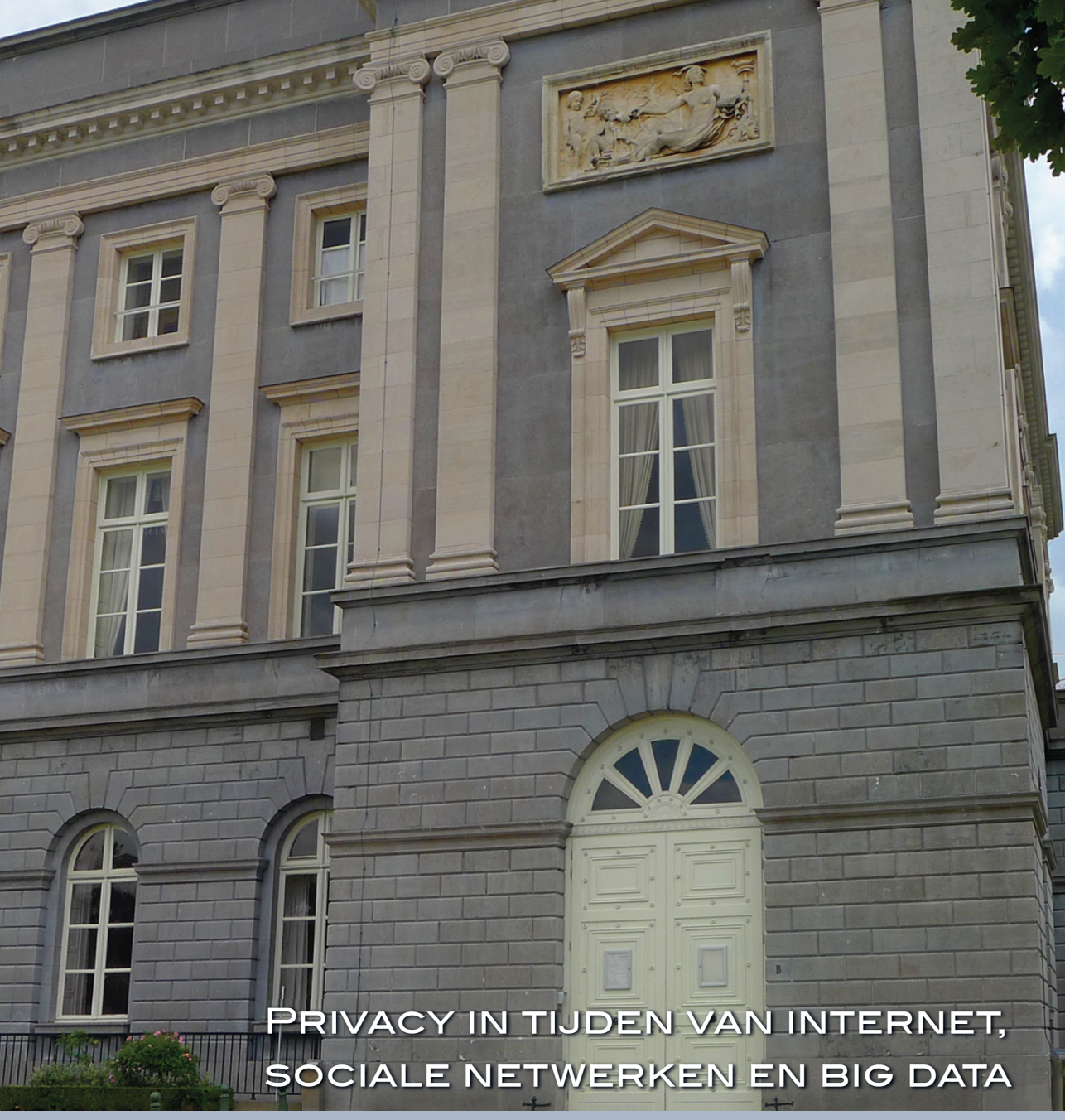


Privacy in tijden van internet, sociale netwerken en big data

Yolande Berbers
Mireille Hildebrandt
Joos Vandewalle e.a.

Koninklijke Vlaamse Academie van België
voor Wetenschappen en Kunsten, 2017
Standpunten 49

Privacy in tijden van internet, sociale netwerken en big data

KVAB Press

Uitgaven
van
de Koninklijke
Vlaamse Academie
van België
voor
Wetenschappen
en Kunsten

Standpunten nr. 49

KVAB Press

Hertogsstraat 1
1000 Brussel
Tel. 02 550 23 23
www.kvab.be
info@kvab.be

Privacy in tijden van internet, sociale netwerken en big data

Yolande Berbers
Willem Debeuckelaere
Paul De Hert
Yvo Desmedt
Frank De Smet
Mireille Hildebrandt
Karolien Poels
Jo Pierson
Bart Preneel
Joos Vandewalle

Aanbevolen citeerwijze: Yolande Berbers, Mireille Hildebrandt, Joos Vandewalle (e.a.), *Privacy in tijden van internet, sociale netwerken en big data,* KVAB Standpunt 49, 2017.

D/2017/0455/05
ISBN 978 90 6569 917 29

Foto en ontwerp cover: Anne-Mie Van Kerckhoven

Privacy in tijden van internet, sociale netwerken en big data

INHOUD

Samenvatting

Het gebruik van internet, sociale media en big data brengt vandaag de dag de privacy in gevaar, ook van gewone gebruikers. Dit Standpunt richt zich vooral tot privégebruikers van alle generaties die geen bijzondere ICT-scholing of -opleiding genoten, die intensief gebruik maken van deze diensten en middelen, en die zich zorgen maken over de gevaren waaraan hun privacy blootgesteld is. Weten of die zorgen al dan niet terecht zijn vergt niet alleen inzicht in de technologische mogelijkheden en beperkingen. Er zijn ook de commerciële belangen en hun relatie tot de inperking van en de gevaren voor onze persoonlijke privacy bij het gebruik van de vele, vaak waardevolle en nuttige diensten. Andere rapporten en Standpunten behandelen meer specifieke aspecten van de privacy en haar regelgeving: het betreft dan patiënten, of ondernemingen en instellingen die bestanden met gegevens van personen, werknemers, studenten of klanten bijhouden en verwerken.

De ICT-wereld goochelt graag met jargon-woorden waarvan de draagwijdte niet doordringt naar het bredere publiek, en in de media worden soms angstaanjagende situaties beschreven en vaak weinig onderbouwde stellingen verkondigd. Daarom bespreken we eerst de voornaamste begrippen. Wat is of zijn 'machineleren', dataontginning en big data? Welke privacyproblemen doen zich voor? Welke marsrichtingen zijn er naar een betere privacy?

Om het voor de oningewijde lezer concreter te maken bespreken we vervolgens een aantal concrete situaties waar zich de meestal ongekende privacy-gevaren schuilhouden: het digitale leven van een gezin, de aanwending van big data bij het profileren van passagiers, het internet der dingen in de context van slimme steden, gedistribueerde informatie versus centrale collectie, de zelfrijdende wagen en de informatievergaring over locaties. Dit alles speelt zich af in de leefwereld van vandaag. De evolutie is nog volop bezig en jong en oud gebruiken steeds nieuwe diensten en toepassingen.

Ook gewone gebruikers kunnen nu reeds hun gedrag bijsturen. We eindigen dit rapport daarom met een tiental aanbevelingen voor diverse doelgroepen: ICT-verantwoordelijken, alerte burgers, de bouwers van ICT- en Internet der Dingen-apparaten, overheden en bedrijven. De aanbevelingen betreffen ook de 'voorzienbaarheid', profieltransparantie en doelbinding, het machtsonevenwicht, het vermijden van onwenselijke *data bias*, grenzen aan het gebruik van big data door de overheid, *digital clearing house* en de taak van het onderwijs.

Over dit onderwerp bestaat er heel wat wetenschappelijke literatuur. Er zijn ook heel wat recente en breder toegankelijke teksten beschikbaar, onder andere op websites. De bibliografie zet de geïnteresseerde lezer op weg.

Executive summary

Privacy in times of internet, social media and big data

The current use of the internet, social media and big data severely affects the privacy of ordinary users. This positioning paper is primarily aimed at the private user young and old who did not have special education or training regarding ICT but still uses these services intensively and who, whether or not, rightly worries about the hazards to which his or her privacy is exposed. This requires not only a better and deeper understanding of the technological possibilities and limitations, but also the commercial interests, and their relation to the constraints and threats of our personal privacy when using the many often valuable services. The specific aspects of privacy as patients, or the privacy regulations for companies and institutions that track and process files with data from individuals, employees, students, or customers, is not dealt with but is referred to other reports. This positioning paper has been conceived by a working group of members of KVAB and external experts covering the different aspects of this interdisciplinary subject, that have met regularly over a period of one year.

Since the ICT world is often overwhelmed with "jargon" words, the scope of which does not penetrate or because the newspapers sometimes describe very frightening lowly-backed situations, we first discuss the main concepts both at the level of the machine learning, data extraction and the big data, as well as the privacy issues that arise, and finally the ways in which a better privacy can be acquired.

In order to make this more concrete for the modal reader, we discuss important privacy hazards in a number of concrete situations, such as the digital life of a family, the big data police in passenger profiles, the internet of things, the context of smart cities, distributed information versus central collection, autonomous vehicles, and location information. Although this digital revolution is not over yet, the modal user can already modify his behavior.

There is extensive scientific literature on this subject, but there are also many widely accessible texts available recently, including websites, to which the interested reader is referred to in the bibliography.

The ten recommendations mainly focus on various target groups and situations.

Recommendation 1: Responsibilities. Privacy in the big data is an issue for citizens, engineers, consumers, companies, institutions, media and governments. This calls for the provision of sufficient resources to the supervisors, especially with regard to companies that derive their earnings model from big data analysis.

Recommendation 2: Alert citizens. Citizens, whose data are being processed, should try to maximize their rights under the GDPR. The verification of personal data requires that the individual gains insight into the use and misuse of the data, as a precondition for genuine freedom of choice. Precisely because it is extremely difficult for individuals, we recommend that those concerned use the opportunity to exercise their claims through mandating to consumer or privacy organizations (Article 80 GDPR).

Recommendation 3: Providence, Profile Transparency, and Goal Binding. Although the profiles themselves are not related to a particular person and thus are not personal data themselves, the fundamental right to data protection (GDPR) applies to a person who fits within the "validation" of the profile. The right to profile transparency implies the obligation to inform stakeholders and explain how they are profiled and this beyond a correlation or statistical relationship.

Recommendation 4: Power Unbalance. If the person responsible for an ICT service relies on the consent for the use of personal data, then it must be easy to withdraw, with a limitation of permission in time. They will not apply a manifest power imbalance between the data subject and the controller or processor, e.g. because the responsible person provides the dominant (or only) service in the market. The controller must demonstrate that there is no power imbalance or that this imbalance cannot affect the consent of the person concerned.

Recommendation 5: The builders of ICT and IoT devices must make use of technologies that maintain privacy and allow transparency for the end user. They need to work on 'privacy by design', taking privacy from the start of the design as an important requirement, and not being "stuck" afterwards. The service providers must allow users to assemble services of different origins. The designers of algorithms must write their algorithms to ensure users' privacy. Application designers need to allow transparency, work on efficient and effective technologies that allow users to authorize their data usage. Additionally, one must make certification of applications so users are sure that the applications are safe. Typically privacy must be default.

Recommendation 6: Role of government and companies. It is the duty of government and companies to check for each big data solution whether the risks for the protection of personal data and the risks to society as a whole outweigh the benefits. In doing so, one should always check if it is not possible to achieve the same goal by using less data or aggregating data.

Recommendation 7: Preventing unwanted data bias. The responsible designers and service providers must always check whether inaccurate or unfair 'data bias', 'algorithm bias' or 'output bias' is hidden in the data sets with which algorithms are being trained, either in mathematical models themselves or in the output (indirect discrimination).

Recommendation 8: Limits to the use of big data by the government. The use of public sector big data, both in the field of detection of tax and social security fraud and in the context of national security, crime and law enforcement, should always be subject to a review by the relevant supervisors. In addition, the legitimacy and the related proportionality must be paramount, which also requires a marginal efficiency test. It is essential that legislation be provided that determines how and when the result of data mining and statistical analyzes (correlations) by the government may or may not be used as legal evidence to make decisions in individual cases (e.g. in dealing with fraud, law enforcement ...).

Recommendation 9: Establishing a digital clearing house. It is advisable to set up a Digital Clearing House (DCH) that monitors the quality of the various digital market regulators.

Recommendation 10: Task of education. Specific to young people, education has a task of bringing awareness, attitudes, skills and behavior from the actual life spheres such as home, school and friends (e.g. youth associations). It is important to point out to young people the "pitfalls" of their own behavior, as expressed, for example, in the privacy paradox.

Voorwoord

Over de reeks Standpunten

De reeks Standpunten van de Academie is een bijdrage tot een wetenschappelijk onderbouwd debat over actuele maatschappelijke en artistieke thema's. De auteurs, leden en werkgroepen van de Academie schrijven in eigen naam, onafhankelijk en met volledige intellectuele vrijheid. De goedkeuring voor publicatie door een of meerdere Klassen van de Academie waarborgt de kwaliteit van de publicatie. Dit Standpunt werd goedgekeurd voor publicatie door de klassenvergadering van de Klasse van de Technische Wetenschappen op 18 mei 2017.

1. Inleiding en situering

De informatie- en communicatietechnologie (ICT) is de voorbije vijftig jaar razendsnel geëvolueerd. In 1969 werden voor het eerst een paar korte boodschappen uitgewisseld tussen twee computers op 800 kilometer afstand van elkaar. In 1983 werd het internet geboren, met het TCP/IP-protocol om boodschappen in pakketten door te sturen. Mede dankzij de Vlaming Robert Calliau werd in 1990 het world wide web (www) voor het eerst gebruikt bij CERN. En in 2008 braken de sociale netwerken door. Vandaag de dag gebruiken jongeren en ouderen deze zeer gebruiksvriendelijke en zelfs verleidelijke technologieën wereldwijd. De evolutie gaat voort.

ICT heeft veel gunstige gevolgen voor mens en samenleving, maar veel maatschappelijke neveneffecten waren niet gepland en niet voorspelbaar tijdens het ontwerp- en ontwikkelingsproces. Zo kon men moeilijk erop anticiperen dat de technologieën zouden leiden tot een concentratie van persoonsgegevens in de handen van enkele grote spelers. Het oogsten en gebruiken van privégegevens en big data opent potentieel heel interessante opportuniteiten voor nieuwe en betere diensten en producten van bedrijven en het creëert onder meer de verwachting van meer veiligheid voor de burgers. Maar dit kan ook een bedreiging vormen voor de persoonlijke levenssfeer van gebruikers en leiden tot het ongewenst gebruiken en verzamelen van gegevens, ongewenste reclame, chantage en ook tot computermisdaden. Tim Berners-Lee, de uitvinder van het web, luidt zelfs de noodklok over de ondergang van privacy.

Vaak zijn gebruikers zich niet voldoende bewust van de gevaren. Bovendien zijn ICT-bedrijven maar matig geïnteresseerd in de problematiek en is de maatschappij onvoldoende gewapend, zowel organisatorisch, juridisch als technisch, om haar onder controle te houden. KVAB besliste eind 2015 een werkgroep op te starten die de meest pregnante aspecten vanuit technisch, maatschappelijk en juridisch standpunt in kaart kon brengen. Dit Standpunt is daar een resultaat van. Het formuleert aan het eind onderbouwde aanbevelingen voor de diverse actoren, met daarbij ook de overheid, de bedrijfswereld, de onderwijswereld en het bredere publiek. Voor de ontwikkeling van de huidige en de volgende generatie van technologische diensten, de zogenaamde Vierde Industriële Revolutie, is het immers wenselijk dat deze diensten wereldwijd op een duurzame manier de waarden van de grondrechten van de mens respecteren. Het individu moet er zeker van kunnen zijn dat zijn gegevens op een rechtmatige wijze worden verwerkt en moet over voldoende mogelijkheden beschikken om een niet-legitieme verwerking te blokkeren. Daartoe is effectieve transparantie nodig, tot achter de muur van bedrijfsgeheimen en/of nationale veiligheid.

Historische vergelijkingen zijn altijd slechts partieel, maar ze kunnen ons wel alert maken voor de mogelijke ernstige impact van het gebruik van collecties met grote

hoeveelheden persoonsgegevens ('big data'). Reeds in 1933 organiseerde nazi-Duitsland een nationale volkstelling bij 41 miljoen Duitsers, waarin onder andere persoonsgegevens over hun etnische oorsprong werden bijgehouden. In het kader van de holocaust hield men een centraal archief bij met ponskaarten (Hollerith-kaarten) die vlot mechanisch uitgelezen konden worden met informatie over 17,5 miljoen mensen uit Duitsland en de bezette gebieden. Dat vulde ruim 27 kilometer archiefplanken met documenten: lijsten, inventarissen, persoonsbeschrijvingen, verslagen van medische experimenten, verordeningen enz. Je ziet er de ambtelijk aangestuurde moordmachine en zijn omvang. De persoonlijke gegevens waarover grote ICT-bedrijven van big data in onze tijd beschikken, zijn vele keren groter. Bovendien kunnen de huidige computers de gegevens heel snel automatisch raadplegen en machinaal verwerken.

De lezer die niet thuis is in de begrippen en de context van dit onderwerp, leest het best eerst het tweede hoofdstuk. Daarin worden begrippen als 'big data', 'machinaal leren' en 'data-ontginning' (*data mining*) toegelicht, worden de diverse, vaak nog ongekende problemen geformuleerd en de technologische, organisatorische, maatschappelijke en juridische oplossingen besproken. In hoofdstuk 3 passeren concrete casussen de revue, waarin de problemen die in hoofdstuk 2 geïntroduceerd werden, in realistische contexten geplaatst worden. In hoofdstuk 4 ten slotte worden conclusies getrokken en aanbevelingen geformuleerd voor de diverse betrokken spelers. Lezers die liever sneller een concreet beeld krijgen van de diverse situaties waarin hun privacy in gevaar is, kunnen onmiddellijk naar hoofdstuk 3 zappen en voor de begrippen die ze nog niet kennen, even terugspoelen naar de relevante sectie in hoofdstuk 2. Ten slotte kunnen lezers, die snel wensen te weten hoe de diverse actoren moeten omgaan met deze evoluties, meteen naar hoofdstuk 4 springen. De onderbouwing van de conclusies en aanbevelingen daar vinden ze in de vorige hoofdstukken.

Dit Standpunt behandelt het generieke gebruik van internet, sociale netwerken en big data. Het gebruik in de gezondheidszorg komt in een ander Standpunt van KVAB aan bod [Verdonck, Van Hulle e.a. 2017]. Vooral voor verantwoordelijken bij het verwerken van persoonsgegevens werd recent een meer specifiek rapport [CBPL 2017] met 33 aanbevelingen uitgebracht door de Belgische *Commissie voor de Bescherming van de Persoonlijke Levenssfeer* (CBPL).

2. Omschrijving en context van de belangrijkste begrippen

Aangezien de belangrijkste begrippen nogal nieuw zijn en vaak slechts vaag of onvolledig gekend zijn, worden ze eerst concreet omschreven, met hun context en de ontwikkelingen die ze al achter de rug hebben. Eerst verkennen we een aantal basisbegrippen, zoals big data, machinaal leren en *data mining*, daarna worden een aantal problemen in verband met privacy omschreven en ten slotte bekijken we mogelijke oplossingen.

2.1 *Begripsverkenning*

Big data en *data mining*

Big data zou 'anders' zijn dan eerdere vormen van gegevensverzameling omdat er sprake is van een ander Volume (exponentieel grote hoeveelheid), een andere Velocity (snelheid van verwerken kan zelfs realtime zijn) en een andere Variety (veel verschillende soorten data kunnen nu geïntegreerd worden verwerkt). Dat zijn de 3 V's. Het gaat in ieder geval om grote hoeveelheden gestructureerde en ongestructureerde data. Dat laatste wil zeggen: data in allerlei formats (tekst, plaatjes, geluid) en vanuit allerlei bronnen, zoals e-mails, video's, brieven, rapporten, blogs, postings, cijfers, archieven, sensoren, camera's enz. Sommige data zijn verbonden met identificeerbare personen, bijvoorbeeld met een welbepaald aspect van hun identiteit. Dat zijn de zogenaamde *persoonsgegevens*, waarvoor een speciaal juridisch regiem geldt. Dat kunnen vrijwillig aangeleverde data zijn (bv. ingevulde onlineformulieren), geobserveerde data (bv. door software of sensoren uitgelezen gedragsgegevens) of daaruit afgeleide gegevens (bv. een profiel voor de kredietwaardigheid). Andere data betreffen het beheer van de levenscyclus van producten, transport of kritische infrastructuur (bv. metaalmoeheid, waterstanden of klimaatverandering). De term 'big data' verwijst intussen steeds naar machinaal leesbare digitale informatie die door computersystemen kan worden verwerkt en direct is verbonden met technieken die het doorzoeken en analyseren mogelijk maken van grote hoeveelheden data die niet noodzakelijk op voorhand zijn gesorteerd. Dit zijn de zogenaamde '*analytics*', analyses van digitale gegevensbestanden met behulp van *digitale analysetechnieken, rekenschema's of algoritmes*. In deze context spreekt men ook van '*data mining*' of gegevensontginning, naar analogie met de ontginning van andere grondstoffen.

