Ethische aspecten van big data

Big data heeft niet alleen geleid tot uitdagende technische vraagstukken, ook gaat het gepaard met allerlei nieuwe ethische en morele kwesties. Om verantwoord met big data om te gaan, moet ook over deze kwesties worden nagedacht. Want slecht datagebruik kan nadelige gevolgen hebben voor grote groepen mensen en voor organisaties. In de slotaflevering van deze serie verkennen Klaas Jan Mollema en Niek van Antwerpen op een pragmatische manier de ethische kant van big data, zonder te blijven steken in de negatieve effecten ervan.

Niek van Antwerpen MA. en Klaas Jan Mollema MSc.

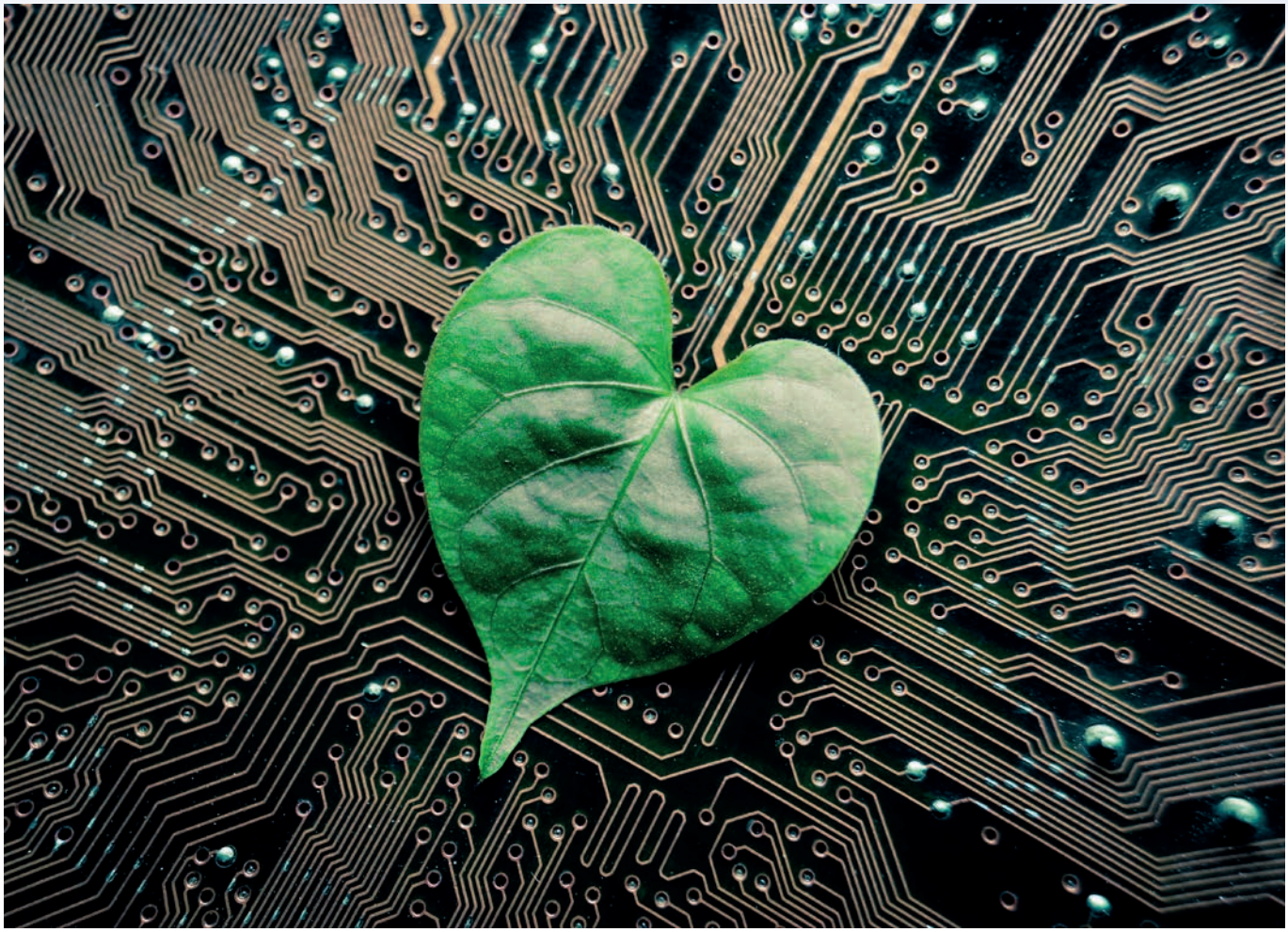
Ons leven wordt steeds meer vastgelegd in data. Het resultaat: een datarealiteit die bestaat uit waarheden en onwaarheden over wie we zijn en wat we doen. Voor iemand die de juiste technische en financiële mogelijkheden heeft én de toegang heeft tot deze data, is ons heden en verleden vrijwel geheel transparant geworden. Zo is het voor verzekeringsmaatschappijen steeds makkelijker om via data-analyse een premie te bepalen op basis van persoonlijke omstandigheden. Soms is dit wenselijk, maar heel vaak ook niet. Volgens de Consumentenbond is de autoverzekering van zwarte auto's bijvoorbeeld hoger dan de premie van hetzelfde type auto met een andere kleur, omdat uit data blijkt dat zwarte auto's kwetsbaarder zijn. Met deze transparantie neemt onze kwetsbaarheid toe. Ons gedrag, psychische gesteldheid, karakter of bepaalde gewoonten worden inzichtelijk en voorspelbaar – en kunnen strategisch worden geëxploiteerd. Dit terwijl andersom de methoden, algoritmen en profielen die het bedrijfsleven en de overheid daartoe inzetten, voor ons nauwelijks transparant en inzichtelijk zijn.

