too smart. How digital capitalisme is extracting data, controlling our lives, and taking over the world*. MIT Press.

Stern, E. (2020, 27 d'octubre). Núvols i aire fred per a la computació. *Divulcat*. https://www.enciclopedia.cat/divulcat/nuvols-i-aire-fred-per-a-la-computacio

Travizano, M. (2018, 28 de setembre). The tech giants get rich using your data. What do you get in return? *Entrepreneur*. https://www.entrepreneur.com/article/319952

Tweedale, A. (2021). The carbon footprint of the internet: What's the environmental impact of being online? *OVO Blog*. https://www.ovoenergy.com/blog/green/the-carbon-footprint-of-the-internet.html

Véliz, C. (2020). *Privacy is power*. Bantam Press.

Warzel, C., & Thompson, S. A. (2021). They stormed the Capitol. Their apps tracked them. *The New York Times*. https://www.nytimes.com/2021/02/05/opinion/capitol-attack-cellphone-data.html

XAVIER DURAN. Químic i periodista científic, redactor especialitzat en ciència i tecnologia als serveis informatius de TV3. Entre els seus últims llibres hi ha: *L'individu transparent* (Pagès Editors, 2016), *L'imperi de les dades* (Bromera, 2018) i *La ciència en la literatura* (Universitat de Barcelona, 2018).