Soms wordt gesuggereerd dat big data zoveel gegevens omvat dat de analyses geen fouten meer bevatten. Mogelijke fouten zouden als het ware worden weggefilterd doordat men – statistisch gezien – eigenlijk alle data heeft die relevant zijn. Dat is een misverstand. Zo is het vaak lastig om relevante data te bemachtigen. Het verkrijgen ervan kan duur zijn, technisch onmogelijk of stuiten op privacybezwaren, bedrijfsgeheimen of intellectuele eigendom. De verleiding is dan groot om te werken met data die gemakkelijk verkregen worden (zogenaamd '*low hanging fruit*'), maar die kunnen irrelevant of incompleet zijn, of vooroordelen (vertekening of *bias*) bevatten, wat al snel tot foutieve of irrelevante voorspellingen leidt. Soms zijn deze data enkel maar sporen van of verwijzingen naar feiten, en geen echte metingen van het fenomeen dat men wil bestuderen. Verder werkt men per definitie met data uit het verleden of heden (*streaming*); toekomstige datapunten zijn altijd voorspellingen. De vraag waar het dan ook altijd om gaat is: in hoeverre zijn de beschikbare data representatief voor nieuwe of toekomstige data? Die vraag kan alleen worden beantwoord in het licht van het

doel waartoe machinaal leren wordt toegepast. Er bestaat niet zoiets als 'de juiste representatie van de werkelijkheid door middel van data'. Wie belastingfraude wil voorspellen, zou data moeten verzamelen die het mogelijk maken om een onderscheid te maken tussen toekomstige fraudeurs en brave betalers. Alleen: het is niet bekend welke data zo'n onderscheid mogelijk maken, en dat verleidt sommigen wellicht om zomaar te beginnen met de data waar de belastingdienst de hand op kan leggen (*low hanging fruit*). Bijkomend punt is dat het ook niet bekend is welke belastingfraude tot nu toe onder de radar is gebleven. Zo wordt het een hachelijke zaak om steeds meer datapunten te verzamelen en daarin patronen te zoeken; de inbreuk op de privacy wordt steeds groter en succes is niet verzekerd. Mogelijk speelt hier ook nog een *selffulfilling prophecy* een rol. Wanneer wordt ingezoomd op degenen waarvan op enig moment bekend is dat zij frauderen, veronderstelt men ten onrechte dat de relevante patronen ook daar worden gevonden [Harcourt 2007].

Machinaal leren

Machinaal leren is een subdiscipline van de computerwetenschappen die voor een revolutie heeft gezorgd in de kunstmatige intelligentie. Om de impact van big data te verstaan is een goed begrip van machinaal leren cruciaal. De meest eenvoudige maar heldere definitie [Tom Mitchell 1997]) luidt: '*We zeggen dat een machine leert met betrekking tot een specifieke taak T, prestatiemaatstaf P en type ervaring E, als het systeem de eigen prestatie P, ten aanzien van taak T, naar aanleiding van ervaring E, op betrouwbare wijze verhoogt.*'

Belangrijk is dat computersystemen die kunnen leren, in staat moeten zijn 'ervaringen' op te doen, zodat ze de invloed van hun eigen gedrag kunnen doormeten en dat gedrag naar aanleiding daarvan kunnen bijstellen. Dat gaat volautomatisch op basis van zogenaamde algoritmes. Een *algoritme* is een stappenplan, een soort recept of instructie waarmee computers uit de voeten kunnen. Bij machinaal leren bestaan die stappenplannen deels uit wiskundige functies die verbanden zoeken tussen verschillende datapunten, of profielen zoeken van personen of groepen van personen, of organisaties. Zo kan bijvoorbeeld blijken uit de data van een kredietverlener dat personen die na 12 uur 's nachts een bestelling doen, veel vaker verzuimen de rekening te betalen. Een webshop die deze kredietverlener inschakelt kan op grond daarvan besluiten dat personen die dit gedrag vertonen niet achteraf mogen betalen. Om dit verband te ontdekken heeft zo'n kredietverlener data nodig om het algoritme te trainen: de zogenaamde 'trainingset'. Als het goed is, worden de gevonden patronen regelmatig getest (en eventueel opnieuw getraind bij vermindering van de performantie) op nieuwe data; die vormen de testset. Een waardevol algoritme zal dus in staat zijn goed te *generaliseren*, m.a.w. ongeziene data correct te classificeren. De betrouwbaarheid van machinaal leren hangt in belangrijke mate af van de relevantie en compleetheid van de trainingset en de testset, en daarnaast ook van de algoritmes die getraind worden om voorspellingen te doen en van de snelheid waarmee resultaten worden verwacht.

We kunnen twee valkuilen aanwijzen. Enerzijds kan het voorkomen dat algoritmes heel gedetailleerde verbanden vinden die de trainingsset vrij nauwkeurig in kaart brengen, maar minder goed voorspellen welke verbanden in een volgende set voorkomen. Dit wordt '*overfitting*' genoemd: de '*fit*' tussen data en wiskundig model is te 'goed', waardoor generalisering niet goed mogelijk is. Anderzijds kunnen de verbanden heel algemeen zijn, maar gaan ze voor individuele gevallen vaak niet op. Dat heet '*overgeneralization*'. Tussen de omvang, volledigheid en relevantie van de trainingssets, de snelheid waarmee het resultaat kan worden gegenereerd en de gedetailleerdheid van de verbanden zijn altijd afwegingen (*trade-offs*). Je kunt nooit op alle fronten goed scoren. In de praktijk worden dan ook allerlei keuzes gemaakt die de resultaten negatief kunnen beïnvloeden. Het is dus zaak steeds goed te kijken naar het doel van het gebruik van machinaal leren in de toepassing. Zeker waar het gaat om beslissingen ten aanzien van personen, maakt het veel uit of de voorspellingen zijn genomen op basis van een voldoende rijke dataset, dan wel of ze maar een heel grove indicatie geven van mogelijk toepasselijke verbanden. In het voorbeeld van de kredietverlening is het denkbaar dat, wanneer er extra data beschikbaar komen, het verband tussen nachtelijke bestelling en wanbetaling gerelativeerd moet worden, bijvoorbeeld omdat dit verband alleen bij mannen geldt. In het voorbeeld van de belastingdienst kan het zijn dat een open oog voor nog onbekende fraudegevallen en het stellen van prioriteiten bij de aanpak van deze of gene fraude steeds tot nieuwe inzichten zullen leiden.

Nood aan kwaliteitsvolle data

Gecontroleerd en ongecontroleerd machinaal leren. Uiteraard zijn er verschillende vormen van machinaal leren, zoals er diverse types van algoritmes voorhanden zijn. Het is niet altijd evident om te achterhalen welke vorm van machinaal leren en welke types *algoritme*s de beste zijn. Op dit moment implementeren de meeste systemen voor machinaal leren wiskundige formules die het verband tussen de voorhanden zijnde ingang en de gewenste uitgang (input en output) van een systeem wiskundig proberen te beschrijven en ook te optimaliseren. Bij iedere vorm van machinaal leren is een zogenaamde hypotheseruimte aan de orde, ook al kan het dat de hypotheses mede door de software worden ontwikkeld. Het is van belang om data-gestuurde *praktijken in te bedden in empirisch en theoretisch gestuurde praktijken*, waarbij data-gestuurd onderzoek niet de hele methode bepaalt. Zo kan worden voorkomen dat men vaart op data en algoritmes, in plaats van op feiten en inzichten. Er is een neiging om de twee te verwarren, alsof data feiten zijn en algoritmes een soort 'wonderolie'. Dat is zeker niet het geval. Data zijn sporen van, verwijzingen naar of representaties van feiten. Niet meer en niet minder. Zoals eerder aangegeven, is het niet zeker is of de data up-to-date, compleet en/of relevant zijn. Juist vanuit de computerwetenschappen is het bewustzijn van dit alles groot. Helaas is het voor beleidmakers, adverteerders en allerhande leveranciers van diensten niet eenvoudig de methodologische valkuilen

te vermijden die het nut van data-gestuurde praktijken besmetten.
Het mag duidelijk zijn dat er veel en kwaliteitsvolle data nodig zijn om goede beslissingen te nemen, op basis van betrouwbare verbanden. Daarmee raken we aan drie andere aandachtspunten:

- wanneer het gaat om beslissingen ten aanzien van personen, moeten al snel veel persoonsgegevens worden verwerkt, mogelijk uit heel verschillende contexten. Dat kan vervolgens leiden tot gedetailleerde profielen van individuen, die een grote inbreuk vormen op hun privacy;
- ook wanneer die profielen op een abstract niveau blijven, bijvoorbeeld omdat ze zijn afgeleid uit geaggregeerde gegevens, kunnen ze bij toepassing wel degelijk een grote impact hebben op de persoonlijke levenssfeer;
- in gevallen waar de rechten en vrijheden van individuen aangetast kunnen worden, kan een betekenisvol en systematisch optreden van een fysiek persoon noodzakelijk blijven bij het nemen van beslissingen op basis van data (waar toch altijd een mogelijkheid bestaat op verkeerde beslissingen). Zo kan machinaal leren gebruikt worden om dagelijks een eerste screening te maken van miljoenen kredietkaartoperaties om daaruit de potentieel frauduleuze uit te halen, die dan verder manueel onderzocht worden. Niet alleen machines leren: ook de mensen die ermee werken moeten leren wanneer machinale beslissingen tot onterechte inbreuken leiden.

2.2 Probleemverkenning

In dit onderdeel worden problemen en mechanismen besproken die vaak onder de radar blijven, maar die de gebruiker sterk kunnen beïnvloeden of onbewust sturen, zoals onder meer: vertekening of *bias*, *choice architecture*, 'zacht duwtje' of *nudging*, AB testing, beïnvloeding van de consument, *tracking* en *search engine advertising*.

Vertekening of bias

Een van de problemen die big data en machinaal leren oproepen is de mogelijke vooringenomenheid van (1) de data, (2) de analysetechnieken en/of (3) de uitkomsten van de analyse. In de literatuur wordt dit probleem meestal gevat onder de noemer vertekening of '*bias*': veronderstellingen of vooroordelen die in een bepaalde richting wijzen en zo de zogenaamde 'hypotheseruimte' zowel mogelijk maken als inperken. Een databestand zonder *bias* (met een willekeurige of random-verdeling) bestaat niet (tenzij het zo is gefabriceerd, maar zelfs dat is niet eenvoudig). Het gaat er altijd om of de *bias* relevant en betrouwbaar is, en in het verlengde daarvan of de *bias* wellicht voortkomt uit een problematische maatschappelijke verdeling, bijvoorbeeld naar inkomen, opleiding, strafblad, gezondheid… De hypotheseruimte is een geheel van wiskundige functies waarmee patronen in databestanden kunnen worden opgespoord; zonder hypotheseruimte kan geen toegevoegde waarde uit data worden afgeleid. Zo'n ruimte kan heel

eenvoudig zijn en maar een paar makkelijk te lokaliseren verbanden zoeken (bijvoorbeeld alle lineaire correlaties tussen de data) ofwel is hij heel complex (door ook niet-lineaire verbanden of achterliggende causale relaties te zoeken). Tussen de lengte van de voeten van kinderen en het niveau van hun algemene ontwikkeling is er geen causaal verband, maar wel een correlatie, die dan ook door andere factoren wordt veroorzaakt. Dit voorbeeld begrijpt iedereen, maar wanneer het gaat om andere verbanden is de verleiding groot om de correlatie te behandelen alsof het om oorzaak en gevolg gaat, ook al kan vaak pas na extra onderzoek worden vastgesteld of er verbanden zijn en hoe ze liggen. Denk aan correlaties tussen eetgewoontes en obesitas, of tussen een genetisch profiel en de vatbaarheid voor ziektes. Dergelijke correlaties hangen af van een complex samenspel van oorzaken; het nemen van beslissingen op basis van een simpele correlatie kan daarom zowel gevaarlijk zijn (als de eigenlijke oorzaak buiten beeld blijft) als nutteloos (en tot verspilling van middelen leiden).

Zoals hiervoor bij de begripsverkenning van machinaal leren is uiteengezet, moeten bij het afleiden van nieuwe inzichten uit databestanden allerlei beslissingen worden genomen die neerkomen op een afweging. Bijvoorbeeld over de omvang van de databestanden, hun relevantie, compleetheid en juistheid, het soort datapunten en/of het format van de data. Al deze beslissingen hebben financiële implicaties (meer en betere data kunnen simpelweg te duur zijn of niet haalbaar), maar ze hebben uiteraard ook gevolgen voor de betrouwbaarheid van de uitkomsten (als relevante datapunten geen deel uitmaken van de dataset, of niet worden gespot door de ontworpen hypotheses). Afwegingen tussen kosten, de snelheid van het verkrijgen van de resultaten en de betrouwbaarheid ervan zijn onvermijdelijk waar het gaat om concrete toepassingen. Wie alle mogelijke data bij elkaar legt en met zeer complexe algoritmes doorzoekt (grote hypotheseruimte), zal tot een zeer gedetailleerde beschrijving van de data komen. Zoals hierboven besproken, kan die beschrijving zo precies zijn dat de gevonden patronen niet generaliseerbaar zijn naar andere data. Wie het probeert toe te spitsen op een meer generieke beschrijving, kan wellicht tot betere voorspellingen komen, maar ook in dat geval zullen die statistisch van aard zijn. Dat betekent bijvoorbeeld dat gedragsprofielen wel opgaan voor de gemiddelde persoon die onder het profiel valt, maar waarschijnlijk niet voor de concrete personen. Als iemand in een profiel past dat gemiddeld 70% kans op darmkanker heeft, betekent dit niet dat zij *dus* 70% kans op darmkanker heeft. Dit heeft te maken met de distributie van patronen in databestanden, die pas interessant worden als ze afwijken van het gemiddelde. Mocht zij verwanten hebben met darmkanker, dan kan het dat haar kans boven de 70% ligt; mocht zij bejaard zijn zonder dat de kanker zich heeft gemanifesteerd, dan zou haar kans wel eens onder de 70% kunnen liggen. De afwijking van een willekeurige (random-)distributie is de productieve *bias* die het überhaupt mogelijk maakt om patronen te onderscheiden. Zoals filosofen [bv. Gadamer 2010] en wetenschappers [bv. Wolpert 2013] al sinds jaar en dag opmerken, is die *bias* de mogelijkheidsvoorwaarde voor het maken van de

onderscheidingen die kennis en inzicht mogelijk maken. *Bias* is dus niet alleen onvermijdelijk, maar vormt in zekere zin de grond waar alle waarneming en cognitie op staat. Dat betekent echter niet dat *anything goes*. Onzorgvuldig voorbereide en slordig getoetste *bias* levert onhoudbare resultaten op die een verkeerd beeld van de werkelijkheid opleveren. In sommige gevallen is dat gevaarlijk (denk aan kritische infrastructuur), in andere gevallen kan het leiden tot ongerechtvaardigde discriminatie (denk aan het gebruik van software door Amerikaanse rechters om de hoogte van de straf te bepalen, waarbij dezelfde statistische significantie bij zwarte daders tot een verhoging, en bij blanke daders tot een verlaging van de straf leidt [Angwin 2016]).

Verboden of moreel niet te verantwoorden bias Het feit dat *bias* onvermijdelijk en productief is, sluit dus niet uit dat een specifieke *bias* tot verboden of moreel niet te verantwoorden discriminatie leidt. Wanneer de verdeling van het inkomen tussen mannen en vrouwen ongelijk is omdat vrouwen ten onrechte minder betaald krijgen voor hetzelfde werk, zal het databestand waar een algoritme op wordt getraind een *bias* vertonen die terugkeert in de resultaten. Stel dat iemand wil onderzoeken of vrouwen net zo capabel zijn als mannen en het gemiddelde inkomen als maatstaf neemt. Een algoritme dat op de correcte data wordt getraind zal aangeven dat vrouwen minder capabel zijn. Het is dan ook zaak steeds in de gaten te houden of de databestanden zelf een *bias* vertonen die zichtbaar moet worden gemaakt vooraleer men beleid baseert op onjuiste vooronderstellingen. Daartoe zijn inmiddels technieken ontwikkeld, zoals *discrimination aware data mining* [Berendt and Preibusch 2014]. In het voorbeeld gaat het om de vooronderstelling dat wie meer verdient meer capabel is. Dit type vooronderstellingen is voortdurend aan de orde en vraagt om een waakzaam oog voor achterliggende verbanden die indirecte verboden discriminatie in de hand werken of moreel niet te verantwoorden zijn.[1]

Choice architecture, 'zacht duwtje' of nudging, AB testing

In advertising, marketing en beleidssferen wordt steeds meer gedacht in termen van *choice architecture* en *nudging*. De gedachte is hier dat het mogelijk is om de keuzes die personen maken als het ware voor te sorteren, zodat de kans dat ze de 'gewenste' keuze maken toeneemt. Dat kan bijvoorbeeld door de gewenste keuze als standaardinstelling (bij verstek of *default*) voor te stellen. Men kan daar weliswaar van afwijken, maar de ervaring leert dat de meeste mensen die moeite niet nemen. Het maakt daarom nogal uit of de standaardinstellingen van laptops, smartphones, slimme energiemeters of sociale netwerken het delen van persoonsgegevens minimaliseren (met een opt-*in* voor verdere verwerking) of juist maximaliseren (met een opt-*out* voor verdere verwerking). Standaardinstellingen

[1] Over de gevolgen van het zich verlaten op machinaal lerende systemen in de medische diagnostiek zie [Cabitza, 2016].

die dataverwerking minimaliseren zijn een vorm van gegevensbescherming 'bij verstek'. Bedrijven van wie het verdienmodel afhangt van de verwerking van grote hoeveelheden gedragsgegevens zullen geneigd zijn een *choice architecture* aan te bieden die standaard toelaat alle gegevens te verzamelen. Overheden die menen dat zij dankzij big data allerlei problemen kunnen oplossen, zullen ertoe geneigd zijn om zo ruim mogelijke bevoegdheden te scheppen voor het onderscheppen, opvragen en/of hergebruiken van gegevens.

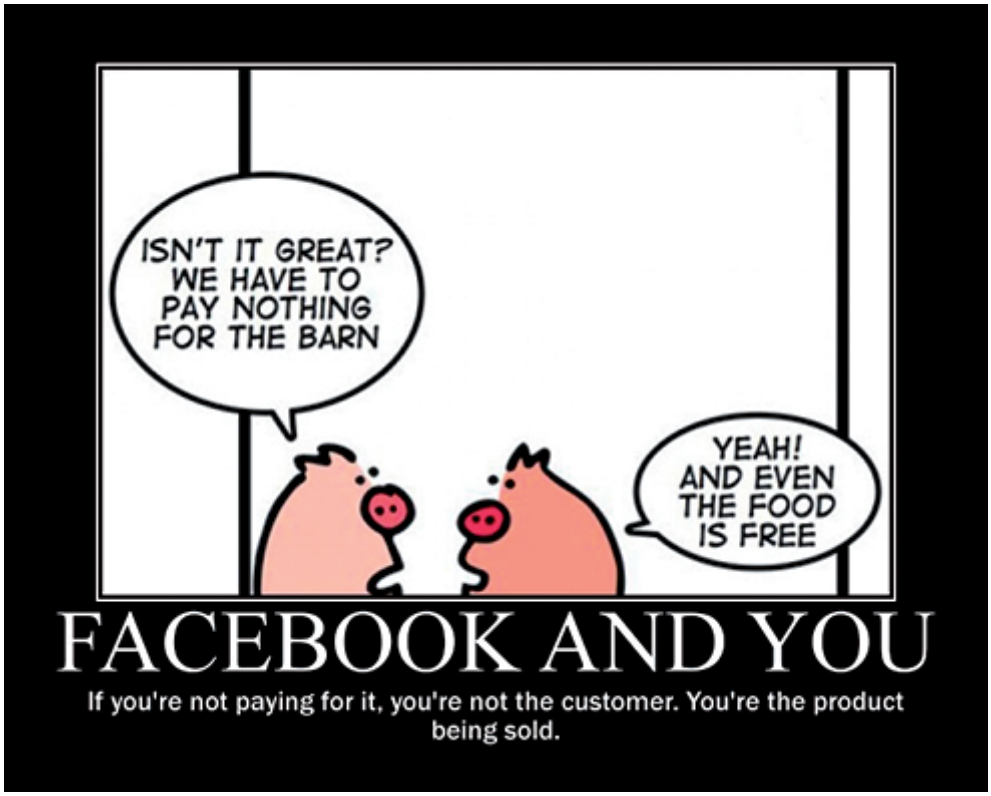
Wanneer machinaal leren wordt gecombineerd met een zacht duwtje of *nudging* kan er gemakkelijk een *choice architecture* worden gebouwd die burgers en consumenten onbewust verleidt tot het delen van ongezien grote hoeveelheden data. *Nudging* is een begrip uit de sociale psychologie en de gedragseconomie dat ervan uitgaat dat mensen zich vaak *irrationeel* gedragen op een *voorspelbare* manier. Wie eenmaal doorheeft welke irrationele neigingen ons gedrag beheersen kan daar handig gebruik van maken. Daarmee verlaat *nudging* de onhoudbare vooronderstellingen van de zogenaamde rationele keuzetheorie, die lange tijd *en vogue* was binnen de economische en beleidswetenschappen – maar blijft men intussen wel denken in termen van rationaliteit en nutsmaximalisatie.