'Big data zet de bescherming van informationele privacy steeds meer onder druk'

Machtsasymmetrie

Het speelveld tussen enerzijds de individuele burger en anderzijds de kapitaalkrachtige dataverzamelaars, zoals de overheid en het bedrijfsleven, is in relatie tot big data dan ook ongelijk. Deze *machtsasymmetrie* tussen bovengenoemde partijen zal onder invloed van technologische innovatie en de mogelijkheden om steeds meer data te verzamelen, te analyseren en te exploiteren, alleen maar toenemen.[1]

In die zin is het debat over privacy in relatie tot big data uitermate belangrijk. Privacy is van oudsher een afweerrecht dat onze persoonlijke levenssfeer beschermt tegen de inmenging en macht van derden en de staat. Privacy vormt daarmee een belangrijke beschermende normatieve en juridische mantel tegen de aantasting van allerlei belangrijke morele waarden zoals rust, intimiteit, integriteit, individualiteit, persoonsvorming en autonomie.[2]



Privacy is echter niet alleen een afweerrecht, maar ook een recht op zelfbeschikking. Juist omdat onze privacy wettelijk beschermd is, hebben we de vrijheid om – tot op zekere hoogte – zelf de inhoud en inrichting van ons privé-leven te bepalen en kunnen we vrij relaties met anderen aangaan en kunnen we bijvoorbeeld ook vrij met iedereen communiceren.

Informationele privacy

Er zijn verschillende privacyvormen, maar in relatie tot big data is met name *informationele privacy* van belang. Deze relatief nieuwe privacyvorm bestaat sinds het midden van de vorige eeuw, toen computers en databanken een belangrijke rol gingen spelen in de samenleving. Het is gerelateerd aan de controle over persoonlijke informatie en betreft een stelsel van normen in de zorgvuldige omgang met persoonsgevoelige informatie.[3]
Theoretici brengen informationele privacy tegenwoordig – binnen een normatieve context – in verband met belangrijke morele waarden als anonimiteit en vrijwaring van beoordeling, manipulatie, stigmatisering en voorspelbaarheid.[4] Deze privacywaarden botsen vaak met andere waarden, zoals efficiency, effectiviteit, winstmaximalisatie of veiligheid, en delven daarbij vaak het onderspit. Binnen een juridische context wordt informationele privacy beschermd door de grondwet en de Wet bescherming persoonsgegevens.