Een andere manier om zo veel mogelijk gedrag zo effectief mogelijk te beïnvloeden is *AB testing*. Dit is inmiddels een veelgebruikte methode om te achterhalen welke opmaak van websites de beste resultaten behaalt. Men stelt zich een website voor die 'geoptimaliseerd' moet worden, en noemt die versie A. Vervolgens passen we de website op een punt aan (nieuwe keuzeknop, andere kleurverdeling, sneller doorklikken, ander taalgebruik). Dat is versie B. Daarna sturen we de ene helft van de bezoekers naar versie A en de andere helft naar versie B, en meten het klikgedrag om te bekijken welke versie tot gewenst gedrag leidt: meer aankopen, beter lezen, dieper doorklikken. We kiezen vervolgens de versie met de gewenste output. Dit kan voortdurend worden herhaald en doordat het bezoekersgedrag machinaal leesbaar is, kan snel worden doorgerekend hoe de bezoeker succesvol kan worden beïnvloed. *AB testing* draagt dus bij aan het ontwerp van de door de opdrachtgever gewenste *choice architecture*.

Interessant is echter dat het juridische kader inzake de gegevensbescherming eisen stelt aan de *choice architecture* en de mogelijkheden inperkt om personen als het ware achter hun rug om te verleiden tot het delen van hun gedragsgegevens. Het gaat daarbij om gegevensbescherming *by default* ('bij verstek') en *by design* ('bij ontwerp'). De vraag is dan ook niet *of* we een *choice architecture* willen ontwerpen maar *welke*.

Beïnvloeding van de consument

Het internet en sociale media zijn onlosmakelijk verbonden met de huidige consument en hoe deze beïnvloed wordt. Mensen geven online enorme



http://geek-and-poke.com

hoeveelheden persoonsgegevens prijs. Dit gebeurt door middel van hun eigen surf-, zoek-, *like*- en klikgedrag, en door de aanmaak van profielen op sociale media, maar ook door allerhande perifere data die een apparaat vrijgeeft (bv. locatie, geluid…). Bovendien zit e-commerce stevig in de lift. Meer en meer zoeken, vergelijken en kopen mensen producten en diensten in een onlineomgeving. Deze diverse vormen van persoonsgegevens worden door bedrijven gretig gebruikt om (potentiële) consumenten te benaderen met gepersonaliseerde reclame en promoties, vaak zonder dat die consumenten dat beseffen. Deze consumentenbeïnvloeding gebeurt opnieuw voornamelijk via onlinekanalen, zoals websites, mobiele applicaties en sociale media. Veel van deze websites en platformen zijn gratis – de consument 'betaalt' met gedragsgegevens – en reclame vormt hun belangrijkste bron van inkomsten. Verder worden deze data door marketeers gebruikt om trends en nieuwe markten in kaart te brengen. Door big data en bijbehorende *data mining*-technieken denken bedrijven meer dan ooit te weten wat 'hun' consument voelt, denkt, wenst en koopt, en kunnen ze op basis van deze inzichten nieuwe of verbeterde producten en diensten ontwikkelen.

Het is dus niet verwonderlijk dat met de massale adoptie van internet en sociale media en de bijbehorende big data de reclame- en marketingwereld een hele reeks transformaties heeft ondergaan.

Tracking

Tracking duidt op het verzamelen van persoonsgegevens van onlinegebruikers door websites en bedrijven. We onderscheiden *first-party* en *third-party tracking*. Het eerste gaat over de data die een bedrijf/organisatie zelf verzamelt over haar gebruikers (dit kan een e-commerce-site zijn, maar ook een ander type website, zoals een nieuwswebsite of een publieke organisatie). Zo wordt bijvoorbeeld bijgehouden wie wanneer een website bezoekt, welke aspecten van de website worden bekeken, aangeklikt, welke producten worden gekocht enzovoort. Een organisatie kan deze data vervolgens gebruiken om gericht naar de gebruikers te adverteren, aanbevelingen te maken bij volgende bezoeken of de ervaring van de eigen website te verbeteren (al dan niet met *AB testing*). Een organisatie kan deze data ook verkopen aan derde partijen. In dat geval kunnen ook zogenaamde *third-party trackers* actief zijn op een website. Dat zijn veelal zogenaamde advertentienetwerken en *data brokers* die op allerhande websites mensen volgen en die zodoende over zeer volledige gegevens beschikken over iemands onlinegedrag. Dit maakt deze *data brokers* erg machtig, vooral wanneer zij de onlinedata ook nog eens koppelen aan offlinegedrag, zoals bijvoorbeeld aankopen in de supermarkt of financiële transacties, of wanneer deze data gekoppeld worden aan locatie- en gezondheids- en lifestylegegevens die zijn verzameld via zogenaamde *wearables*, (bv. smartphones, smartwatch, *activity trackers*, mobiele applicaties). Door het samenbrengen van al deze datastromen kunnen *data brokers* bijzonder veel in kaart brengen over het doen en laten van een individu, ook gevoelige en hyperpersoonlijke data in verband met gezondheid, seksualiteit, politieke voorkeuren, financiën… Deze verzamelde data zijn extreem waardevol voor adverteerders en marketeers en worden bijgevolg duur verkocht. Samengevat [Wareable]:

'*They know all about you. They know who you are and where you live, where you work and how you worship, what magazines you read and what websites you visit, what books you love and bands you loathe, what you earn and what you save, what you like to eat and do and say and see and buy. They're the data brokers, and your business is their business.*'

Gedragsgestuurd adverteren wordt dus mogelijk gemaakt door *tracking*. Welke (banner)reclame iemand te zien krijgt op een website, wordt meer en meer bepaald op basis van het persoonlijke profiel en onlinegedrag van een websitegebruiker. Een belangrijke reclamevorm is de zogenaamde *retargeting*. Dit zijn (veelal) bannerreclames van producten of diensten die een gebruiker eerder bekeek of aankocht. Met andere woorden: deze vorm van reclame is gebaseerd op eerder

surfgedrag. *Retargeting* komt voor op allerhande websites en is een zeer populaire reclamevorm op sociale netwerksites als Facebook. Hoewel recente studies aantonen dat deze vorm van reclame zeer effectief is (mensen zijn meer geneigd te klikken en over te gaan tot aankoop) en vaak als relevanter wordt beschouwd dan 'willekeurige' reclame, zien we tegelijk dat de kennis over de exacte werking van *retargeting* laag is. Wanneer mensen worden ingelicht over het gebruik van voorgaand surfgedrag en de bijbehorende onlinemonitoring, dan verhoogt hun kritische houding en bezorgdheid over privacy ten aanzien van deze reclamevorm.

Populaire websites, zoals bijvoorbeeld onlinekranten, verkopen reclameruimte aan adverteerders, die zo hopen de juiste doelgroep te bereiken. Dit kan op een directe manier gebeuren, zoals bij de klassieke media. Een adverteerder koopt dan reclameruimte op een bepaalde nieuwswebsite omdat het profiel van zijn potentiële klanten aansluit bij de typische lezer van die krant. Meer en meer wordt onlinereclameruimte echter verkocht volgens het principe van *real time bidding.* Dit wil zeggen dat op het moment dat een gebruiker een website bezoekt de zogenaamde advertentieruimte dankzij een geautomatiseerde veiling wordt verkocht aan de hoogst biedende adverteerder. De adverteerder krijgt de persoonsgegevens niet in handen, maar kan dankzij de tussenkomst van een 'advertentienetwerk' bieden om de banner te mogen plaatsen bij een gebruiker met een specifiek profiel, waarna bijvoorbeeld betaald moet worden per '*impression*' (het aantal keren dat de advertentie zichtbaar is voor websitebezoekers) of per 'klik' (het aantal keren dat de bezoekers doorklikken op de advertentie). Zie bijvoorbeeld [Google], [Facebook] of [Coursera]. Bij een aantal mobiele apps gaat men nog een stap verder en worden persoonlijke gegevens, zoals telefoonnummers, locatie- en gezondheidsgegevens, doorgezonden naar adverteerders; dit is mogelijk omdat app-developers softwarebibliotheken met advertentiesoftware aan hun app toevoegen om geld te verdienen. Daarnaast is aangetoond dat die informatie vaak onbeschermd wordt doorgestuurd en ook door derde partijen, zoals telecommunicatiebedrijven en overheden, onderschept kan worden [Demetriou 2016], [Vanrykel 2016].

Op deze manier wordt getracht het hele reclamegebeuren op een website of in een app te personaliseren naar het onlinegedragsprofiel van de individuele gebruiker. Dat alles gebeurt in fracties van seconden, zonder dat de gebruiker dit merkt. In hoeverre dit soort 'gedragsadvertenties' effectiever is dan bijvoorbeeld contextuele advertenties, die afhangen van het soort site waarop wordt geadverteerd, is onduidelijk.

Search engine advertising

Een vergelijkbare, veelgebruikte vorm van onlineconsumentenbeïnvloeding is *search engine advertising*. Hierbij betalen bedrijven de onlinezoekmachines, zoals Google, om hun zoekresultaten bovenaan of op een andere prominente

plaats te tonen, zodat de kans vergroot dat mensen erop klikken en dus bij het bedrijf terechtkomen. Reclame gebaseerd op eerdere zoekopdrachten kan ook terugkomen op andere websites. *Search engine advertising* wordt beschouwd als een vorm van *native advertising*. Dit is onlinereclame die er qua vorm en opbouw uitziet als de niet-commerciële inhoud van een website, zoals bijvoorbeeld een reeks zoekresultaten, een redactioneel artikel (bij onlinekranten) of inhoud geplaatst door andere gebruikers (bijvoorbeeld op de newsfeed van een sociale netwerksite). *Native advertising* omvat echter gesponsorde (lees: betaalde) inhoud door een bedrijf en is moeilijk van de oorspronkelijke inhoud van een website te onderscheiden. De opkomst ervan is mede een reactie op de dalende reclame-inkomsten van traditionele reclamebanners en de massale adoptie van *ad blocking*-software door onlinegebruikers. De *native advertising* die iemand te zien krijgt, wordt meer en meer op een doorgedreven manier aangepast naargelang van het persoonlijke onlineprofiel, de interesses, de zoekacties, het aankoopgedrag en andere individuele, sociale en contextfactoren van een gebruiker [Working party 2010].

Een belangrijke kanttekening is dat de gegevens verzameld door *data brokers* en de gepersonaliseerde inzet ervan door advertentienetwerken niet vrij zijn van problematische vooroordelen en zelfs tot ongewenste discriminatie kunnen leiden. Adverteerders hebben een bepaald idee van hun doelgroep en van wat mensen drijft. Zo zullen vrouwen die zoektermen in verband met zwangerschap of een kinderwens intypen, mogelijk meteen in een categorie van 'aanstaande moeder' terechtkomen en ongevraagd overspoeld worden met reclame die daarop is gericht. Onderzoek toonde verder aan dat zoekacties van persoonsnamen die meer voorkomen bij zwarten tot andere vormen van gepersonaliseerde reclame leiden dan bij persoonsnamen die meer voorkomen bij blanken. Zo leverden de namen die over het algemeen worden geassocieerd met personen met een donkere huidskleur significant meer gepersonaliseerde zoekmachine-reclame op over negatieve zaken, zoals bijvoorbeeld arrestaties, dan bij de 'blank' klinkende voornamen [Sweeney 2013]. Dit is uiteraard problematisch wanneer op persoonsnamen wordt gezocht voor het bekijken van iemands (professioneel) profiel bij een sollicitatie of andere relevante sociale interacties.

Machtige commerciële spelers; ongewapende gebruikers

In deze opsomming van onlineconsumentenbeïnvloeding is het belangrijk stil te staan bij enkele zeer belangrijke, machtige spelers die verschillende sleutelrollen vervullen. Google is een van de meest prominente voorbeelden. Als belangrijkste zoekmachine ter wereld is Google geleidelijk geëvolueerd tot de machtigste speler in het onlinereclamelandschap. Het bedrijf heeft voor een groot deel bepaald hoe consumentenbeïnvloeding online verloopt. Google heeft een gesofisticeerd en zeer lucratief model van *search engine advertising* waarbij het bedrijf betaalde zoekwoorden en betalingen per klik hanteert: veelgezochte woorden worden duur

verkocht aan adverteerders die betalen per klik van een gebruiker. Verder maken ze ook gebruik van biedingen bij het intikken van zoektermen. Als een gebruiker bijvoorbeeld 'zomervakantie in Spanje' intikt, zullen verschillende reisorganisaties bieden om hun advertentie in de vorm van een zoekresultaat als eerste te laten verschijnen [Rathenau Inst. 2010]. Hier komt nog bij dat Google sinds 2008 het belangrijkste onlineadvertentienetwerk Double Click in handen heeft. Double Click is een van de grootste *third-party trackers* en beheert de plaatsing en verkoop van onlineadvertentieruimte via persoonsgegevens en browsergeschiedenissen op talloze websites. Sinds 2016 heeft Google zijn privacybeleid aangepast en maakt het bedrijf het mogelijk om de data van Double Click te koppelen aan de gegevens van de eigen zoekmachine *én* aan persoonlijke accounts van Googlegebruikers (bv. via hun emailservice Gmail). Hierdoor combineert Google zowel *first-party tracking* via de eigen website en diensten met die van een grootschalig *third-party tracking*systeem via de integratie van de gegevens van Double Click (die daardoor in feite *first-party tracker* is geworden). Ook andere grote spelers, zoals Facebook, Amazon en Apple, hebben zeer veel data in handen en dus veel macht op het gebied van consumentenbeïnvloeding. Zo kan Facebook gebruikers volgen op het eigen platform, maar ook via alle websites die mogelijkheden tot het delen en volgen op Facebook aanbieden (via zogenaamde *social plugin* in de vorm van bv. de *like*-knop).

De vraag rijst waarom gebruikers zo massaal persoonlijke, waardevolle data (blijven) delen, en die dus ogenschijnlijk vrijwillig vrijgeven aan commerciële partijen. Ten eerste zijn zij zich hiervan doorgaans niet bewust. Er wordt vaak gesteld dat mensen bij het vrijgeven van hun persoonsgegevens een *privacy calculus* maken. Ze moeten met andere woorden afwegen of het vrijgeven van persoonsgegevens opweegt tegen wat ze ervoor terugkrijgen: meer gepersonaliseerde, relevante aanbiedingen, gratis nieuws of andere informatie of vermaak. Omdat mensen geneigd zijn vooral de directe voordelen te zien en niet weten dat hun data ook nog een eigen leven gaan leiden, kiezen ze er (mogelijks) voor om ze vrij te geven. Door het gebrek aan transparantie en de onoverzichtelijkheid van de hoeveelheid en de aard van *tracking* die er gebeurt, is het voor mensen rationeel onmogelijk om een echte calculus te maken. Meestal maakt men de calculus niet bewust of onvolledig. Verder wordt ook vaak de *privacy paradox* genoemd: ook al geven mensen aan dat ze zich zorgen maken om hun privacy online, toch zijn ze er vaak op de cruciale momenten niet mee bezig, bijvoorbeeld bij het aanmaken van een persoonlijk profiel voor sociale contacten via sociale media, het verrichten van onlineaankopen, hun uitgebreide en zeer persoonlijke zoekgedrag, aantrekkelijke onlinepromoties en -wedstrijden, leuke spelletjes... Het plezier, de sociale behoeften of het ervaren nut van de verkregen informatie overheerst. Het vrijwaren van hun privacy weegt daar niet tegen op, op dat moment toch niet.
Het lijkt er dus op dat mensen, mede door het manipuleren van cognitieve en emotionele beperkingen (gebrek aan kennis en transparantie over de hoeveelheid en de aard van dataverzamelingen, de directe aantrekkelijkheid

van de onlineomgeving en het aanbod), niet voldoende gewapend zijn tegen deze dataverzamelingspraktijken en de bijbehorende en alomtegenwoordige toepassingen van consumentbeïnvloeding. Dit hangt samen met de hiervoor besproken *nudge*-praktijken, waarmee niet alleen wordt ingespeeld op deze beperkingen, maar juist ook de omgeving wordt gecreëerd waarin de beperkingen zich voordoen.

2.3 Verkenning van oplossingsrichtingen

Wat kan tegen al deze problemen gedaan worden? Een multistakeholderbenadering lijkt het meest aangewezen: alle partijen nemen een verantwoordelijkheid op. Er zijn vooreerst de ICT-methodes van cryptografie en beveiliging en anonimisering. Daarnaast is er educatie die inzet op data- en reclamewijsheid/geletterdheid vanaf jonge leeftijd [website ik beslis], maar ook op oudere generaties. Ten slotte zijn er juridische benaderingen met naast de regelgeving Algemene Verordening Gegevensbescherming AVG nog vier delen: profieltransparantie, doelbinding, het gerechtvaardigde belang van de data controller preventieve acties, onschuldpresumptie bij politie en justitie en objectieve en privaatrechtelijke aansprakelijkheid voor onrechtmatige verwerking.

Cryptografie, achterpoortjes, massasurveillantie, beveiliging

Cryptografie is de tak van de wetenschap die zich bezighoudt met het beveiligen van (digitale) informatie. In een historisch perspectief lag de nadruk op de geheimhouding van *communicatie*, wat betekent dat enkel de gespecificeerde ontvanger de informatie kon lezen. Bij digitale informatie groeit het belang van de bescherming tegen het wijzigen van data (integriteit) en het correct identificeren van afzender en ontvanger. In de context van de bescherming van de persoonlijke levenssfeer wil men ook vaak de metadata beschermen: dit betekent dat men de identiteit en locatie van zender en ontvanger verborgen wil houden voor derden. Het toenemende gebruik van computers heeft geleid tot meer aandacht voor de bescherming van informatie terwijl ze *bewaard* wordt in pc's, tablets, smartphones of in de cloud. Een recente evolutie is dat men informatie ook wil beschermen terwijl er *berekeningen* op uitgevoerd worden. Stel dat een gebruiker op basis van zijn gezondheidsgegevens wil berekenen wat zijn risico is op een aantal ziektes: als gebruiker wil je graag je gegevens vertrouwelijk houden, terwijl de dienstverlener misschien zijn berekeningsmethode of algoritmen wil beschermen. Op het eerste gezicht lijkt dit onmogelijk, maar met behulp van speciale cryptografische algoritmen kan men gevoelige gegevens in vercijferde vorm in de cloud opladen, waarna de dienstverlener er berekeningen op uitvoert. Vervolgens kan men het vercijferde resultaat downloaden en ontcijferen.

Aangezien we evolueren naar een wereld van big data, wordt er een toenemende hoeveelheid informatie verzonden, bewaard en bewerkt. Dit betekent ook dat

persoonlijke informatie steeds meer verspreid wordt, met een groeiend risico op misbruik door andere gebruikers, bedrijven en overheden (‘datavervuiling’). De beste methode om informatie effectief te beschermen en datavervuiling onder controle te houden is met behulp van cryptografie. Er zijn ook grenzen aan wat cryptografie kan bereiken. Als de analyse gebeurt op vercijferde data, blijven data goed beschermd maar bestaat nog steeds de mogelijkheid dat de analyse op zich discriminerend is of de privacy schendt. Om dit af te dekken zijn andere oplossingen nodig die eerder in dit document ter sprake kwamen. En in sommige toepassingen, zoals sociale netwerken, is het essentieel dat je informatie deelt met je peers. Cryptografie kan dan helpen om de informatie te beperken tot enkel de peers die jij uitkiest en om ze te beschermen tegen gebruik door de grote dienstverleners, zoals de operator van het sociale netwerk (bv. Facebook) en de netwerkoperator (bv. Vodafone).

Cryptografie herleidt de bescherming van gegevens tot de bescherming van digitale sleutels. Als Alice een geheim bericht wil sturen naar Bob, moeten zij eerst een geheime sleutel afspreken. Alice zal dan het bericht met behulp van de sleutel vercijferen tot een cijfertekst, die ze naar Bob stuurt. Bob kan met dezelfde sleutel de cijfertekst omvormen tot het juiste bericht en nagaan of het bericht wel degelijk van Alice komt. Zonder de sleutel is het bericht niet te lezen. In grote netwerken zoals het internet is het niet haalbaar om met elke dienstverlener of gebruiker vooraf zo’n geheime sleutel voor elk paar van zender en ontvanger af te spreken. In dat geval biedt publieke sleutelcryptografie een oplossing: de sleutel voor het vercijferen is publiek beschikbaar en alleen de sleutel om de informatie weer leesbaar te maken moet geheim blijven. Je kan dit vergelijken met een brievenbus waar iedereen een brief in kan deponeren. Enkel de eigenaar met de sleutel kan de brievenbus leegmaken.
Tot in de jaren 1980 bleef cryptografie voorbehouden voor militairen, overheden en banken. Vandaag de dag is het een massatechnologie: meer dan 30 miljard toestellen gebruiken cryptografie. De grootste toepassing zijn nog altijd bankkaarten, maar cryptografie zit ook in alle mobiele telefoons en Wifi-netwerken, browsers, smartphones, elektronische identiteitskaarten, paspoorten, toegangsbadges, autosleutels, DVD’s en blu-ray-spelers, chatprogramma’s zoals WhatsApp en iMessage enz.

Omdat cryptografie ook militaire toepassingen heeft, valt deze onder de *dual use*-wetgeving, wat betekent dat in vele landen het gebruik, de import en de export van cryptografie gereguleerd zijn. Tot eind jaren 1980 werd cryptografie bijna uitsluitend in dure en omvangrijke hardware geïmplementeerd, wat de controle vereenvoudigde. Met de komst van het web werd een hele reeks nieuwe toepassingen mogelijk die cryptografie nodig hadden: internetshoppen en -bankieren, muziek downloaden, apps downloaden enz. En omdat de rekenkracht van computers bleef groeien, werd het mogelijk om cryptografie in software te implementeren, wat betekent dat controle over cryptografie veel moeilijker werd.

Er zijn verschillende redenen waarom overheden cryptografie het liefst willen controleren. Een eerste is militair: men wil deze strategische kennis en oplossingen niet in verkeerde handen laten vallen. Een tweede reden zit bij de politiediensten: zij hebben altijd de mogelijkheid gehad om – met toestemming van een onderzoeksrechter – brieven te openen of telefoongesprekken af te tappen. Als beide partijen hun communicatie vercijferen, is dat niet langer mogelijk. Een derde reden zijn de geheime diensten: zij willen informatie verzamelen over andere landen en bepaalde groepen in de samenleving; ook zij zien liever niet dat cryptografie gebruikt wordt.