Onder druk

Big data zet de bescherming van informationele privacy echter steeds meer onder druk. Draait het in de huidige privacywetgeving om principes als toestemming of noodzakelijkheid en doelbinding, bedoeld om het ongebreideld delen en verspreiden van persoonlijke informatie tegen te gaan, bij big data gaat het juist om een ongerichte dataverzameling en een gebrek aan doelbinding. Big data is voornamelijk gebaseerd op secundair gebruik van enorme volumes reeds verzamelde gegevens.[5] Het gaat, kortom, om allerlei datastromen die

‘Steeds vaker blijkt dat het anonimiseren van gegevens de privacy van individuen niet voldoende kan waarborgen’

‘Bij ethische vraagstukken spelen behalve big data ook technologieën en zaken als het verzamelen, interpreteren en bewaren een belangrijke rol’

afkomstig zijn uit tal van gekoppelde databronnen en die veelal uit hun context zijn gehaald.
Al betreft het vaak geanonimiseerde datasets, steeds vaker blijkt dat het anonimiseren van gegevens de privacy van individuen niet voldoende kan waarborgen. Dit komt mede door de enorme hoeveelheid gegevens die bij big data beschikbaar zijn, door het koppelen en samenvoegen van data en door moderne dataminingmethoden. Hierdoor kunnen combinaties van onschadelijke geanonimiseerde data toch herleidbaar zijn tot personen.

Een voorbeeld

Een goed voorbeeld is het onderzoek naar 173 miljoen individuele taxiritten over het jaar 2013 in New York. Deze open dataset, vrijgegeven door New York City Taxi & Limousine Commission, bevatte gegevens over de routes, op- en uitstappunten, tijden, locaties, vervoersprijzen en fooien. Het unieke taxi-identificatienummer was geanonimiseerd, maar waren de data daarmee ook ‘anoniem’?
Al snel haalden verschillende onderzoekers gevoelige informatie boven water, zoals het gemiddelde inkomen en zelfs de huisadressen van sommige taxichauffeurs. Een onderzoeker van Neustar Research combineerde data uit de taxidataset met die uit publieke bronnen, zoals celebrityblogs, en bracht zo de routes van acteurs in kaart. Ook wist hij huisadressen te achterhalen van vaste bezoekers van stripclubs. Weer andere onderzoekers slaagden erin taxiritten – en de rustpauzes – te combineren met de vaste gebedstijden van moslims, waardoor ze de routes van moslimtaxichauffeurs uit de dataset konden filteren.[6]
Het blijkt dus dat zelfs uit geanonimiseerde en op zichzelf onschadelijk open datasets potentieel risicovolle en zelfs schadelijke correlaties kunnen worden aangemaakt die de individuele privacy kunnen aantasten.

Groepsprivacy

Privacy staat in relatie tot big data ook op een ander punt onder toenemende druk. Het gaat hierbij om de problematiek met betrekking tot de zogenaamde groepsprivacy.[7] Ondanks de anonimisering van persoonlijke gegevens lukt het derden om algemene conclusies te trekken op basis van groepsprofielen. Individuen die passen binnen een dergelijk groepsprofiel kunnen hierdoor worden gestigmatiseerd of gediscrimineerd. Iemand kan bijvoorbeeld in een ‘verkeerd’ postcodegebied wonen, een bepaalde etnische achtergrond hebben of tot een bepaalde leeftijdscategorie behoren. Mensen worden hierdoor digitaal ‘gestript’ van hun individualiteit en puur op basis van bepaalde kenmerken in een groepsprofiel geplaatst.[8]
Digitale burgerrechtenorganisaties, privacyspecialisten en de Wetenschappelijke Raad voor Regeringsbeleid waarschuwen dat dit op termijn een verkoelend effect kan opleveren in de relatie tussen de staat en burger.[9] Anders gezegd: burgers kunnen de overheid gaan wantrouwen en vormen van zelfcensuur gaan toepassen.
Ook kan de individuele rechtsbescherming in het gedrang komen. De burger zal mogelijk moeten gaan bewijzen ten onrechte met een bepaald profiel te worden geassocieerd in plaats van omgekeerd.

Ethiek van kunstmatige intelligentie

Ethische vraagstukken blijven niet beperkt tot big data zelf. Ook de technologieën en de praktijken die bij het verzamelen, bewaren, analyseren en interpreteren van big datasets betrokken zijn spelen een belangrijke rol. Big data-uitspraken zijn gebaseerd op het gebruik van algoritmes en kunstmatige intelligentie. Computers worden steeds krachtiger. Niet alleen neemt hun rekenkracht nog steeds exponentieel toe en kunnen deze steeds meer data verwerken. Onder invloed van nieuwe ontwikkelingen in de kunstmatige intelligentie krijgen computers bovendien de capaciteit om door ervaring zelf *dingen te leren*: machine learning.
Aan de hand van algoritmes kunnen computers patronen herkennen in big data en op basis daarvan nieuwe verbanden zoeken, voorspellingen doen of zelfstandig beslissingen nemen. Machines zijn daardoor steeds beter in staat om van ons gedrag te leren en complexe keuzes voor ons te maken. Ze hebben daarmee een sturende werking op onze besluiten. Algoritmes zullen in toenemende mate beïnvloeden hoe bijvoorbeeld onze vra-

gen worden beantwoord, wie we moeten *daten*, welke films we moeten bekijken of wat voor nieuws we onder ogen krijgen. Ze zullen eveneens beïnvloeden – of zelfs bepalen – tot welke banen we toegang krijgen, binnen welke risicoprofielen we worden geplaatst of wat de hoogte gaat worden van onze verzekeringspremies.

Ethische uitdagingen

Deze ontwikkelingen leveren tal van ethische uitdagingen op. En niet alleen vanwege de enorme omvang van de analyses en de complexiteit rondom de besluitvorming door algoritmes en de toenemende impact van deze algoritmes op ons individuele leven of de samenleving als geheel.
Het proces op basis waarvan deze algoritmes worden ontworpen of hoe ze tot hun besluit komen is weinig transparant en vanwege de toenemende complexiteit en autonomie van algoritmes ook nog eens moeilijk controleerbaar.[10] Algoritmes opereren als het ware in een 'black box': wat er onder de motorkap gebeurt wordt veelal zakelijk en politiek afgeschermd en blijft onzichtbaar voor de buitenwereld.[11] Hier ligt een belangrijk moreel vraagstuk. We dragen steeds meer gezag over aan algoritmes, maar het noodzakelijke toezicht op hun ontwerp, de keuzes die ze maken of bijvoorbeeld de nadelige effecten die ze kunnen oproepen ontbreekt vooralsnog. En dat terwijl de impact en complexiteit van de problematiek van de algoritmes snel toenemen.[12] Zorgelijk is dat veel mensen een rotsvast vertrouwen hebben in data en computerintelligentie. Wat machines tonen op het beeldscherm wordt zo sturend voor hun handelen. De betrouwbaarheid van de besluitvorming door algoritmes kan echter op allerlei manieren negatief worden beïnvloed. Bijvoorbeeld door de kwaliteit van de datasets. De verzamelde data kunnen incompleet, incorrect of verouderd ('garbage in, garbage out') zijn. Er kunnen bovendien verkeerde conclusies worden getrokken op basis van de gevonden correlaties. Techniek is daarnaast niet neutraal: zij bevat altijd een waardeoordeel omdat het een product is ontworpen door mensen. De besluitvorming door algoritmes kan dus worden beïnvloed door de bias die zowel in de datasets zelf als in de algoritmes aanwezig is. Dit kan tot uitkomsten leiden die bepaalde groepen kunnen discrimineren of juist bevoordelen.[13]
Bijkomstig probleem is dat naarmate algoritmes onder invloed van kunstmatige intelligentie en zelflerend vermogen meer zelfstandig gaan opereren in complexe en omvangrijke netwerken, het bepalen van de (morele) verantwoordelijkheid rondom hun handelen naar alle waarschijnlijkheid steeds lastiger wordt.[14] Hoewel de ontwikkeling rond kunstmatige intelligentie al een tijdje bestaat, ontbreekt bij de overheid vooralsnog een fundamentele visie hoe bovengenoemde problematiek rondom big data te reguleren.