In de huidige maatschappij is het nog moeilijk denkbaar om het gebruik van cryptografie te verbieden. Ten eerste heeft elk land er strategisch belang bij dat de communicatie van zijn burgers, bedrijven en overheden voldoende beveiligd is tegen criminelen en tegen andere landen. Daarnaast wordt een groeiend aantal kritische sectoren in de economie sterker gedigitaliseerd. Sectoren zoals de elektriciteitsvoorziening, het transport en de gezondheidszorg hebben een heel sterke nood aan veilige systemen en communicatie. Dat kan enkel met cryptografie.

De overheden hebben een aantal strategieën bedacht om dit interne belangenconflict te beheersen. Een eerste oplossing was een strenge controle dan wel een verbod op cryptografie, maar zoals vermeld is dit onmogelijk geworden door de massale omschakeling naar cryptografie in software. Een volgende stap was het opleggen van onveilige cryptografie (bijvoorbeeld korte sleutels), wat betekent dat overheidsdiensten deze systemen wel kunnen breken maar anderen niet. Deze oplossing is problematisch omdat ze burgers en bedrijven blootstelt aan aanvallen van andere landen en de georganiseerde misdaad. Daarnaast evolueert deze technologie heel snel, wat betekent dat wat nu nog enige veiligheid biedt binnen tien jaar compleet onveilig kan zijn. Ondanks de problemen wordt deze aanpak nog altijd gehanteerd. Voor de export van cryptografische hardware uit de EU mag men sleutels van ten hoogste 56 bits gebruiken. Deze keuze is meer dan twintig jaar oud. Vandaag de dag kan een overheid zo'n systeem breken in een paar seconden en een academische onderzoeksgroep kan dit binnen een paar uur doen.

De volgende strategie van de overheid was het opleggen van cryptografische systemen met een achterpoortje, dat toegankelijk is voor de overheid. Zo had men in het begin van de jaren 1990 in de VS de *Clipper Chip*, waardoor de Amerikaanse overheid toegang zou hebben tot vercijferde communicatie. Onder druk van de industrie en de academische wereld is dit voorstel ingetrokken. Dat betekent echter niet dat de overheden de controle hebben opgegeven. Zo weten we uit de Snowden-documenten dat de NSA (National Security Agency, VS) een achterdeur heeft ingebouwd in een gestandardizeerd softwaresysteem voor het genereren van sleutels. Ook kan de overheid op basis van een *security letter* aanbieders van

diensten verplichten om geheime sleutels te geven. Eind 2016 werd in de UK de *Investigatory Powers Act* goedgekeurd. Op basis van *section* 217 van deze wet kan de UK overheid de aanbieders van diensten of producten verplichten om een achterpoort in te bouwen.

Experts zijn het erover eens dat het inbouwen van dergelijke achterpoorten zeer gevaarlijk is: er bestaat namelijk het risico dat een ander land of een criminele organisatie achterpoort ontdekt en gebruikt; op die manier wordt het hele systeem onveilig. In het geval van *Juniper routers* is dit risico al werkelijkheid geworden [Juniper routers trapdoor]. Daarnaast blijft bij dergelijke achterpoorten een massale interceptie mogelijk, waarbij alle gegevens van iedereen bewaard en geanalyseerd worden. Dat is in strijd met het Europees Verdrag voor de Rechten van de Mens en opent de deur naar grootschalig misbruik. Experts twijfelen er ook aan of massale interceptie wel doeltreffend is in de bestrijding van georganiseerde misdaad en terrorisme. Door een gebrek aan transparantie bestaat hierover geen publiek debat.

Een laatste oplossing is het gebruik van malware: dit betekent dat men de computer of telefoon van een doelwit besmet en op deze manier aan de data komt die nodig zijn voor onderzoeken. Hier wordt vaak gesproken van *remote hacken*, wat het mogelijk maakt om gegevens te vatten voordat ze zijn geëncrypteerd of nadat ze zijn gedecrypteerd. Dit gebeurt op basis van daarvoor in de wet vastgelegde bevoegdheden. In principe is dit een betere oplossing. Een probleem is dat het toezicht op het gebruik hiervan zeer moeilijk is. Daarnaast bestaat er een risico op proliferatie. Als de malware ontdekt wordt door anderen, kan ze heel gemakkelijk aangepast worden en gericht tegen andere doelwitten, met inbegrip van de overheden.
Conclusie: het controleren of doelbewust verzwakken van cryptografie is geen goede oplossing. Sterke cryptografische oplossingen zijn nodig om de maatschappij doeltreffend te beschermen. Het alternatief, het op afstand hacken van computersystemen, vereist streng na te leven wettelijke waarborgen.

Juridische benaderingen voor gegevensbescherming bij big data

Op 4 mei 2016 werden de officiële teksten van de regelgeving Algemene Verordening Gegevensbescherming AVG (2016/679) gepubliceerd in het *EU Official Journal* (Document 32016R0679), [AVG 2016], [GDPR, 2016]. De regelgeving trad in werking op 24 mei 2016 en is toepasbaar vanaf 25 mei 2018. De Verordening verzoent twee doelstellingen: een betere bescherming van persoonsgegevens voor particulieren, en meer opportuniteiten voor het bedrijfsleven in de digitale eengemaakte markt door de directe werking van één regelgeving in alle lidstaten. Aanvullende wetgeving op nationaal niveau is alleen toegestaan voor zover daartoe discretionaire ruimte is gelaten en/of voorbehouden zijn gemaakt.

De AVG is met name meer toegesneden op big data dan de huidige Richtlijn. Dat heeft alles te maken met de uniforme tekst die overal van kracht is, waardoor transnationale bedrijven hun bedrijfsprocessen gemakkelijker kunnen inrichten in overeenstemming met de Verordening. Daarenboven zijn de maximale boetes zeer hoog (vergelijkbaar met die in de mededingingswetgeving), namelijk 4% van de wereldomzet. Dit schept de juiste *incentive*-structuur voor stakeholders om rechtmatig te werken. Ten slotte is ook de privaatrechtelijke aansprakelijkheid voor onrechtmatige verwerking beter geregeld, onder meer door het recht van de betrokkenen om hun claim te mandateren aan een ngo.

Kort gezegd: de Verordening biedt inhoudelijk grotendeels dezelfde bescherming als de huidige Richtlijn [EU Data Protection Directive 1995] die nog in voege is tot mei 2018, maar ze is effectiever en meer toegesneden op de massaliteit en complexiteit van big data en machinaal leren. Het gaat bijvoorbeeld om: gemakkelijker toegang tot de eigen persoonsgegevens;

1. zwaardere eisen die worden gesteld aan toestemming;
2. het intrekken van toestemming moet even gemakkelijk zijn als het verlenen ervan;
3. duidelijk begrijpbare informatie over wat er gebeurt met de gegevens,;
4. het recht om persoonsgegevens – onder bepaalde voorwaarden – te laten verwijderen;
5. het recht op gebruiksvriendelijke overdraagbaarheid van de ene verantwoordelijke naar de andere of naar zichzelf (*portability*);
6. de verplichting om de bescherming in te bouwen in het ontwerp van de betrokken computersystemen, ook wel 'gegevensbescherming per ontwerp' genoemd;
7. de verplichting om de standaardinstellingen van deze systemen steeds zo in te stellen dat alleen de noodzakelijke verwerking plaatsvindt (gegevensbescherming 'bij verstek');
8. het stellen van grenzen aan 'profilering' (de geautomatiseerde verwerking van persoonsgegevens met als doel persoonskenmerken te evalueren).

Hieronder worden enkele bijzonder relevante onderdelen van het gegevensbeschermingsrecht besproken, voor zover ze betrekking hebben op big data en machinaal leren.

2. Profieltransparantie

Klassiek ligt het zwaartepunt van de waarborgen die gevraagd worden inzake privacy bij het verzamelen van de persoonsgegevens. Bij big data verhuist dat naar de analyse en het gebruik van persoonsgegevens [WRR 2016]. Dat is met name en vooral het geval als geautomatiseerde beslissingen worden genomen die een belangrijke invloed hebben op de betrokken persoon. Cruciaal is dan – zowel ten opzichte van degene op wie de data betrekking hebben als tegenover

de toezichthouder, de Autoriteit Persoonsgegevens (in België de Commissie voor de Bescherming van de Persoonlijke Levenssfeer, CBPL) – dat er transparantie wordt geboden over drie zaken. Ten eerste moet duidelijk worden gemaakt dat de beslissing is genomen op grond van dit soort analyses (vaak '*profiling*' genoemd), wat de precieze doeleinden of finaliteiten zijn van de verwerking en wie hiervoor de verantwoordelijke is. Ten tweede moet de onderliggende logica in begrijpelijke taal worden uitgelegd. Ten derde moet worden aangegeven wat de voorziene gevolgen zijn van het profileren. Wat de onderliggende logica betreft, is het cruciaal dat er informatie beschikbaar is over de methodologische keuzes die bij de analyse zijn gemaakt. Het gaat dan bijvoorbeeld om de keuze van (trainings)algoritmen en de modelstructuur, eventuele parameters, de variabelen die in overweging worden genomen, het type data dat is gebruikt om te trainen. Dankzij de nodige informatie zullen gegevensverwerkingen zo reproduceerbaar mogelijk zijn (transparantie- en informatieplicht). Het best worden de resulterende beslissingsalgoritmen (met de concrete ingangs-uitgangsrelatie die het resultaat is van de gebruikte methodologie) ook openbaar en begrijpbaar gemaakt, zodat de betrokkene voldoende informatie krijgt om te kunnen begrijpen welke logica ten grondslag ligt aan een beslissing. Belangrijk is ook dat men de nauwkeurigheid/performantie van eventueel gebruikte modellen (op onafhankelijke testdata) zo precies mogelijk weergeeft, waardoor de foutenmarge van de profilering duidelijk wordt. Dit moet de betrokkene in staat stellen om zich indien nodig te verdedigen tegen beslissingen die over hem/haar worden genomen en waarin analyses van big data een rol in hebben gespeeld. Van belang is dat niet van het individu kan worden gevraagd om dit alles zelf boven te spitten en erover met de beslisser in discussie te gaan. Hier ligt een evidente taak voor de toezichthouders, zowel die voor de gegevensverwerking als die voor de consumentenbescherming, of voor spelers uit het middenveld.

Bij deze transparantie moet rekening worden gehouden met het uitgangspunt dat er geen onnodige afbreuk wordt gedaan aan de rechten of vrijheden van anderen, met inbegrip van het zakengeheim of de intellectuele eigendom, en met name aan het auteursrecht dat de software beschermt [Overweging 63, AVG 2016]. Dat probleem is overigens minder ernstig dan het lijkt, want de betrokkene heeft recht op *betekenisvolle* informatie over de gebruikte profielen en algoritmes. Het gaat er juist niet om hem te bezwaren met technische taal waar hij niets mee kan aanvangen. Dat neemt niet weg dat de hierboven besproken discussie over de methodologische integriteit van *profiling* en het onderzoek naar mogelijk onrechtmatige *bias* wel degelijk ook door het betrokken individu gevoerd moet kunnen worden. Dat zou wel eens tot lastige afwegingen ten aanzien van het bedrijfsgeheim en IP-rechten kunnen leiden. In ieder geval zal de Autoriteit Persoonsgegevens onder de nieuwe AVG de bevoegdheid krijgen om documentatie op te vragen over de werking van deze systemen en ter controle ook toegang kunnen eisen tot de servers en inzage in de gebruikte methode.

3. Doelbinding en het gerechtvaardigde belang van de data controller

Een belangrijke voorwaarde bij de verwerking van persoonsgegevens, ook wanneer het gaat om big data, is dat vooraf duidelijk is (1) voor welk specifiek doel de data worden verwerkt en (2) dat de verwerking zich ook daartoe beperkt. Dit is het finaliteitsbeginsel oftewel de eis van doelbinding. Gezien de complexiteit van de gegevensstromen wordt deze eis door veel verantwoordelijken als zeer problematisch ervaren. Bij big data is het adagium dikwijls: eerst zo veel mogelijk en overal gegevens verzamelen en achteraf kijken we wel wat we ermee kunnen doen. Zoals hiervoor beschreven leidt dit gemakkelijk tot het werken met *low hanging fruit*, wat de betrouwbaarheid van de uitkomsten niet ten goede komt. Doelbinding is dus niet alleen een eis voor een rechtmatige, maar ook voor een methodologisch betrouwbare verwerking. Denk bijvoorbeeld aan de datastromen die bij het gebruik van een smartphone op gang komen: naar de hardwareprovider, de aanbieder van het *operating system*, de *firmware*, de browser, allerhande applicaties. Het Internet of Things (IoT), de slimme energie-infrastructuur en de robotica voegen daar nog eens vele gegevensstromen aan toe, waaronder vooral ook machinaal leesbare gedragsgegevens. Het doel moet echter, ook en juist bij big data, steeds legitiem zijn en steeds voldoende specifiek en bovendien expliciet – dat wil zeggen kenbaar – zijn. Verwerkingen voor een ander doel mogen alleen plaatsvinden als het doel verenigbaar is met het oorspronkelijke, zodat het voor de betrokkenen redelijkerwijs voorzienbaar blijft waar hun data voor gebruikt kunnen worden.

Bij big data is vaak hergebruik aan de orde door andere partijen en/of voor andere doelen. Denk aan energiegebruiksgegevens die worden ingezet om sociale-zekerheidsfraude op te sporen, of aan locatiegegevens die nodig zijn om een dienst te verlenen en die worden hergebruikt om gezondheidsrisico's in te schatten. In dat geval is in beginsel een nieuwe juridische grond nodig. Dat hoeft zeker niet altijd toestemming te zijn, het kan ook een wettelijke verplichting betreffen (fraudedetectie) of het legitieme belang van de verantwoordelijke (zoals een economisch belang). Zo bieden veel bedrijven gratis diensten aan (zoekmachines, sociale netwerken), die ze kunnen onderhouden en verbeteren dankzij advertentie-inkomsten die de verwerking van gedragsgegevens vragen.
Bij big data lijkt het vragen van toestemming dan ook vaak ondoenbaar en oneerlijk. Voor individuele consumenten is het vrijwel ondoenbaar in te schatten welke patronen en verbanden uit hun data worden afgeleid en wat de gevolgen daarvan kunnen zijn (uitsluiting van krediet, premieverhoging, afwijzing voor een baan of opleiding). Om die reden zal toestemming meestal niet voldoen aan de eisen van ondubbelzinnige, geïnformeerde *consent*. In de rechtspraak van het Hof van Justitie van de EU, die hier doorslaggevend is, wordt daarom aangenomen dat, wanneer het verdienmodel van een bedrijf afhankelijk is van de verwerking van persoonsgegevens, de toepasselijke grondslag 'het gerechtvaardigde belang van de verantwoordelijke' zal zijn. Die grond vraagt altijd een afweging van

dit belang tegen de rechten en vrijheden van de betrokkenen, die uiteraard moeten worden gerespecteerd. Veel zal daarbij afhangen van de technische en organisatorische maatregelen die de verantwoordelijken nemen: om de data te beschermen tegen niet-toegelaten gebruik en tegen hacking, om ongewenste targeting te voorkomen en om de hierboven besproken profieltransparantie te realiseren. Denk aan pseudonimisering en eenvoudige manieren om gegevens in te zien of de verwerking stop te zetten (het intrekken van de toestemming moet net zo eenvoudig zijn als het verlenen ervan, aldus de AVG). Zoals hierboven al opgemerkt wordt dit soort maatregelen onder de AVG verplicht gesteld als gegevensbescherming 'per ontwerp' en 'per verstek'.

4. Onschuldpresumptie bij politie en justitie

Big data en machinaal leren worden niet alleen in de privésector ingezet. Politie en justitie gebruiken inmiddels technieken als *crime mapping* om zogenaamde *hot spots* te detecteren waar specifieke problemen verwacht kunnen worden (voor de openbare orde, strafbare feiten, rampen…). Ook wordt geïnvesteerd in softwaretoepassingen die scores geven over de kans dat veroordeelden opnieuw in de fout zullen gaan. Daarmee kan rekening worden gehouden bij de hoogte en modaliteit van de strafeis. Ook hier gaat het om de verwerking van persoonsgegevens, maar dan binnen een ander juridisch kader, waar de transparantie anders is georganiseerd omdat geheimhouding vaak noodzakelijk is om de taak te vervullen. Dit alles raakt aan de onschuldpresumptie, juist omdat ook hier de neiging bestaat om groepen personen op basis van de uitkomsten van big data-analyse en machinaal leren systematisch te gaan monitoren. Dit gebeurt dan bijvoorbeeld op grond van een mogelijke verdenking van mogelijk nog te plegen strafbare feiten of – nog breder – op grond van vermoedens dat er sprake zal zijn van ongewenst gedrag [Hildebrandt 2016].
Sinds de 'openbaringen' van Snowden is duidelijk dat ook de inlichtingen- en veiligheidsdiensten zich intensief bezighouden met het verzamelen en analyseren van allerhande communicatie- en gedragsgegevens, om tot accurate voorspellingen te komen van voorgenomen terroristische aanvallen. Tot nog toe is onduidelijk of de vele lijsten met mogelijk gevaarlijke personen bijdragen aan de daadwerkelijke preventie van aanslagen. Hierbij speelt onder meer mee dat deze fenomenen een onvoldoende regelmatig karakter hebben om een goed profiel te maken.

Dat leidt tot minstens drie heikele kwesties: (1) de publiek-private samenwerking die leidt tot het doorspelen van big data vanuit de privésector naar de strafrechtelijke autoriteiten en veiligheidsdiensten; (2) de hierboven al besproken crypto-oorlog rond de vraag of de nationale en internationale veiligheid nu al dan niet gediend is met 'achterdeuren' waarlangs overheden toegang kunnen krijgen tot de opslag en de communicatie van gegevens van eigen en andere burgers; (3) burgers kunnen steeds minder vooraf weten welk gedrag aanleiding is voor nader onderzoek. Het kan daarbij bovendien lastig zijn om verdenkingen, aanwijzingen en vermoedens aan de kaak te stellen, omdat de betrokken profielen geheim blijven.

Ten aanzien van de veiligheidsdiensten mag duidelijk zijn dat dit een nationale aangelegenheid is, die dus niet onder de werking van het gegevensbeschermingsrecht van de Europese Unie valt, wel onder het Europees Verdrag voor de Rechten van de Mens. Het Europees Hof voor de Rechten van de Mens spreekt zich dan ook regelmatig uit over de strenge voorwaarden waaronder profilering door veiligheidsdiensten is toegestaan.

5. Privaatrechtelijke aansprakelijkheid voor onrechtmatige verwerking

Naast de publiekrechtelijke aansprakelijkheid van de burger tegenover de overheid is er ook een privaatrechtelijke aansprakelijkheid van de burgers tegenover medeburgers. Indien de Autoriteit Persoonsgegevens in de diverse lidstaten (wellicht onder dwang van het Hof van Justitie van de EU) voldoende budget, technische expertise en staf kan inzetten, mogen we verwachten dat de hoge boetes en de Unie-brede toepassing van de Verordening een redelijk effectieve bescherming zullen bieden tegen uitwassen. De Verordening eist echter ook effectieve privaatrechtelijke aansprakelijkheid voor de schending van het juridische kader. Daarbij kan gedacht worden aan de schending van de verplichting om een datalek te melden, maar ook aan de schending van de veiligheidseisen of simpelweg aan onrechtmatige verwerking (zonder geldige grond of die voorbij het aangegeven doel gaat). Deze privaatrechtelijke aansprakelijkheid eist dat er sprake is van aantoonbare materiële of immateriële schade die – ook weer aantoonbaar – veroorzaakt moet zijn door de betrokken schending van rechtsplichten of rechten. De Verordening spreekt echter van een 'effectief rechtsmiddel', dat daadwerkelijk bescherming moet bieden wanneer rechten van betrokkenen zijn geschonden. Wereldwijd lijken rechters vaker bereid om schadevergoeding toe te kennen wanneer er sprake is van immateriële schade, bijvoorbeeld onzekerheid over mogelijke identiteitsfraude of reputatieschade. Zoals hierboven vermeld, eist de Verordening bovendien dat de lidstaten de mogelijkheid scheppen voor betrokken individuen om hun aanspraken te mandateren aan ngo's, waardoor aanspraken kunnen worden gebundeld.

We kunnen verwachten dat op enig moment zal worden overwogen of een risicoaansprakelijkheid hier uitkomst kan bieden, zodat de getroffene niet opgezadeld wordt met de onmogelijke bewijslast dat de onrechtmatige verwerking de schade daadwerkelijk heeft veroorzaakt. De bewijslast kan dan bijvoorbeeld worden omgedraaid. Het is niet waarschijnlijk dat een enkele schending, op zichzelf genomen, gelijk kan worden gesteld met schade. Zo blijft de bescherming van dit type aansprakelijkheid beperkt tot zaken waarbij aannemelijk kan worden gemaakt dat er sprake is van materiële of immateriële schade. Daar komt bij dat het aansprakelijkheidsrecht nationaal recht is, en dus door de nationale wetgever en rechter wordt bepaald. Het 'Europese privaatrecht' is vooralsnog een mengeling van rechtsvergelijking en relevante Richtlijnen, zoals die van de productaansprakelijkheid. We mogen echter verwachten dat het zich verder

zal ontwikkelen, nu het van cruciaal belang is voor een goede werking van de interne markt en de effectieve bescherming van consumenten die diensten van transnationale bedrijven gebruiken.