Rol voor informatieprofessionals

Informatieprofesionals kunnen in dit speelveld een heldere positie innemen. Door zijn rol is de informatieprofessional neutraal in de informatievoorziening. Het controleren van big data-conclusies, beschermen van gevoelige informatie en het valideren van de analyse zou in die rol passen. De kwaliteit, betekenis en waarheid van informatie is immers een van de kernproducten van de informatiespecialist.

Noten

1] Andrejevic (2014).
2] Westin (1967), Gerstein (1978), Cohen (2000).
3] Koops (2014).
4] Nissenbaum (2010), Floridi (2004), Introna (1997).
5] Wetenschappelijke Raad voor het Regeringsbeleid. Rapport 95.
6] Metcalf, Crawford (2016).
7] Floridi (2014).
8] Hildebrandt (2011).
9] Zie rapport WRR.
10] Burrell J (2016).
11] Pasquale, F., (2015).
12] Wetenschappelijke Raad voor het Regeringsbeleid. Rapport 95.
13] Friedman, Nissenbaum H. (1996).
14] Floridi (2012).

Klaas Jan Mollema MSc. (www.zijlmo.nl/mo) is specialisatiecoördinator Business Data Management aan de opleiding Informatica van Hogeschool Leiden. Niek van Antwerpen MA. is docent Information & Media Studies aan opleiding HBO-ICT aan De Haagse Hogeschool.

Literatuurlijst

] Andrejevic, M., (2014) Big data, big questions: The big data divide. I*nternational Journal of Communication* 8(0): 17.
] Burrell, J., (2016) How the machine 'thinks:' Understanding opacity in machine learning algorithms. *Big Data & Society* 3(1): 1–12.
] Cohen,J.(2000) Examined Lives: Informational Privacy and the Subject as Object. *Stanford Law Review* 53(3): 1373–1438.
] Floridi L. (2012). Distributed morality in an information society. *Sci. Eng. Ethics* 19, 727–743.
] Floridi L. (2014) T*he Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford: OUP.
] Floridi L. (2014). Open data, data protection, and group privacy. *Philos. Technol.* 27, 1–3.
] Friedman, B., Nissenbaum H (1996) Bias in computer systems. *ACM Transactions on Information Systems* (TOIS) 14(3): 330–347.
] Gerstein, R., (1978) 'Intimacy and Privacy', *Ethics*, 89: 76–81.
] Hildebrandt M (2011) Who needs stories if you can get the data? ISPs in the era of big number crunching. *Philosophy & Technology* 24(4): 371–390.
] Koops, E. J. (2014). Privacy, informatieveiligheid en een onzichtbare medaille. In S. Kok (editor), Informatieveiligheid s.l.: Taskforce Bestuur & Informatieveiligheid Dienstverlening.
] Introna (1997),D. 1997. Privacy and the Computer: Why we Need Privacy in the Information Society. *Metaphilosophy* Vol. 28, Nos. 3, July 1997.
] Metcalf, J., Crawford K. (2016) Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Societ*y January–June 2016: 1–14.
] Nissenbaum H. (2010) *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Palo Alto, CA: Stanford University Press.
] Pasquale, F., (2015) *The Black Box Society: The Secret Algorithms that Control Money and Information*, Cambridge: Harvard University Press.
] Tene, O., Polonetsky, J.(2013). Big Data for All: Privacy and User Control in the Age of Analytics. *North Western Journal of technology & Intellectual Property*. Vol 11. Iss 5.
] Westin A.F,.(1967) *Privacy and Freedom*, New York: Athenum 1967.
] Wetenschappelijke Raad voor het Regeringsbeleid, *Big data in een vrije en veilige samenleving*. University Press Amsterdam, gepubliceerd op Rapport 95.

'Zorgelijk is dat veel mensen een rotsvast vertrouwen hebben in data en computerintelligentie'