Anonimiteit en anonimiseren

Anonimiteit speelt een centrale rol in vraagstukken over privacy en big data. Vooral juridisch is het een boeiend begrip. Persoonsgegevens worden beschermd omwille van de privacy, maar als je ze anonimiseert, dan zijn het niet langer persoonsgegevens en worden ze ook niet langer juridisch beschermd.
Hoe komt dat? De verklaring is eenvoudig: geanonimiseerde gegevens verwijzen niet langer naar individuen. Er is voor niemand een privacyprobleem bij de uitspraak: 'Er is iemand met een baard'. Niemand weet wie die persoon is en de uitspraak wordt daarom ook niet als relevant voor de privacy aangemerkt. Dat gaat niet langer op als er maar weinig mensen met een baard zijn, tegenover veel mensen zonder baard, want dan is het al snel mogelijk de baarddrager te identificeren. Het anonimiseren van persoonsgegevens lijkt dus erg goed voor de privacy, voor zover geanonimiseerde gegevens niet gebruikt kunnen worden met miskenning van iemands privacy. In onze kennismaatschappij wordt er geanonimiseerd met het oog op het verdere gebruik en delen van beschikbare gegevens. Vaak bereiken bedrijven en instellingen met anonieme gegevens al hun doelen en blijken persoonsgegevens niet echt nodig te zijn.
Big data-activiteiten zijn dan ook vaak haalbaar op basis van geanonimiseerde gegevens. Een beheerder van deelfietssystemen wil weten waar in zijn stad, bij welke fietsstations, er deelfietsen te veel of te weinig zijn. Weten wie de fiets gebruikt of waar een specifieke fiets zich bevindt, is daarvoor niet nodig. Natuurlijk blijft het interessant, en soms nodig, om te werken met persoonsgegevens. De beheerder van fietsen die ook weet welke gebruiker waar fietst en met welke fiets, weet ontegensprekelijk meer. De vraag is of hij dat moet weten. Juridisch kan er gewerkt worden met persoonsgegevens, maar alleen als het echt moet en niet anders kan, zegt de wet. Het spreekt vanzelf dat er hierover veel discussies zijn.

Een andere, meer fundamentele discussie, gaat over de anonimisering zelf. In steeds meer gevallen maakt de technologie het mogelijk om de anonimiteit van gegevens te doorbreken en te weten wie de persoon is achter een uitspraak als 'er is iemand met een baard'. In de gegevensbeschermingswetgeving is het bovendien zo dat gegevens waar de kans op (her)identificatie redelijk is, per definitie persoonsgegevens zijn. Zodra het mogelijk is, bijvoorbeeld door het samenvoegen van data om een datapunt met een identificeerbare persoon te verbinden, is er sprake van een persoonsgegeven. Zo worden zowel dynamische als statische IP-adressen beschouwd als persoonsgegevens voor zover ze redelijkerwijs gekoppeld kunnen worden aan de persoon. Juridisch gezien leidt de inzet van de meeste anonimiseringstechnieken dan ook niet tot volledige anonimisering, maar tot pseudonimisering, juist omdat er bijvoorbeeld een decryptiesleutel beschikbaar

is. (Zelfs als die door anderen wordt beheerd, zal er vaak geen sprake zijn van anonieme data, zolang de verantwoordelijke redelijkerwijs toegang kan krijgen tot die sleutel, bijvoorbeeld door een rechterlijke tussenkomst.). Ook pseudonieme data zijn juridisch dus per definitie persoonsgegevens, maar effectieve pseudonimisering kan wel een goede manier zijn om tegemoet te komen aan de eisen van de Algemene Verordening Gegevensbescherming AVG [AVG 2016]. Het is dan een vorm van gegevensbescherming bij ontwerp. In dat geval moeten de aanvullende data waarmee de pseudoniemen geïdentificeerd kunnen worden, zowel technisch als organisatorisch apart worden gehouden.

Intussen is het in beginsel ook mogelijk om iedere gebruiker van de data een eigen sleutel te geven, zodat de gegevens die verschillende partijen verkrijgen niet bij elkaar kunnen worden gelegd. Dat is met name van groot belang bij medische en studie-gerelateerde gegevens, waar professionals een eigen verantwoordelijkheid hebben voor de vertrouwelijkheid van de data terwijl er een grote behoefte is aan toegang tot big data voor medisch onderzoek of de ontwikkeling van datagestuurde leeromgevingen [Verheul e.a. 2016]. De privacybescherming van pseudonimsering is echter beperkt [deMontjoye2013].

Onlineplatformen en gedeelde verantwoordelijkheid

Internet, en bij uitbreiding onlineplatformen, dringen in toenemende mate door tot in de vezels van de samenleving en economie. ‘Onlineplatformen’ zijn de ‘technologische, economische en sociaal-culturele infrastructuur voor het faciliteren en organiseren van online sociaal en economisch verkeer tussen gebruikers en aanbieders, met (gebruikers)data als brandstof’ [van Dijck et al, 2016]. Denk aan informatie en communicatie via sociale media (bv. Facebook), maar ook aan allerlei soorten dienstverlening via de deeleconomie, zoals transport (bv. Uber), de hotelmarkt (bv. Airbnb), onderwijs (bv. Coursera), gezondheidszorg (bv. Patientslikeme) en vele andere sectoren. De (meta)data van gebruikers zijn een cruciaal onderdeel van het verdienmodel van deze platformen, waardoor er steeds grotere belangen aan de orde zijn. Het komt er daarbij op aan een evenwicht te vinden tussen de opportuniteiten van data voor de samenleving en de bedrijfswereld enerzijds en de risico’s die daaraan verbonden zijn op het vlak van de fundamentele rechten van burgers (privacy, databescherming enz.) en (Europese) publieke waarden (gelijke behandeling, inclusie, diversiteit enz.).

Om dit evenwicht te realiseren is het noodzakelijk alle relevante publieke en privépartijen samen te brengen en in dialoog te laten treden. Dit betekent dat je een coöperatieve of samenwerkende verantwoordelijkheid organiseert tussen beleidsmakers, industrie, gebruikers en middenveldorganisaties. In de bestaande wetgeving bestaat de neiging om verantwoordelijkheid toe te wijzen aan één centrale speler (bv. *data controller*, redacteur, dienstverlener), om het toezicht en de aansprakelijkheid efficiënt te kunnen organiseren. Daarmee zijn alle problemen

echter niet opgelost. We zien dan ook dat de AVG spreekt van '*joint controllers*' en dat daarin ook de bewerkers van persoonsgegevens die in opdracht werken aansprakelijk worden gesteld voor onrechtmatige verwerking. Vergelijkbare kwesties van gedeelde verantwoordelijkheid spelen echter ook op het vlak van de vrije meningsvorming, en de problemen die zich daar voordoen zijn niet noodzakelijk oplosbaar via de weg van gegevensbescherming. In de context van internet en onlineplatformen zien we dat verschillende partijen verantwoordelijkheid kunnen en moeten opnemen, niet alleen inzake privacy en databescherming, maar ook voor onderwerpen zoals omstreden inhoud, haatdragende boodschappen, diversiteit en transparantie. De platformeigenaars zorgen voor de infrastructuur waardoor gebruikers met elkaar in contact treden en informatie delen, de gebruikers kiezen ervoor om – soms in groten getale – welbepaalde inhoud en data te delen en de overheid voorziet in een beleidsmatig en legaal kader waarbinnen de data verzameld, opgeslagen, verwerkt en beschermd worden.

Om zo'n samenwerkende verantwoordelijkheid te kunnen organiseren is er nood aan een vorm van multistakeholdersoverleg, waarbij rekening gehouden wordt met ethische, legale en sociale aspecten (ELSA). Dit vereist vooreerst de aanvaarding door alle stakeholders (inclusief sociale media en onlineplatformen) dat zij een verantwoordelijkheid dragen, die verschillend is naargelang van de context. Vervolgens dienen de betrokken spelers tot een gedeelde visie te komen over de invulling van en de omgang met privacy. Hiervoor kunnen waardevolle lessen getrokken worden uit de lange traditie van *(constructive) technology assessment* (technologisch aspectenonderzoek). Ten slotte dient iedere partij een voornemen om te zetten in concrete praktijken. Voor sociale media en onlineplatformen betekent dit bijvoorbeeld het inbouwen van mogelijkheden in het systeem waardoor er voor gebruikers voldoende transparantie is en zij controle hebben over hun data. Enkel zo kunnen breed maatschappelijk gedragen oplossingen ontstaan rond toekomstige digitale media en datatechnologieën. Het mag daarbij duidelijk zijn dat er een '*incentive*'-structuur moet bestaan waarbinnen de stakeholders gedwongen zijn hun verantwoordelijkheid te nemen, vooral ook vanwege het transnationale en globale karakter van de grote spelers. Hoge boetes en privaatrechtelijke aansprakelijkheid, zoals neergelegd in de AVG, vormen een eerste aanzet om hier een '*level playing field*' te creëren dat bedrijven en overheden toestaat om hun verantwoordelijkheid te nemen zonder uit de markt te worden geduwd.

Datageletterdheid van gebruikers

Vaak wordt in de juridische en technische studies over privacy, data en internet over de gebruikers gesproken vanuit een expertperspectief en te weinig met (of door) de gebruikers, vanuit het standpunt en de dagelijkse leefwereld van de gebruikers zelf. Dit leidt soms tot een eenzijdige en 'arme' invulling van 'de' gebruiker, terwijl de werkelijke omgang met de digitale omgeving veel rijker en complexer is wat

bewustwording, attitudes, vaardigheden en gedrag betreft. Specifiek voor jongeren moet men rekening houden met de verschillende levenssferen waarin zij zich begeven: thuis, school, de vriendenkring (bv. jeugdbewegingen). Enkel door op een dergelijke genuanceerde wijze de gebruiker empirisch te onderzoeken is het mogelijk om 'empowerment' van gebruikers te realiseren en 'disempowerment' te vermijden. 'Empowerment' is het proces van versterking van gebruikers/burgers/consumenten, waardoor ze greep krijgen op de eigen situatie en hun omgeving, door het verwerven van controle, het aanscherpen van kritisch bewustzijn en het stimuleren van participatie.

Gezien het belang van 'empowerment' in socio-technologische veranderingen rond internet en big data moet de gekende notie van mediawijsheid worden aangevuld met 'datageleterdheid'. Dat begrip verwijst naar het inzicht, de controle en het vertrouwen dat men heeft in de wijze waarop (persoonlijke) data worden verzameld, opgeslagen, verwerkt en gebruikt (of hergebruikt). Het gaat om data die men bewust (foto's, interesses, adresgegevens enz.) dan wel onbewust (surfgedrag, locatie, cookies enz.) vrijwillig vrijgeeft, maar zeker ook om afgeleide data (kredietscore, *profiling*, emoties enz). Hiertoe dient men niet enkel te kijken naar digitale media (zoals internet en sociale netwerken), maar ook naar een toenemend aantal andere technologieën die data genereren en verwerken (*wearables*, drones, Internet of Things, *smart grids* enz.). Een van de grootste uitdagingen in dit verband is het verstaanbaar en betekenisvol maken van de complexiteit van technologische aspecten die impact hebben op (meta) data en privacy, zoals algoritmes, APIs, *machine learning*, privacysettings bij verstek (default) en AB testing.

3. Analyse van de privacy aan de hand van relevante casussen

De casussen hieronder worden behandeld met een blik op de toekomst en op het hele systeem zoals de modale gebruiker het ervaart. Hierbij wordt telkens ingegaan op de facetten van big data en lerende systemen die de situatie voor de privacy grondig veranderen. We wijzen ook elke keer op de afwegingen die worden gemaakt tussen gebruiksvriendelijkheid, functionaliteit, privacybescherming, kosten, kwaliteit en volledigheid.

Casus 1: het digitale leven van een gezin

De penetratie van het internet en daarmee verbonden apparaten ligt in Vlaanderen zeer hoog [Digimeter 2016]. Hieronder schetsen we een fictieve, maar realistische situatie uit het dagelijkse leven van een gezin met moeder, vader en twee dochters van 9 en 13 jaar. Allerhande aspecten uit de begrips- en probleemverkenning die deze tekst uitvoert komen erin tot uiting.

Het is weekend. Iedereen is wakker en het ontbijt staat op tafel. De regel is dat er tijdens het eten geen tablets of smartphones worden gebruikt. Na het afruimen

gaat elk gezinslid aan de slag met de smartphone of de tablet. Moeder plant de zomervakantie naar Frankrijk. Ze wil nog even wat campingwebsites bekijken. Vader bereidt een fietstochtje met wielervrienden voor. Hij is alvast benieuwd naar de route die ze gepland hebben en hij wil stiekem ook kijken waar zijn vrienden de voorbije week allemaal geoefend hebben. De jongste dochter houdt van Youtube. Ze is moe van de voorbije schoolweek en heeft gewoon zin om wat filmpjes te bekijken op haar tablet om zich te ontspannen. De oudste dochter heeft een eigen smartphone en trekt zich terug in haar kamer. Haar vrienden zijn immers ook wakker en zijn al volop berichten aan het sturen via Instagram.

Moeder begint te surfen. Ze gebruikt Google om campings te vinden. Ze tikt 'kleinschalige camping Zuid-Frankrijk' in. Als eerste verschijnt er een 'privacy-herinneringsbericht' van Google. Ze bekijkt het snel. Het is een lange tekst.

Ze leest iets van 'gegevens bijhouden, Youtube, andere services, ervaringen verbeteren, gepersonaliseerde zoekresultaten…'. Ze leest ook dat je alles kan aanpassen, maar besluit dat ze te weinig tijd heeft om zich hiermee bezig te houden. Ze klikt 'ik ga akkoord' en gaat verder. De *privacy paradox* in werking. Moeder vindt de eerste campings uit de zoeklijst al direct interessant. Ze vindt het ook handig dat er meteen een kaart verschijnt met de ligging van de campings en beoordelingen van andere toeristen. Ze klikt de eerste campings van de lijst aan. *Advertenties* (*search engine advertising*) dus. Het valt haar op dat ze op de website van een van de campings reclame zag voor de schoenen die ze deze week online had bekeken. Ze twijfelt nog of ze die gaat bestellen. Dat ze de schoenen overal ziet verschijnen, geeft haar toch geen behaaglijk gevoel. Ze besluit dat dit onvermijdelijk is en dat ze dan maar beter zaken ziet die ze ook effectief interessant vindt. De werking van *retargeting*. Helaas betekent het vaak ook dat ze dit soort advertenties blijft ontvangen nadat ze de bestelling gedaan heeft, wat het wel enigszins storend maakt.

Vader kijkt uit naar de wekelijkse fietstocht. Hij heeft een nieuwe hartslagmeter gekocht en wil die nog snel installeren voor hij vertrekt. Hij downloadt de app, die vraagt om de hartslaggegevens te koppelen aan Strava, de app die hij gebruikt om zijn fietsactiviteiten en –prestaties te meten. 'Dat is leuk!', denkt hij. 'Zo kan ik mijn hartslag vergelijken met mijn vrienden. Wie van ons is in de beste conditie?' Of hij deze gegevens ook openbaar wil maken? 'Hmm, misschien nu nog niet. Maar als blijkt dat mijn hart in topvorm is dan kan ik daar echt wel mee uitpakken.' Hij kijkt even naar de profielen van zijn vrienden. Ze hebben duidelijk goed getraind deze week. Een is zelfs in de Vlaamse Ardennen gaan fietsen. Hij surft nog even naar de website waar hij zijn fietstoebehoren koopt. Ze bevelen hem een nieuw soort banden aan. Die blijken tijdelijk in promotie. Hij kijkt ook naar de fietshandschoenen die hij allang wil kopen. 'Nog maar enkele in voorraad.' Hij beslist zowel de banden als de handschoenen meteen te bestellen. *Nudging* in de praktijk.

De jongste dochter installeert zich in het salon met haar tablet. Ze gaat meteen naar Youtube en kijkt of er nog nieuwe filmpjes zijn van de mensen die ze volgt. Eén ervan is een zogenaamde vlogger. Een grappige jongen die enorm veel volgers heeft en twee keer per week een filmpje maakt over zijn leven. Vandaag heeft hij het over een tripje naar een nieuw pretpark. 'Dat ziet er supercool uit!' De dochter beslist om straks aan haar ouders te vragen of ze daar deze zomer samen naartoe gaan. Ze werd dus beïnvloed door een *social influencer*, die mogelijk een gratis trip naar het betreffende pretpark kreeg aangeboden. Nadien speelt ze nog een gratis spelletje (mama en papa betalen niet graag voor spelletjes op de tablet). Het is een racespel waarbij je zo veel mogelijk flesjes van een frisdrank moet omverrijden. 'Heel grappig!' Ze deelt haar score met vrienden en nodigt nog enkele andere vrienden uit om het spel ook te spelen. 'Waar zou je die frisdrank kunnen kopen?' Ze speelde een *advergame* gemaakt en betaald door een commercieel bedrijf. Voor jonge kinderen is het niet duidelijk dat dit reclame is.

De oudste dochter trekt zich terug in haar kamer. Ze opent Instagram. Haar vrienden posten volop grappige foto's en filmpjes, en becommentariëren wat ze hebben gezien. Als ze door de nieuwsberichten scrollt, ziet ze een bericht van een sneakerwinkel waar ze gisteren nog binnenwandelde met haar vriendinnen. Ze hebben in de winkel staan aanduiden welke sneakers ze graag zouden kopen. Via het bericht kan je meteen naar de onlinewinkel surfen. Ze gaat dat straks tonen aan haar moeder. Die had immers beloofd dat ze nieuwe sneakers zou krijgen. Het bericht van de sneakerwinkel is *reclame gebaseerd op geolocatie*. Doordat het social medium Instagram via smartphones weet wie waar is geweest, kunnen ze aan adverteerders aanbieden om zo gericht mensen te bereiken. Plots ziet de oudste dochter een foto van haarzelf opduiken van een aantal jaren geleden. Een foto waar ze zich erg over schaamt. Blijkbaar heeft iemand de foto ooit op een website geplaatst en kan je hem dus via Google vinden als je haar naam intikt. Ze weet niet hoe de foto verwijderd kan worden, maar hoopt dat dit beeld haar ooit niet meer zal achtervolgen. 'Iedereen doet toch wel eens domme dingen?' Het recht om vergeten te worden is met andere woorden essentieel.
Ondertussen is het middag. Het gezin gaat weer aan tafel. Er wordt gepraat over de zomervakantie, het geplande fietstochtje, een nieuw pretpark en coole sneakers. Over de bewuste foto wordt niet gerept.

Casus 2: Big data bij het profileren van passagiers

Na de aanslagen in New York op 9/11 was het voor de autoriteiten niet meer dan logisch om te kijken naar de wijze waarop de luchtvaart georganiseerd was. De Verenigde Staten, meer bepaald *Homeland Security*, ontwikkelde de idee om een zwarte lijst aan te leggen van passagiers die het best niet meer op vliegtuigen opstapten. Alle luchtvaartmaatschappijen gebruiken bij het boeken en organiseren van vluchten hetzelfde *Passenger Name Records*-systeem (PNR). Maatschappijen die op de V.S. vliegen werden verplicht de gegevens voor het vertrek te mailen: geen gegevens, geen vluchtlicentie. In deze berg gegevens zoekt *Homeland Security* naar verdachte passagiers. Om zwarte lijsten aan te leggen gebruikten ze niet alleen de namen van gekende verdachten maar ook profilering. Op basis van een verdachte aankoop van de vlucht (bv. met cash betalen, zeer laat een vliegticket boeken enz.) wordt de lijst met echte verdachten aangevuld met een lijst van mogelijke verdachten. Typisch big data.

Het systeem wordt vandaag de dag overal ter wereld gebruikt, ook in Europa. Echte overzichtscijfers erover bestaan jammer genoeg niet. Het is dus niet geweten of het werkt. We weten wel dat er op die zwarte lijst mensen staan die helemaal niets met terrorisme te maken hebben, maar toevallig geprofileerd zijn. Dat overkwam onder meer de zanger Cat Stevens. De bevoegde overheidsdiensten noemen dat pech voor de betrokken persoon. Een echte regeling om uit de zwarte lijsten te geraken is er niet. Het systeem is bovendien niet echt vriendelijk te noemen. Je krijgt pas te horen dat je op de zwarte lijst staat bij de aanmelding met de koffers op de luchthaven.

Wat te denken van deze big data-toepassing? Zonder effectrapportage is er geen discussie mogelijk, maar los daarvan valt op dat er weinig inspanningen gedaan worden om dit big data-project ook voor burgers aantrekkelijk te maken. Een eerste stap is eerlijk zijn en aangeven dat het systeem ten dele op gokken berust. Fouten zijn bijgevolg mogelijk. Een tweede stap is: goed en concreet nadenken over passagiers die ten onrechte worden tegengehouden en te voorzien in een nagenoeg automatische compensatie bij fouten. Dit systeem bestaat nu reeds wanneer vluchten geannuleerd worden of bij grote vertragingen. Op die manier kan al wat ergernis worden weggenomen. Derde stap: het systeem echt gebruiksvriendelijk maken. Breng mensen die op een zwarte lijst staan en een ticket kochten, per mail op de hoogte van het feit dat ze niet kunnen vliegen. Waarom niet deze transparantie bieden? Overheidsgeheimen? De opgelijste burgers komen het bestaan van hun vermelding op de lijst toch te weten, maar dan in moeilijke omstandigheden (op de luchthaven). Het zou een ander verhaal zijn wanneer ze reeds bij aankoop van het ticket vernemen dat de zwarte lijst bestaat. Dan is bevraging mogelijk en kan men een advocaat inschakelen. Minstens kunnen dan afspraken in het buitenland worden geannuleerd. Door op een juiste manier met big data om te gaan kunnen de rechten van passagiers worden gerespecteerd, zelfs in de sfeer van terrorismebestrijding. De onvermijdelijke fouten van de algoritmes worden dan meer acceptabel [DH 2011].

Casus 3: Internet der Dingen in de context van slimme steden

Twee tendensen maakten de doorbraak van het Internet der Dingen (Internet of Things, IoT) mogelijk: de hedendaagse elektronica reikt goedkope, minuscule sensoren en microprocessen aan, en de alomtegenwoordige netwerking zorgt ervoor dat alles met alles geconnecteerd kan worden. Objecten kunnen via hun sensoren data verzamelen en met elkaar en de buitenwereld interageren. Dat biedt veelvuldige nieuwe toepassingen, gaande van gezondheid (waar persoonlijke parameters zoals bloeddruk, temperatuur en andere klinische parameters van individuen opgevolgd kunnen worden), over intelligente huizen en gebouwen (waar alles, van lichten tot verwarming en deuren, gemonitord en gecontroleerd wordt), tot intelligente steden (waar mobiliteit en transport, maar ook watergebruik, afvalproductie en energiegebruik gemonitord en gecontroleerd kunnen worden).

De typische sensoren in IoT leveren een continue gegevensstroom op (een groot volume en een grote variëteit van gegevens in een hoge frequentie), die geanalyseerd kan worden. De gegevens kunnen ook gecombineerd worden met andere data, bijvoorbeeld uit sociale media (Facebook, Twitter…). Zo vormt het Internet der Dingen een fantastische bron voor big data-toepassingen. Aan de hand van allerhande algoritmes zal men in de veelheid aan data patronen en correlaties proberen te herkennen, om zo contextueel verrijkte informatie te produceren,

waaruit men conclusies kan trekken. Zelfs uit zogenaamde geanonimiseerde data kan men uit de grote hoeveelheid informatie een uniek profiel herkennen en zo te weten komen op welk individu de gegevens betrekking hebben.
Deze casus focust op IoT in slimme steden, *smart cities*. Die term wordt vaak gebruikt voor steden waar ingezet wordt op het stimuleren van innovatie en creativiteit, al dan niet met ICT, en van duurzaamheid, ondernemerschap en ICT-onderwijs. Verder wordt hij ook toegepast op steden die een burger-gecentreerd model hanteren van stedelijke ontwikkeling, met aandacht voor sociale innovatie en cohesie, gelijkekansenbeleid en participatief stadsbeheer.
Waarom die interesse voor *smart cities*? Wereldwijd zien we dat een groeiend aantal mensen in steden leeft: in 2014 al meer dan 50% van de wereldbevolking. De Verenigde Naties verwacht dat in 2050 dit percentage zal gestegen zijn tot 66% [VN2014]. Om de steden leefbaar te houden zet men in op betere diensten en omstandigheden. *Smart cities* moeten daaraan bijdragen, onder meer door te voorzien in betere, goedkopere, efficiëntere en duurzamere diensten [Ballon 2016]. Enkele voorbeelden van het gebruik van IoT maken dit duidelijk.

- Slimme mobiliteit: sensoren meten de mobiliteitsstromen in de stad, zowel van voetgangers, fietsers en automobilisten als van het openbaar vervoer. Zo kom je bijvoorbeeld te weten waar er files staan of op welke buslijnen versterking nodig is. En ook de automobilist op zoek naar een parkeerplaats wordt geholpen. Sommige steden in Nederland denken eraan om bij regenweer de stoplichten aan te passen en zo fietsers extra voorrang te geven. Autodelen en fietsdelen wordt gemakkelijker gemaakt door slimme apps, die verbonden zijn met het netwerk van de te delen vervoermiddelen. Heel het mobiliteitsgebeuren wordt globaal gemonitord, gecontroleerd en waar mogelijk gestuurd.
- Hogere beveiliging met snellere en beter gerichte spoedinterventies: een netwerk van slimme camera's en andere sensoren, gecombineerd met gecoördineerde hulpdiensten en controlekamers, kan leiden tot een snellere detectie en efficiëntere interventies bij problemen, natuurrampen of terreurdaden.
- Intelligent afvalbeheer: door sensoren te plaatsen in vuilniscontainers en –bakken kunnen de ophaaldiensten hun routes aanpassen aan de reële noden.
- Monitoring van het milieu: een netwerk van sensoren kan luchtvervuiling, geluidsoverlast, waterzuiverheid en het waterniveau in beken en rivieren, monitoren, om sneller in te grijpen of preventieve acties te ondernemen.
- Slimme verlichting: in plaats van straatverlichting altijd te laten werken kan men met sensoren meten wanneer er beweging is, en alleen dan de straatverlichting (op volle kracht) te laten schijnen.
- Elektriciteit wordt overal in de stad gebruikt en mogelijk ook opgewekt door bijvoorbeeld zonnepanelen of andere kleinschalige groene energieprojecten. Alle gebouwen voorzien van slimme meters is één facet van de zogenaamde *smart grids* of intelligente energienetwerken. Deze slimme meters bevatten intelligentie en kunnen dus ook beslissingen nemen [Belmans e.a. 2016].

Bron: www.energiesparen.be

° Met een slimme meter kan elk gebouw niet alleen elektriciteit van het net verbruiken, maar de meter kan ook de opgewekte elektriciteit in het gebouw op het net brengen. Het hele energienetwerk wordt hierdoor dynamisch, in tegenstelling tot vandaag meestal.
° Met een intelligente meter kan een bewoner makkelijker zijn gebruik optimaliseren en bijvoorbeeld alleen in de daluren zware apparaten zoals wasmachines laten draaien of zijn elektrische auto laden.
° Elektriciteit wordt vaak gebruikt in pieken (bijvoorbeeld het begin van de avond, wanneer veel mensen thuiskomen, koken en televisie kijken). Op die piekmomenten de airco of de wasmachine even laten pauzeren kan een slimme beslissing zijn die genomen kan worden door een slimme meter.
° Globaal gezien verwacht men van *smart grids* dat ze zullen zorgen voor een meer betrouwbaar systeem (want gedecentraliseerd, waar men niet afhankelijk is van één bron, omdat meerdere gebouwen in de stad ook elektriciteit kunnen leveren) en voor een efficiënter systeem. Daarenboven

bevorderen *smart grids* hernieuwbare energie – en dus duurzaamheid – omdat die gemakkelijker opgenomen kan worden in het elektriciteitsnetwerk.

- In slimme steden staan slimme huizen, waar de verwarming per kamer ingesteld kan worden, afhankelijk van de actie die er al dan niet plaatsvindt; ook kan de verwarming per kamer op afstand geregeld worden. Wat geldt voor verwarming, geldt ook voor licht of airco. Allerhande apparaten die 'op het internet hangen' kunnen op afstand bediend worden.

De sensoren, zoals ze worden gebruikt in de vermelde voorbeelden, leveren grote hoeveelheden informatie aan, die soms erg gevoelig is. Het doen en laten van de bewoners van een slimme stad wordt namelijk op verschillende manieren gecapteerd, met een ongeziene intensiteit. De data worden gebruikt om 'slimme' diensten aan te bieden en veel gegevens kunnen worden gekoppeld aan individuen, zelfs al zijn ze vooraf geanonimiseerd. Door het combineren van data vanuit verschillende invalshoeken kan het leven van individuen redelijk nauwgezet in kaart gebracht worden! Om deze reden werd in Nederland de regelgeving rond slimme elektriciteitsmeters aangepast. Eerst was gepland dat het invoeren ervan verplicht zou worden bij nieuwbouw en renovatie, en bij elke vervanging van een oude meter. Maar op grond van privacyaspecten werd die verplichting afgevoerd: een gebruiker mag nog altijd kiezen voor een traditionele 'domme' meter. In het Verenigd Koninkrijk maakte men zich meer zorgen over de veiligheid. Men was bang dat slimme meters, die meestal draadloos communiceren, tot een ongewenste blootstelling aan gepulste, radiofrequente straling in de woning zouden leiden. Ook in het Verenigd Koninkrijk is de verplichting afgevoerd. Maar de EU [Smart metering in EU 2014] en de Vlaamse regering [Digitale meters 2017] blijven de uitrol van slimme meters aanmoedigen.

De uitdagingen voor de privacy die zijn gekoppeld aan de toepassingen van *smart cities* zijn groot.

- Een groot deel van de uitdagingen heeft te maken met transparantie en openheid.
    - Elke individu moet de toestemming kunnen geven voor het gebruik van zijn data, op een efficiënte en effectieve manier. Het gaat daarbij niet alleen over het directe gebruik. Data die bedoeld zijn voor één toepassing worden vaak later nog eens gebruikt voor andere toepassingen. Hetzelfde geldt voor het combineren van gegevens: gegevens die voortvloeien uit verschillende *smart city*-toepassingen kunnen gecombineerd worden om nieuwe informatie aan te leveren.
    - Elk individu heeft het recht te weten wat er gebeurt met de data die over hem of haar zijn verzameld. Dat is niet evident. De gebruikte algoritmen zijn vaak niet transparant. Het lijkt er meer op alsof alle data in een toverhoed gaan en er dan op een duistere manier een beslissing uitkomt. Wanneer we spreken over het gebruik van data, denken we ook aan hergebruik en aan het combineren van data uit verschillende bronnen.

- De gebruiker zou vrij moeten zijn in zijn keuze van apparatuur en toepassingen, en die zelf moeten kunnen combineren. Momenteel zijn de leveranciers van apparatuur, de serviceproviders en de schrijvers van toepassingen weinig transparant en proberen ze een monopolie te creëren om de gebruiker aan zich te binden.

- Beveiliging. Zelfs als men akkoord gaat dat bijvoorbeeld de gegevens uit een intelligent huis gebruikt worden door toepassingen die bijvoorbeeld een efficiënt energieverbruik bewerkstelligen, wenst men niet dat deze data in handen komen van dieven, die uit deze gegevens perfect de gewoonten van de bewoners kunnen afleiden en zo het beste moment om in te breken kunnen bepalen.
- Wettelijk kader: bovenstaande normen zijn steviger dan voorheen verankerd in de AVG, die over de hele Europese Unie een direct effect zal hebben en effectieve handhaving mogelijk zal maken. Dit is van cruciaal belang, nu vele ICT-toepassingen, ook allerhande apps die in het kader van *smart cities* aangeboden worden, gedownload kunnen worden van het internet, dat geen nationale grenzen kent.

Er is veel werk aan de winkel:

- de *bouwers van IoT-apparaten* moeten technologieën ontwikkelen die de privacy respecteren en transparantie toelaten voor de eindgebruiker. Ze moeten werk maken van privacy *bij ontwerp*, waarbij privacy vanaf het begin van het ontwerp als een belangrijke vereiste meegenomen wordt en er niet achteraf aan 'geplakt' wordt;
- de *service providers* moeten toelaten dat gebruikers diensten samenstellen van verschillende oorsprong;
- de *ontwerpers van algoritmes* moeten hun algoritmes zo schrijven dat ze de privacy van de gebruikers garanderen;
- de ontwerpers van *toepassingen* moeten transparantie toelaten en werk maken van efficiënte en effectieve technologieën, waarbij gebruikers de toestemming kunnen geven voor het gebruik van hun data. Verder moet men werk maken van het certificeren van toepassingen, zodat gebruikers zeker zijn dat de toepassingen veilig zijn. Voor zowel de apparatuur als de toepassingen is er een grote nood aan betere en fijnmazige methodes om instellingen en voorkeuren op te geven. Hier moet privacy de verstek-keuze zijn (privacy 'bij verstek');
- er is tot slot ook veel werk voor de *regelgeving*. Die moet zorgen voor standaardisatie (zonder een rem te zijn voor innovatie) en een basis aanreiken voor certificatie [Perera 2015] [DG Int.Pol. 2015] [VN2014].

Casus 4: Gedistribueerde informatie versus centrale collectie

Er zijn verschillende redenen die bedrijven en overheden aanzetten om steeds meer informatie te centraliseren in een *data cloud*; dat kan een publieke cloud zijn,

bij een grote speler als Amazon, Google of Microsoft, of de eigen infrastructuur. In eerste instantie brengt dit een gevoelige kostenreductie met zich mee, voornamelijk door schaalvoordelen. Je kunt hierdoor ook op een zeer flexibele manier inspelen op veranderingen in datavolumes, zonder grote investeringen. Een tweede argument is dat op deze manier de gegevens altijd en overal beschikbaar zijn voor wie ze nodig heeft. Men kan hierbij denken aan foto's die men wil bekijken op smartphones en tablets, maar die ook gemakkelijk gedeeld kunnen worden. Een derde reden is de mogelijkheid tot data-analyse, het zogenaamde *data mining*: dit maakt het mogelijk om nieuwe informatie en behoeften te ontdekken, maar ook om gericht te adverteren – de inkomsten hiervan voor Google en Facebook zijn gigantisch.

Deze aanpak is niet zonder risico's: elke maand worden persoonsgegevens van miljoenen gebruikers gelekt door aanvallen met hacks of nalatigheid [website-breaches]. Blijkbaar is het niet altijd mogelijk om de grote hoeveelheden gevoelige data in de (publieke) cloud adequaat te beschermen. Daarnaast eindigt deze informatie ook in de handen van overheden: in juni 2013 heeft Edward Snowden onthuld dat het PRISM-programma de NSA toegang geeft tot de data die zijn opgeslagen in de cloud bij spelers als Microsoft, Facebook, Google en Apple [PRISM]. Ten slotte brengt de centralisatie van informatie een groeiend risico voor secundair gebruik met zich mee; zelfs als de gebruiker dit had toegezegd, kan hij de implicaties zeer moeilijk overzien.

Het is in principe ook mogelijk om heel wat diensten op basis van lokale informatie aan te bieden (*fog computing* en *edge computing*). Zo zou een toestel enkel op basis van informatie op dat toestel zelf kunnen laten weten in welke advertenties de gebruiker geïnteresseerd is, zonder dat er gevoelige informatie in de cloud opgeslagen hoeft te worden. Ook sociale netwerken online zouden op deze manier ontworpen kunnen worden. Gebruikers verwachten vandaag de dag een automatische synchronisatie tussen smartphones, tablets en laptops en een automatisch mechanisme voor back-ups. Dit is mogelijk met een lokale server die gedeeld wordt binnen een gezin of tussen vrienden. Op deze manier blijft alle informatie lokaal. Als ze toch in de cloud wordt opgeslagen, kan dit in vercijferde vorm gebeuren: dit betekent dat de *cloud provider* de informatie niet kan lezen, wat misbruik voorkomt.

Voor een aantal toepassingen is het essentieel dat data centraal verwerkt worden. Zo kan men denken aan complexe analyses op grote medische datasets. In dat geval heeft men strikte regels nodig voor de beveiliging, toegang, loggen en audit. Maar ook cryptografie kan hier helpen: er wordt de jongste jaren steeds meer vooruitgang geboekt in het uitvoeren van bewerkingen op vercijferde data. Dit betekent dat een server kan zoeken in data die vercijferd zijn, op voorwaarde dat de zoektermen op het moment van de vercijfering bekend waren. Ook is het mogelijk om statistische functies (gemiddelden, correlaties, eenvoudige veeltermen)

efficiënt te berekenen op vercijferde data. Het resultaat is een 'cijfertekst', die enkel ontcijferd kan worden door wie de sleutel kent. In zijn meest algemene vorm (volledig homomorfe vercijfering) kan men zo elke mogelijke operatie uitvoeren, maar helaas is dit nog heel traag en duur. Tot slot laat cryptografie toe dat verschillende partijen samenwerken om over een netwerk berekeningen uit te voeren op al hun data samen, terwijl er toch geen informatie lekt over die gegevens. Met deze 'magische' oplossingen is het mogelijk om data enerzijds sterk te beschermen en er anderzijds toch nuttige informatie uit te halen. De kosten zijn nog altijd heel wat hoger dan de huidige oplossingen, waarbij alle data in een grote database terechtkomen, maar voor een aantal toepassingen en datasets moeten er maatschappelijke keuzes gemaakt worden om de data op een gedistribueerde manier te verwerken.

Casus 5: *Connected and Autonomous Driving (CAD)*[2]

In de uitgelekte communicatie [Commission EU Parliament] van de Europese Commissie over *Het bouwen van een Europese data economie* werd '*connected and autonomous driving*' (CAD) voorgesteld als een testcase om te onderzoeken hoe de uitwisseling van gegevens binnen de interne markt van de EU zo efficiënt, effectief en verantwoord mogelijk plaats kan vinden. CAD is vooral interessant als voorbeeld van het al vermelde zogenaamde Internet van de Dingen, een cyberfysieke infrastructuur die 'dingen' (in dit geval voertuigen) online verbindt en via de cloud met elkaar in gesprek laat gaan (machine-2-machine communicatie). Het is daarnaast een mooi voorbeeld van gegevensstromen die deels bestaan uit persoonsgegevens en deels uit niet-persoonsgegevens (de toestand van de weg, de gladheid van de banden, verkeersdrukte), terwijl het onderscheid in veel gevallen voorlopig zal zijn omdat ook niet-persoonsgegevens op enig moment gerelateerd kunnen worden aan een identificeerbare persoon, waardoor ze alsnog als persoonsgegevens gekwalificeerd worden.

Zowel in België als in andere landen wordt overwogen om de politie bij ernstige verkeersongelukken toegang te geven tot de data uit de zwarte doos, die vanaf 2018 verplicht is in alle nieuwe auto's om automatische noodoproepen bij ongevallen mogelijk te maken (eCall). De zwarte doos of *event data recorder* (EDR) kan in beginsel allerhande gegevens opslaan en/of verzenden. Voorlopig is uitsluitend het doorgeven van een 112-noodoproep verplicht, maar eenzelfde zwarte doos zou gebruikt kunnen worden om data vast te leggen van de laatste vijf seconden voor een crash. Dat geeft inzicht in de snelheid en het remgedrag, die vroeger op basis van fysieke sporen werden vastgesteld. De introductie van *Antilock Brake Systems* en *Electronic Stability Programs* leidt ertoe dat dergelijke sporen steeds vaker ontbreken, waardoor een EDR uitkomst kan bieden over de

[2] Deze tekst is een modulatie van [Hildebrandt 2017], op uitnodiging geschreven voor het Nederlandse tijdschrift *Ars Aequi*.

toedracht van een ongeluk [Post-Crash Voertuig Diagnose]. Zonder toegang tot deze data zou een bestuurder bij roekeloos gedrag zijn gerechte straf kunnen ontlopen – of ten onrechte verdacht blijven.

Naast een EDR bevatten auto's inmiddels een reeks van andere systemen die data registreren en eventueel bewaren en/of doorsturen. Denk aan de boordcomputer (het besturingssysteem van de auto, met de hierboven genoemde rem- en stabilisatiesystemen), het navigatiesysteem, eventuele rijstijldetectoren (sensoren) en zelfs de meegebrachte mobiele telefoon. Zij vormen een schier onuitputtelijke bron van data. Sommige daarvan betreffen de toestand van de auto (de tank is bijna leeg, het oliepeil is te laag, de banden zijn te glad), andere hebben meer direct te maken met het gedrag van de inzittenden en/of de bestuurder: de riemen zijn al dan niet aangegespt, de auto is niet afgesloten, de verlichting is uit. Weer andere hebben direct betrekking op het gedrag van personen: de snelheid, het remgedrag, vermoeidheidssymptomen, eten of bellen tijdens het rijden, of zelfs gesprekken met medepassagiers die uit de hand lopen. Locatiegegevens betreffen intussen de mobiliteit van zowel de auto als die van de bestuurder en andere inzittenden.

Het samenstel van gegevens waar het hier om gaat laat toe om zeer specifieke profielen te ontwerpen van individuele personen: van reisgedrag en rijstijl tot en met een inschatting van persoonlijkheidskenmerken. Daarenboven kunnen uit de geaggregeerde rijgedragsgegevens allerlei predicties worden afgeleid over het individuele rijgedrag en het rijgedrag van bepaalde types weggebruikers: oudere bestuurders, rokers, jonge mannen, allochtonen, vegetariërs, werklozen, brildragers enz. Het is van belang om twee types profielen te onderscheiden: (1) een individueel profiel bestaande uit *historische persoonsgegevens* en (2) een individueel of groepsprofiel dat bestaat uit *predicties* van toekomstig gedrag, gebaseerd op geaggregeerde databestanden van een veelheid van voertuigen, bestuurders, inzittenden en de slimme omgeving van de weggebruikers. Dit soort predicties kunnen aanleiding zijn tot risicoanalyses en tot al dan niet geautomatiseerde beslissingen. Bij de overgang van *connected driving* naar *autonomous driving* zal het steeds vaker gaan om geautomatiseerde beslissingen die gebaseerd zijn op het remgedrag van een autonoom voertuig dat 'in gesprek is' met andere voertuigen en zo een botsing voorkomt. Voor zover het dan niet gaat om een evaluatie van de inzittende, noch om een oordeel over diens rijvaardigheid, risicoperceptie of risicovol gedrag zijn dat soort beslissingen niet direct relevant voor de privacy van de inzittenden. Maar zodra het gaat om evaluaties van personen komen een aantal fundamentele rechten in het vizier: naast privacy ook non-discriminatie en mogelijk ook de onschuldpresumptie. Denk aan verzekeringspremies die fluctueren op basis van voorspeld gevaarlijk rijgedrag, of aan inschattingen van de persoonlijkheid van de betrokkene (neiging tot impulsief gedrag) die doorsijpelen naar *data brokers* die ze weer doorverkopen aan werkgevers of opsporingsdiensten. Dit is geen sciencefiction, al staan we nog maar aan het begin van deze en soortgelijke ontwikkelingen [Viereckl 2016] .

Dit soort voorspellende profielen, die op basis van voortdurende datastromen steeds worden aangepast, scheppen een nieuw type 'doorzichtigheid'. Personen worden 'doorzichtig' in de zin dat hun profiel het toestaat om dwars door de persoon heen naar – statistisch – vergelijkbare personen te kijken en op basis daarvan een aantal geautomatiseerde beslissingen te nemen. Denk opnieuw aan het verhogen of verlagen van de verzekeringspremie, het ontzeggen of opschorten van de rijbevoegdheid, ingrepen in snelheid of remgedrag, maar ook aan het monitoren van individuele personen op grond van de voorspelde gevaarzetting. Dit laatste tast niet alleen de autonomie van individuele personen aan, maar creëert een heel nieuwe 'keuzearchitectuur', waardoor het gedrag van burgers en consumenten in toenemende mate bewust wordt aangestuurd, ingeperkt en bijgestuurd. Burgers en consumenten zijn zich daar intussen niet van bewust, zij hebben weinig zicht op deze 'sturing' [Sunstein 2016] [Yeung 2017].

De privacyproblematiek spitst zich dan ook toe op het feit dat, terwijl het gedrag van personen steeds meer voorspelbaar wordt gemaakt, de gevolgen van hun gedrag voor henzelf steeds moeilijker te voorzien zijn. Dat leidt tot een fundamenteel soort onzekerheid, en een problematische vorm van mogelijke manipulatie. Met name wanneer CAD-data – eventueel via *data brokers* – zouden worden doorgespeeld naar andere contexten (werkgevers, gezondheidszorg, politie), lijkt de privacy volkomen zoek.

Oplossingsrichtingen: *Voorzienbaarheid en profieltransparantie*. Zoals hierboven aangegeven kunnen voorspellingen worden afgeleid uit het machinaal leesbare rijgedrag van een persoon ofwel uit de geaggregeerde data van heel veel personen. Hoewel de profielen in dat laatste geval zelf *niet* gerelateerd zijn aan een bepaalde persoon en dus zelf *geen* persoonsgegeven zijn, valt de toepassing op een persoon die binnen de 'gelding' van het profiel past wel onder het fundamentele recht op gegevensbescherming. Het is van groot belang om te kunnen voorspellen welke geautomatiseerde beslissingen op grond van dit type profielen kunnen worden genomen en een beeld te krijgen van de logica waarmee profielen zijn afgeleid. Zoals in het eerste deel van dit Standpunt is uiteengezet, mogen van big data en machinaal leren geen wonderen worden verwacht. Het is dus zaak dat de toegepaste kunstmatige intelligentie kan worden getest en, voor zover daarop gebaseerde beslissingen een serieuze invloed hebben op het leven van betrokkenen, ook dat die beslissingen in rechte kunnen worden aangevochten. Zowel de huidige privacywetgeving als de komende Algemene Verordening Gegevensbescherming (AVG) bieden een recht op profieltransparantie, waarbij de verplichting om betrokkenen op de hoogte te stellen en uit te leggen hoe zij worden geprofileerd vooropstaat. Nu de AVG, anders dan de huidige wetgeving, in een effectief handhavingskader voorziet, lijkt deze profieltransparantie een cruciale oplossingsrichting te bieden.

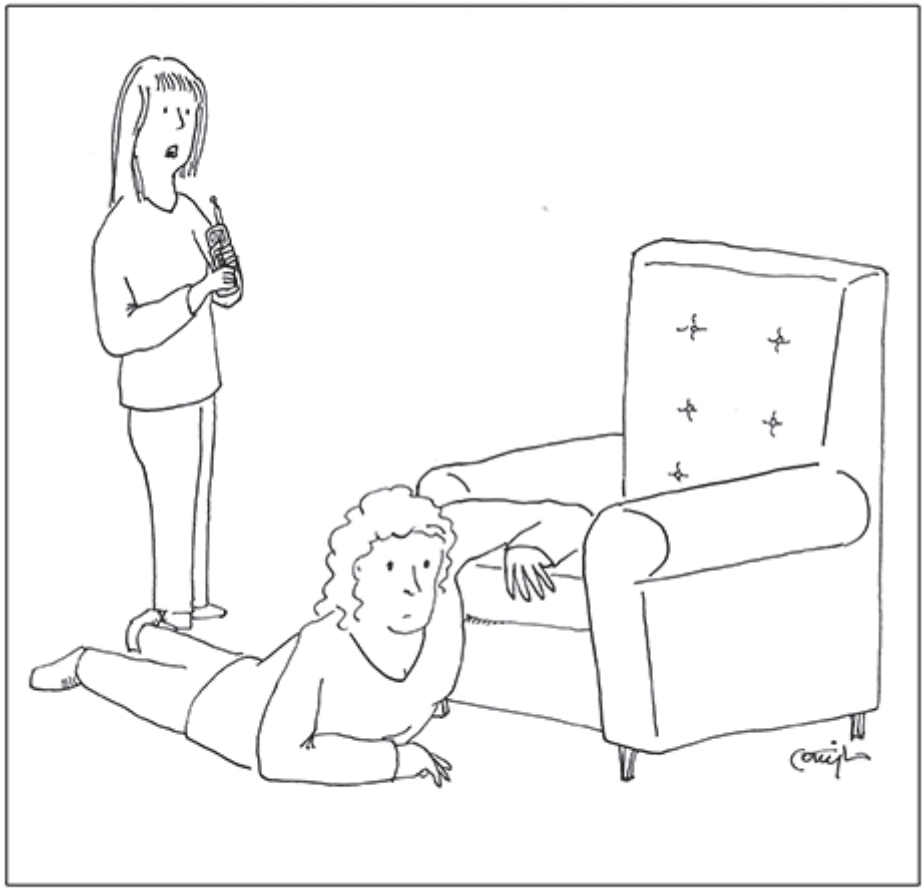
Voorzienbaarheid en doelbinding Daarnaast is het van belang dat burgers en consumenten een goed beeld hebben van het doel waartoe hun CAD-data worden

verwerkt. Sommigen menen dat doelbinding niet meer van deze tijd is, omdat het big data en machinaal leren zou hinderen. Alleen het ongelimiteerde verzamelen en analyseren van CAD-data zou de toegevoegde waarde van de data kunnen ontsluiten, juist omdat men dan alle relevante data zou kunnen verkrijgen (n=all). Dat is zeker niet het geval. Zoals uiteengezet in de begripsverkenning rond big data en machinaal leren, vereist een effectieve data-analyse een zorgvuldige afweging van allerhande trade-offs, waarbij juist het doel van de analyse bepalend zal zijn voor de methodologische keuzes. Dat doel maakt het bovendien mogelijk om wie de voordelen van de data wenst te oogsten bij de les te houden en eventueel ter verantwoording te roepen op basis van de voorzienbare inzet van die gegevens.

Casus 6: Informatievergaring over locaties

Een belangrijke trend in het aanbieden van diensten is het aanpassen van die diensten aan de context. Elementen die hierbij een rol spelen, zijn het algemene profiel van de gebruiker en het toestel van de gebruiker en ook de locatie. Bij vaste toestellen kan de locatie leiden tot prijsdiscriminatie (in duurdere landen of buurten worden hogere prijzen aangerekend); in meer extreme gevallen wordt bepaald of een dienst al dan niet beschikbaar is (zoals Netflix of BBC iPlayer). Bij mobiele toestellen kan de locatie meer nauwkeurig bepaald worden en wordt het mogelijk om restaurants, winkels of benzinestations in de buurt weer te geven. Meer gesofisticeerde diensten zijn ook mogelijk, zoals het aanbieden van verkeersinformatie, advertenties of kortingsbonnen afhankelijk van de locatie of de melding wie uit de vriendenkring in de buurt is. Het lijdt geen twijfel dat de verwerking van locatie-informatie, samen met de respons van de gebruiker op deze diensten, commercieel zeer waardevol is. Anderzijds kan locatie-informatie zeer gevoelig zijn: zo kun je het thuis- en werkadres achterhalen, maar ook de sociale klasse, religie en informatie over de gezondheidstoestand. Onderzoek toont aan dat locatie-informatie op grote schaal verzameld wordt door de grote internetspelers, maar ook verwerkt wordt door ontwikkelaars van mobiele apps en bedrijven die aan deze ontwikkelaars advertentiesoftware aanbieden. Zelfs als de naam van de gebruiker niet bewaard wordt, vormen onze locaties een unieke identificatie: zo toont een studie van 2013 aan dat vier locaties en tijdstippen volstaan om 95% van de gebruikers te identificeren [deMontjoye13].

Naast de juridische bescherming, die o.m. vereist dat de gebruiker expliciet de toestemming geeft om zijn informatie te verwerken, is het mogelijk om door technische maatregelen locatiegegevens te beschermen bij het aanbieden van locatie-afhankelijke diensten [Shokri11]. In het meest extreme geval verbergt men zijn locatie; met technieken als *private information retrieval* kan men informatie uit een databestand halen terwijl de dienstenaanbieder niet weet welke informatie er gevraagd is. Men kan zijn locatie aanpassen door een fout van een paar honderd meter of een paar kilometer toe te voegen: een grotere aanpassing biedt meer privacy, maar maakt de dienst minder nuttig. Een andere oplossing is het toevoegen van locaties waar men helemaal niet geweest is: dit is aanvaardbaar



It's Google. They say you left your keys in the left-hand pocket of your other pants.

voor het zoeken van een restaurant, maar minder geschikt als men op deze manier vrienden in de buurt zoekt. Tot slot kan men de nauwkeurigheid van de locatie-informatie verminderen, door aan te geven dat men zich in een bepaald gebied bevindt, bijvoorbeeld op minder dan twee kilometer van het centrum van Brussel, zonder verdere details te geven.

Welke oplossing het beste is, hangt af van de specifieke dienst. Voor vaste verbindingen kan men met Tor zijn IP-adres verbergen, en dus ook de locatie. Het Tor-netwerk is ontworpen om te verhinderen dat anderen door analyse van het berichtenverkeer kunnen achterhalen wat de herkomst en bestemming

van berichten is. Tor biedt bescherming tegen eigenaars van websites en telecommunicatieoperatoren, maar niet tegen overheden die over een groot aantal interceptiepunten beschikken. Met Android-telefoons heeft Orbot hetzelfde doel, maar dat betekent niet dat de locatie ook verborgen is. In de praktijk kan men er bij een aantal diensten voor kiezen om zijn locatie niet te tonen aan derden, maar is het vrijwel onmogelijk om zijn locatie echt verborgen te houden. Apple of Google kennen de locatie van een groot aantal Bluetooth- en Wifi-netwerken en kunnen op deze manier je locatie bepalen, zelfs als je de Wifi- of GPS-functie uitschakelt. De enige manier om dit te beletten is je iPhone te *jailbraken* of Android-telefoon te *rooten* en vervolgens de nodige tools te installeren. Je bent dan wel meer kwetsbaar voor hackers. Op een vergelijkbare manier kunnen netwerkoperatoren nagaan in welke 2G/3G/4G-cel een gebruiker zich bevindt; zeker in de steden zijn deze cellen zeer klein. In dit geval is de enige[3] oplossing de telefoon in vluchtmode te zetten of uit te schakelen, maar dan verliest hij uiteraard een belangrijk deel van zijn functies.

De conclusie is dat locatiegegevens zeer gevoelige gegevens zijn, terwijl het in de praktijk vrijwel onmogelijk is een mobiel toestel te gebruiken en toch ook zijn locatie te verbergen voor de grote internet- en telecommunicatiespelers [deMontjoye13], [Shokri11].

4. Conclusies en aanbevelingen gericht aan doelgroepen

Het toenemende gebruik van sociale media, de cloud, internet en smartphones, en de kracht van het machinaal leren uit de big data bij ICT-bedrijven en grote instellingen, creëert een heel nieuwe context wat de privacy van jong en oud betreft. Hoewel de evolutie nog volop bezig is, vergt dit een geheel van maatregelen bij verantwoordelijken voor de verwerking, en meer bewustwording bij het brede publiek.

4.1 Conclusies

Wat kan tegen al deze problemen ondernomen worden? De controle van de persoonsgegevens vereist eerst en vooral inzicht bij het individu in het gebruik en misbruik van deze gegevens, en echte keuzevrijheid. Deze voorwaarden zijn noodzakelijk voor de bescherming van de grondrechten, meer bepaald het fundamenteel recht op de bescherming van persoonsgegevens. We moeten evenwel onder ogen zien dat de gevolgen van big data en machinaal leren complex en vaak onoverzichtelijk zijn, waardoor een goed geïnformeerde keuze voor individuen een illusie lijkt. De partijen die garen spinnen bij de analyse van big data zullen dan ook hun verantwoordelijkheid moeten opnemen, en daartoe

[3] Er zijn ook wettelijke garanties die het gebruik van locatiegegevens door operatoren beperken. [Wet op de Elektronische Communicatie].

ook wettelijk worden verplicht. Men kan immers niet verwachten dat partijen zich ethisch gedragen als ze daardoor uit de markt worden gedrukt – er moet een gelijk speelveld komen, waarin 'bij verstek' bescherming wordt geboden.

Uiteraard moeten we bij oplossingen denken aan allerhande ICT-methodes, zoals cryptografie, en andere vormen van beveiliging, zoals anonimisering of pseudonimisering. Daarnaast is er de educatie: data- en reclamewijsheid/geletterdheid vanaf jonge leeftijd, maar ook voor oudere generaties. Verder dient een aanpak van empowerment en verzet zich aan. Ten slotte zou dit alles bekrachtigd moeten worden in een stevig juridisch kader: enerzijds is er de al genoemde gegevensbeschermingswetgeving op Europees niveau (transparantie, doelbinding, strafrecht en veiligheidsdiensten, data-minimalisering en de scheidingsbenadering) en anderzijds de privaatrechtelijke handhaving. Ze moeten worden afgestemd op de sociale en technologische context, en rekening houden met het gebrek aan kennis bij individuen. [Council of Europe 2017].

De concentratie aan data bij de grote spelers in het ICT-domein van het web en de sociale media, en de krachtige profileringstechnieken en businessmodellen die zij hanteren, vereisen een internationale regulering en actieve toezichthouders die over voldoende middelen beschikken om die regulering te handhaven.

De verantwoordelijke verwerkers spelen een belangrijke rol om de waarschijnlijke gevolgen van hun beoogde verwerking van gegevens af te wegen tegen de grondrechten en de fundamentele vrijheden van de betrokkenen. De Europese wetgever eist daarbij dat ze de risico's van elke verwerkingsactiviteit met big data en het potentiële negatieve resultaat voor de rechten van het individu en de fundamentele vrijheden identificeren en evalueren, in het kader van het recht op de bescherming van persoonsgegevens, met het oog op het recht op non-discriminatie en rekening houdend met de sociale en ethische gevolgen. Door passende maatregelen, zoals privacy 'per ontwerp' en 'per verstek', kunnen de risico's beperkt worden. Ten slotte zullen de verantwoordelijken de effectiviteit van de oplossingen moeten controleren en goedkeuren, onder het toeziend oog van de autoriteit inzake gegevensbescherming.

'Empowerment' is het proces van het versterken van gebruikers/burgers/consumenten, waardoor ze greep krijgen op hun eigen situatie en hun omgeving door het verwerven van controle, het aanscherpen van het kritische bewustzijn en het stimuleren van participatie. Een van de grootste uitdagingen in dit verband is het verstaanbaar en betekenisvol maken van de complexiteit van technologische aspecten die impact hebben op (meta)data en privacy, zoals algoritmes, API's, *machine learning*, (verstek) privacy settings en AB testing. Onderwijs en opleiding kunnen mensen op alle leeftijden helpen om de implicaties van het gebruik van hun persoonsgegevens in het kader van big data te leren begrijpen. Daarom moeten scholen en onderwijsinstellingen informatie- en digitale geletterdheid als een essentiële educatieve vaardigheid beschouwen.

Met name over de zogenaamde metadata is een meer diepgaande bewustwording vereist van de zeer persoonlijke profilering die zij mogelijk maken. Zeker locatiegegevens vergen de nodige aandacht. Er bestaan vrijwel geen diensten waardoor een gebruiker zijn locatie kan verbergen voor de aanbieder van de diensten, de grote spelers in de internetwereld en de mobiele operatoren, zonder uitgesloten te worden van de betreffende dienstverlening (waardoor ook die van de hardware en het operating system).

Omdat de verantwoordelijkheden over diverse actoren verspreid zijn, is er ten slotte nood aan een publiek debat en multistakeholderoverleg met alle betrokken instanties. Een belangrijke verantwoordelijkheid ligt bij de overheid en de industrie, maar altijd in nauwe samenspraak met toezichthouders, wetenschappers, middenveldorganisaties en burgers. Alleen zo kan er een vorm van 'samenwerkende verantwoordelijkheid' gerealiseerd worden om tot doeltreffende en gedragen oplossingen te komen die de publieke waarden en fundamentele rechten in de digitale samenleving in stand houden en versterken.

4.2 Aanbevelingen

De aanbevelingen hieronder worden geformuleerd met het oog op de rol van de burger, de overheid, de ICT-bedrijven en de ontwerpers en onderzoekers van ICT-diensten. Het recent uitgebrachte rapport van de Belgische Privacycommissie [CBPL 2017] bespreekt uitgebreid de privacyproblematiek van big data. Het bevat 33 waardevolle aanbevelingen, die zich voornamelijk richten op de verantwoordelijken voor de verwerking in de organisaties en bedrijven, en op de overheden in het licht van AVG. De aanbevelingen 2, 8, 11, 12, 17 en 23 uit dat rapport keren hieronder terug en worden toegelicht. Daarnaast worden nog vier aanbevelingen gegeven voor de burger, het onderwijs, en de ICT-ontwerpers van diensten en de overheid.

Aanbeveling 1: Verantwoordelijkheden. Privacy wat big data betreft is een zaak van de burgers, ingenieurs, consumenten, bedrijven, instellingen, media en overheden. Dit neemt niet weg dat grote spelers die baat hebben bij big data-analyse, een grotere verantwoordelijkheid hebben, terwijl de overheid een eindverantwoordelijkheid heeft voor het respecteren van de mensenrechten en er dus zorg voor moet dragen dat de juiste inspanningen op het juiste niveau worden geleverd. Dit vraagt om het beschikbaar stellen van voldoende middelen aan de toezichthouders, met name ten aanzien van bedrijven die hun verdienmodel uit de analyse van big data halen, waarbij bovendien van de overheden zelf het goede voorbeeld wordt verwacht. Het gaat daarnaast om het ontwikkelen van onderwijs over privacy, waarbij goede praktijken en producten worden gestimuleerd en onder de aandacht worden gebracht. Onder meer consumenten- en andere middenveldorganisaties hebben een belangrijke rol te spelen, bijvoorbeeld bij het vergelijken van de privacyvoorwaarden van big data- en internet diensten, maar

bijvoorbeeld ook bij het uitoefenen van gemandateerde rechten van inzage en verzet.

Aanbeveling 2: Alerte burgers. Burgers van wie gegevens worden verwerkt moeten hun rechten onder de AVG maximaal proberen uit te oefenen. De controle van de persoonsgegevens vereist inzicht van het individu in het gebruik en misbruik van deze gegevens, omdat alleen dan sprake kan zijn van echte keuzevrijheid. Omdat dit voor individuen buitengewoon moeilijk is, bevelen wij aan dat de betrokkenen gebruik maken van de mogelijkheid om hun aanspraken uit te oefenen via een mandatering aan consumenten- of privacy organisaties (art. 80 AVG). Langs die weg zal het effectief mogelijk zijn om de verantwoordelijke voor de verwerking te contacteren (bv. in het kader van het recht van inzage, kopie of beperking), wanneer de beschikbare informatie over het gebruik van hun gegevens te weinig transparant, onvolledig of te vaag is, of wanneer iemand de toewijzing van een bepaald kenmerk (bv. fraudeur) op zijn persoon betwist.

Aanbeveling 3: Voorzienbaarheid, profieltransparantie en doelbinding. Hoewel de profielen zelf *niet* gerelateerd zijn aan een persoon en dus zelf *geen* persoonsgegeven zijn, valt de toepassing op een persoon die binnen de 'gelding' van het profiel past, w*é*l onder het fundamentele recht op gegevensbescherming (AVG). Het recht op profieltransparantie houdt de verplichting in om de betrokkenen op de hoogte te stellen en uit te leggen hoe zij worden geprofileerd. Dit gaat verder dan een correlatie of statistisch verband. Daarnaast is het van belang dat burgers en consumenten een goed beeld hebben van het doel waartoe hun persoonsgegevens worden verwerkt.

Aanbeveling 4: Machtsonevenwicht. Indien de verantwoordelijke van een ICT-dienst zich beroept op de toestemming voor het gebruik van persoonsgegevens, dan moet die toestemming makkelijk in te trekken zijn, en steeds beperkt zijn in de tijd. Ze zal bovendien niet gelden bij een manifest machtsonevenwicht tussen de betrokkene en de verantwoordelijke of verwerker, bv. omdat de verantwoordelijke de dominante (of enige) dienst in de markt levert. De verantwoordelijke zal moeten aantonen dat er geen machtsonevenwicht is of dat dit onevenwicht de toestemming van de betrokkene niet kan beïnvloeden.

Aanbeveling 5: De *bouwers van ICT- en IoT-apparaten* moeten werk maken van technologieën die de privacy behouden en die transparantie bieden aan de eindgebruiker, zoals privacy bij ontwerp, waardoor privacy van bij het begin van het ontwerp als een belangrijke vereiste meegenomen wordt en er niet achteraf aan 'geplakt' wordt. De *service providers* moeten toelaten dat gebruikers diensten van verschillende oorsprong samenstellen. De *ontwerpers van algoritmes* moeten die zo schrijven dat ze de privacy van de gebruikers garanderen. De ontwerpers van *toepassingen* moeten transparantie toelaten en werk maken van efficiënte en effectieve technologieën, waarbij gebruikers de toestemming kunnen geven

voor het gebruik van hun data. Verder moet men werk maken van het certificeren van toepassingen, zodat de gebruikers zeker zijn dat de toepassingen veilig zijn. Voor zowel de apparatuur als de toepassingen is er een grote nood aan betere en fijnmazige methodes om instellingen en voorkeuren op te geven. Hier moet maximale privacy 'bij verstek' de regel zijn (*privacy 'by default'*). Er is ook nog veel werk voor verdere *regelgeving*. Die moet zorgen voor standaardisatie (zonder een rem te zetten op innovatie) en een basis geven aan certificatie.

Aanbeveling 6: Rol van de overheid en de bedrijven. Het is de taak van de overheid en de bedrijven om voor elke big data-oplossing grondig af te wegen of de voordelen opwegen tegen de risico's voor de bescherming van persoonlijke gegevens en voor de maatschappij als geheel (wat als deze gegevens uitlekken?). Daarbij moet men steeds nagaan of het niet mogelijk is om hetzelfde doel te bereiken door minder gegevens te gebruiken of gegevens te aggregeren. Dit volgt uit de grondbeginselen van de Europese regelgeving: gegevensbescherming door ontwerp (*data protection by design*) en door standaardinstellingen (*data protection by default*). In het licht van het grote aantal gelekte dataverzamelingen is het zaak dat de autoriteit Gegevensbescherming haar nieuwe bevoegdheden inzet om effectieve oplossingen met gegevensbescherming door ontwerp af te dwingen, waarbij data zo veel mogelijk lokaal blijven en er zo weinig mogelijk centraal worden verzameld.

Aanbeveling 7: Vermijden van onwenselijke data bias. De verantwoordelijke ontwerpers en dienstenleveranciers moeten steeds nagaan of er onjuiste dan wel oneerlijke *data bias*, *algoritme bias* dan wel *output bias* verscholen zitten in de datasets waarmee algoritmes worden getraind. Dat kan in de wiskundige modellen zelf zijn, of in de output (indirecte discriminatie). Vragen zullen moeten worden beantwoord zoals onder meer: waarom worden bepaalde bevolkingsgroepen uitgesloten? Welke datapunten zijn minder zichtbaar bij training of tests? Hierbij kan gebruik worden gemaakt van *discrimination aware data mining*.

Aanbeveling 8: Grenzen aan het gebruik van big data door de overheid. Het gebruik van big data in de publieke sector – zowel voor de detectie van belasting- en sociale-zekerheidsfraude als in het kader van de nationale veiligheid, criminaliteitsbestrijding en ordehandhaving – moet steeds worden onderworpen aan een onderzoek door de relevante toezichthouders. Daarin zullen de rechtmatigheid en de daarmee verbonden proportionaliteit voorop moeten staan, wat ook steeds een marginale doelmatigheidstoets vereist. Het is van groot belang dat er een wettelijke regeling komt die bepaalt hoe en wanneer het resultaat van *data mining* en statistische analyses (correlaties) door de overheid al dan niet gebruikt kan worden als juridisch bewijsmateriaal in individuele dossiers (bv. bij de aanpak van fraude, de ordehandhaving...).

Aanbeveling 9: Digital clearing house oprichten. Het verdient aanbeveling om een *Digital Clearing House* (DCH) op te richten, dat de kwaliteit bewaakt van de diverse regulatoren op de digitale markt.

Aanbeveling 10: Taak van het onderwijs. Specifiek voor jongeren heeft het onderwijs een taak te vervullen op het vlak van bewustwording, attitudes, vaardigheden en gedragingen in de concrete levenssferen waarin zij verkeren: thuis, op school, in de vriendenkring (bv. jeugdverenigingen)... Het is belangrijk jongeren te wijzen op de valkuilen van hun gedrag, zoals die bijvoorbeeld tot uiting komen in de privacyparadox.

Bibliografie

[Angwin 2016] J. Angwin e.a., 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks', *ProPublica*, 23 May 2016, zie www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[AVG 2016] Algemene Verordening Gegevensbescherming, Regulation EU 2016/679. Aangenomen op 27 april 2016 en treedt in werking op 25 mei 2018. http://eur-lex.europa.eu/legal-content/NL/TXT/?uri=CELEX%3A32016R0679. Engelse versie zie [GDPR, 2016].

[Ballon 2016] P. Ballon, *Smart Cities: Hoe technologie onze steden leefbaar houdt en slimmer maakt*, Tielt: Lannoo, 221.

[Berendt and Preibusch 2014] B. Berendt and S. Preibusch, 'Better Decision Support through Exploratory Discrimination-Aware Data Mining: Foundations and Empirical Evidence', *Artificial Intelligence and Law* 22, no. 2 (1 June 2014): 175–209, doi:10.1007/s10506-013-9152-0.

[Belmans e.a. 2016] R. Belmans, P. Vingerhoets, I. Van Vaerenbergh e.a., *De eindgebruiker centraal in de energietransitie*, KVAB Standpunt 44, 2016.

[Cabitza 2016] F. Cabitza, 'The Unintended Consequences of Chasing Electric Zebras', IEEE SMC Interdisciplinary Workshop HUML 2016, The Human Use of Machine Learning, 12/16/ 2016, Venice, Italy https://www.researchgate.net/publication/311702431_The_Unintended_Consequences_of_Chasing_Electric_Zebras

[CBPL 2017] Commissie voor de bescherming van de persoonlijke levenssfeer, Big Data Rapport AH-2016-0154, https://www.privacycommission.be/nl/publieke-consultatie-big-data-rapport

[Commission EU Parliament] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: 'Building a European Data Economy', www.euractiv.com/wp-content/uploads/sites/2/2016/12/data-communication.pdf.

[Council of Europe 2017] Council of Europe - Guidelines on the Protection of Individuals with Regard to the Processing of Personal Data in a World of Big Data http://rm.coe.int/16806ebe7a

[Coursera] https://www.coursera.org/learn/friends-money-bytes/lecture/CKluM/selling-ad-spaces-through-auctions

[DCS] Digital clearinghouse DCS https://secure.edps.europa.eu/EDPSWEB/edps/site/mySite/Big_data_rights_Lets_get_together

[De Hert 2011] P. De Hert, & R. Bellanova, R., 'Mobility should be fun. A consumer (law) perspective on border check technology', *The Scientific World JOURNAL*, 2011, vol. 11, 490-502.

[Demetriou 2016] S. Demetriou, W. Merrill, W. Yang, A. Zhang, C. A. Gunter: *Free for All! Assessing User Data Exposure to Advertising Libraries on Android*, NDSS, 2016.

[deMontjoye2013] Y.-A. de Montjoye, C.A. Hidalgo, M. Verleysen, V. D. Blondel, *Unique in the Crowd: The privacy bounds of human mobility*. Nature Scientific Reports 3, 2013.

[DG Int.Pol. 2015] Big Data and Smart Devices and Their Impact on Privacy. Study for the LIBE Committee. Directorate General for Internal Policies, Policy Department Citizen's Rights and Contitutional Affairs. http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf

[Digimeter 2016] Digimeter report 2016 http://www.imec-int.com/en/digimeter

[Digitale meters 2017] https://www.vlaanderen.be/nl/nbwa-news-message-document/document/09013557801c194c

[Dual use] http://ec.europa.eu/trade/import-and-export-rules/export-from-eu/dual-use-controls/

[EU Data protection directive 1995] http://ec.europa.eu/justice/policies/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf

[Facebook] https://www.facebook.com/business/help/430291176997542

[Gadamer 2010] H. G. Gadamer, *Gesammelte Werke: Band 1: Hermeneutik I: Wahrheit und Methode: Grundzüge einer philosophischen Hermeneutik*, 7., durchges. A. edition, Tübingen: Mohr Siebeck, 2010.

[GDPR, 2016] General Data Protection Regulation, (GDPR) Regulation EU 2016/679. Aangenomen op 27 april 2016 en treedt in werking op 25 mei 2018. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679. Nederlandse versie zie [AVG 2016].

[Google] https://support.google.com/adwords/answer/2996564?hl=nl

[Hildebrandt 2016] M. Hildebrandt, 'Data-gestuurde intelligentie in het strafrecht, Preadvies Nederlandse Juristen Vereniging 2016', in: E.M.L. Moerel e.a., *Homo Digitalis* (Handelingen Nederlandse Juristen-Vereniging 2016-I), Den Haag: Wolters Kluwer 2016, p. 137-240, zie ook http://njv.nl/preadviezen/preadviezen-2016/

[Hildebrandt 2017] M. Hildebrandt, 'Wat weet mijn auto nog meer? Juridische bescherming by design in tijden van Internet van de Dingen', Ars Aequi, februari 2017, 97-102.

[Juniper routers trapdoor] https://en.wikipedia.org/wiki/Dual_EC_DRBG

[Mitchell, 1997] T. Mitchell, Machine Learning, McGraw Hill, 1997.

[NAP 2016], E. Grumbling, Privacy Research and Best Practices: Summary of a Workshop for the Intelligence Community, The National Academies Press, 2016.

[Perera 2015] C. Perera, R. Ranjan, L. Wang, S. U. Khan, A. Y. Zomaya, 'Big Data Privacy in the Internet of Things Era', IEEE IT Professional Magazine: Special Issue Internet of Anything 2015, Issue No.03 - May-June, 2015 vol.17.

[Post-Crash Voertuig Diagnose] http://www.p-crashvd.nl. Licentiehouder van een aantal merkspecifieke diagnosesystemen, zoals die van ODIS (Volkswagen, Audi, Skoda en Seat), VCDS (VAG-COM), BMW ISTA (BMW en Mini) en BMW Keyreader, Mercedes Xentry (Mercedes personenauto's en lichte bedrijfswagens), Volvo VIDA (Volvo personenauto's).

[PRISM] https://en.wikipedia.org/wiki/PRISM_(surveillance_program)

[Rathenau Inst. 2010] Rathenau Instituut, 'Databases. Over ICT-beloftes, informatiehonger en digitale autonomie' http://www.cs.ru.nl/B.Jacobs/PAPERS/Rapport_Databases_Rathenau_Instituut_nov_2010.pdf

[Royal Society] The Royal Society, 'Progress and research in cybersecurity,; supporting a resilient and trustworthy system for the UK?', 2016. http://royalsociety.org/cybersecurity

[Shokri 2011] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, J.-P. Hubaux, 'Quantifying Location Privacy', IEEE Symposium on Security and Privacy, 2011: 247-262.

[Smart metering in EU 2014] Smart metering deployment in the European Union http://ses.jrc.ec.europa.eu/smart-metering-deployment-european-union

[Sunstein 2016] C.R. Sunstein, The ethics of Influence. Government in the Age of Behavioral Science, Cambridge University Press, 2016.

[Sweeny 2013] https://arxiv.org/ftp/arxiv/papers/1301/1301.6822.pdf

[Tene, Polonetsky 2015] O. Tene, J. Polonetsky, 2015, 'A theory of creepy: Technology, privacy, and shifting social norms', Yale Journal of Law and Technology 16.1:2 (2015), http://digitalcommons.law.yale.edu/yjolt/vol16/iss1/2/

[UK Information Commissioner 2017] UK Information Commissioner, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection', http://iconewsblog.wordpress.com/2017/03/03/ai-machine-learning-and-personal-data/

[van Dijck 2016] J. van Dijck, T. Poell, & M. de Waal, 2016, De platformsamenleving: strijd om publieke waarden in een online wereld, Amsterdam: Amsterdam University Press, 180.

[Vanrykel 2016] E. Vanrykel, G. Acar, M. Herrmann, C. Diaz: 'Leaky Birds: Exploiting Mobile Application Traffic for Surveillance', in: *Financial Cryptography and Data Security – 20th International Conference, FC 2016, Lecture Notes in Computer Science 9603*, Springer-Verlag.

[Verdonck, Van Hulle 2017] M. Van Hulle en P. Verdonck e.a., 'Datawetenschappen en gezondheidszorg', KVAB Standpunt 48, 2017.

[Verheul et al. 2016] E. Verheul et al., 'Polymorphic Encryption and Pseudonymisation for Personalised Healthcare', 2016, https://eprint.iacr.org/2016/411

[Viereckl 2016] R. Viereckl e.a., 'Connected Car Report 2016: Opportunities, Risk, and Turmoil on the Road to Autonomous Vehicles' (PWC), http://www.strategyand.pwc.com/reports/connected-car-2016-study.

[VN2014] UN Department of Economic and Social Affairs, 'World Urbanisation Prospects'. United Nations, New York, 2014 revision, p. 1, http://esa.un.org/unpd/wup/Highlights/WUP2014-Highlights.pdf .

[Wareable] https://www.wareable.com/internet-of-things/whos-watching-your-smartwatch

[website-breaches] http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks

[website ik beslis] https://www.ikbeslis.be

[Wet Elektronische Communicatie] Wet Elektronische Communicatie, Art 123. http://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=nl&la=N&cn=2005061332&table_name=wet

[Wolpert 2013] D. H. Wolpert, 'Ubiquity Symposium: Evolutionary Computation and the Processes of Life: What the No Free Lunch Theorems Really Mean: How to Improve Search Algorithms', *Ubiquity 2013*, no. December: 2:1–2:15, doi:10.1145/2555235.2555237.

[Working Party] Working Party on data protection and privacy, European advisory body, 'Opinion 2/2010 on online behavioural advertising', 2010 http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp171_en.pdf

[WRR, 2016] Wetenschappelijke Raad voor het Regeringsbeleid WRR in Nederland 2016, WRR-rapport 95: *Big Data in een vrije en veilige samenleving* http://www.wrr.nl/publicaties/publicatie/article/big-data-in-een-vrije-en-veilige-samenleving/

[WRR 2017] Wetenschappelijke Raad voor het Regeringsbeleid WRR in Nederland WRR-Policy Brief 6: *Big Data and Security Policies: Serving Security, Protecting Freedom* 2017 https://www.wrr.nl/publicaties/publicaties/2017/01/31/big-data-and-security-policies-serving-security-protecting-freedom

[Yeung 2017] K. Yeung, '"Hypernudge": Big Data as a Mode of Regulation by Design', *Information, Communication & Society* (20) 2017, afl. 1, p. 118-136.

Samenstelling van de werkgroep

Yolande Berbers (KTW, KU Leuven)

Willem Debeuckelaere (Commissie voor de bescherming van de persoonlijke levenssfeer)

Paul De Hert (VUB en Tilburg University TILT)

Yvo Desmedt (KTW, University of Texas, University College London)

Frank De Smet (Commissie voor de bescherming van de persoonlijke levenssfeer)

Jos Dumortier (Advocatenkantoor Time.Lex)

Mireille Hildebrandt (University of Nijmegen, VUB)

Eleni Kosta (TILT Tilburg University)

Karolien Poels (JA, Universiteit Antwerpen)

Jo Pierson (VUB)

Yves Poullet (Centre de Recherche Information, Droit et Société CRID, FUNDP Namur)

Bart Preneel (ESAT COSIC, KU Leuven)

Joos Vandewalle (KTW, KU Leuven)

Karel Velle (KMW, Rijksarchief, UGent)

KTW = Klasse Technische Wetenschappen
KMW = Klasse Menswetenschappen
JA = Jonge Academie

RECENTE STANDPUNTEN (vanaf 2014)

26. Charles Hirsch, Erik Tambuyzer e.a. – *Innovatief ondernemerschap via spin-offs van kenniscentra*, KVAB/Klassen Natuurwetenschappen en Technische wetenschappen, 2014.

27. Giovanni Samaey, Jacques Van Remortel e.a. – *Informaticawetenschappen in het leerplichtonderwijs*, KVAB/Klasse Technische wetenschappen en Jonge Academie, 2014.

28. Paul Van Rompuy – *Leidt fiscale autonomie van deelgebieden in een federale staat tot budgettaire discipline?* KVAB/Klasse Menswetenschappen, 2014.

29. Luc Bonte, Paul Verstraeten e.a. – *Maatschappelijk verantwoord ondernemen. Meedoen omdat het moet, of echt engagement?* KVAB/Klasse Technische wetenschappen, 2014.

30. Piet Van Avermaet, Stef Slembrouck, Anne-Marie Simon-Vandenbergen – *Talige diversiteit in het Vlaams onderwijs: problematiek en oplossingen*, KVAB/Klasse Menswetenschappen, 2015.

31. Jo Tollebeek – *Metamorfoses van het Europese historisch besef, 1800-2000*, KVAB/Klasse Menswetenschappen, 2015.

32. Charles Hirsch, Erik Tambuyzer e.a. – *Innovative Entrepreneurship via Spin-offs of Knowledge Centers*, KVAB/Klassen Natuurwetenschappen en Technische wetenschappen, 2015.

33. Georges Van der Perre en Jan Van Campenhout (eds.) – *Higher education in the digital era. A thinking exercise in Flanders*, Denkersprogramma KVAB/Klasse Technische wetenschappen, 2015.

34. Georges Van der Perre, Jan Van Campenhout e.a. – *Hoger onderwijs voor de digitale eeuw*, KVAB/Klasse Technische wetenschappen, 2015.

35. Hugo Hens e.a. – *Energiezuinig (ver)bouwen: geen rechttoe rechtaan verhaal*, KVAB/Klasse Technische wetenschappen, 2015.

36. Marnix Van Damme – *Financiële vorming*, KVAB/Klasse Menswetenschappen, 2015.

37. Els Witte – *Het debat rond de federale culturele en wetenschappelijke instellingen (2010-2015)*, KVAB/Klasse Menswetenschappen, 2015.

38. Irina Veretennicoff, Joos Vandewalle e.a. – *De STEM-leerkracht*, KVAB/Klasse Natuurwetenschappen en Klasse Technische wetenschappen, 2015.

39. Johan Martens e.a. – *De chemische weg naar een CO_2-neutrale wereld*, KVAB/Klasse Natuurwetenschappen, 2015.

40. Herman De Dijn, Irina Veretennicoff, Dominique Willems e.a. – *Het professoraat anno 2016*, KVAB/Klasse Natuurwetenschappen, Klasse Menswetenschappen, Klasse Kunsten en Klasse Technische wetenschappen, 2016.

41. Anne-Mie Van Kerckhoven, Francis Strauven – *Een bloementapijt voor Antwerpen*, KVAB/Klasse Kunsten, 2016.

42. Erik Mathijs, Willy Verstraete (e.a.), *Vlaanderen wijs met water: waterbeleid in transitie*, KVAB/Klasse Technische wetenschappen, 2016.

43. Erik Schokkaert - *De gezondheidszorg in evolutie: uitdagingen en keuzes*, KVAB/Klasse Menswetenschappen, 2016.

44. Ronnie Belmans, Pieter Vingerhoets, Ivo Van Vaerenbergh e.a. – *De eindgebruiker centraal in de energietransitie*, KVAB/Klasse Technische Wetenschappen, 2016.

45. Willem Elias, Tom De Mette – *Doctoraat in de kunsten*, KVAB/Klasse Kunsten, 2016.

46. Hendrik Van Brussel, Joris De Schutter e.a., *Naar een inclusieve robotsamenleving*, KVAB/Klasse Technische Wetenschappen, 2016.

47. Bart Verschaffel, Marc Ruyters e.a., *Elementen van een duurzaam kunstenbeleid*, KVAB/Klasse Kunsten, 2016.

48. Pascal Verdonck, Marc Van Hulle (e.a.) - *Datawetenschappen en gezondheidszorg*, KVAB/Klasse Technische wetenschappen, 2017.

De volledige lijst met standpunten en alle pdf's kunnen worden geraadpleegd op
www.kvab.be/standpunten

Privacy in tijden van internet, sociale netwerken en big data

Bescherming van onze persoonsgegevens is een grondrecht. De leefwereld van jong en oud wordt vandaag gekenmerkt door een verregaande digitalisering: internet-der-dingen, informatievergaring over locaties, sociale media, het aanwenden van big data in het profileren van passagiers en consumenten… Welke gevaren voor onze privacy houden zich schuil in onze digitale leefwereld?

Dit Standpunt van de Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten richt zich vooral tot privégebruikers van alle generaties die zich, al dan niet terecht, zorgen maken over de gevaren waaraan hun privacy blootgesteld is in een wereld van big data. Lezers krijgen een dieper inzicht in de mogelijkheden en beperkingen van de technologie, maar ook in commerciële belangen en hun relatie tot de inperking van en de gevaren voor onze persoonlijke privacy. Een aantal gevallenstudies maakt de problematiek aanschouwelijk en concreet. Het Standpunt besluit met aanbevelingen voor ICT-verantwoordelijken, alerte burgers, bouwers van ICT- en Internet der Dingen-apparaten, overheden, bedrijven en het onderwijs.

De reeks Standpunten van de Academie is een bijdrage tot het wetenschappelijk onderbouwd debat over actuele maatschappelijke en artistieke thema's. De auteurs, leden en werkgroepen van de Academie schrijven in eigen naam, onafhankelijk en met volledige intellectuele vrijheid. De goedkeuring voor publicatie door één of meerdere Klassen van de Academie waarborgt de kwaliteit van de gepubliceerde studies

Vlaamse overheid

RICHTING MORGEN
MET WETENSCHAP, TECHNOLOGIE, CREATIVITEIT & INNOVATIE