



2 Naar een datagedreven samenleving: uitdagingen

Sunil Choenni, Mortaza S. Bargh en Niels Netten

Samenvatting

De hoeveelheid beschikbare data is groter dan ooit en blijft razendsnel groeien. Tegelijkertijd is de technologie vandaag de dag zo ver gevorderd dat computers krachtig genoeg zijn om deze gegevens snel te verwerken en dat biedt oneindig veel mogelijkheden. Al deze nieuwe technische mogelijkheden roepen tegelijkertijd ook vragen op. Deze vragen blijven niet alleen beperkt tot het gebied van datakwaliteit, privacy en ethiek. De vragen gaan ook over de rol van een datagedreven samenleving op eigentijdse vraagstukken zoals polarisatie. In deze bijdrage bespreken we hoofdzakelijk de uitdagingen van een datagedreven samenleving alsmede de impact die zo'n samenleving op haar beurt heeft op het onderwijs.

1. Inleiding

Steeds meer organisaties willen de mogelijkheden van (big) data benutten en kiezen daarom voor data- en informatiedreven werken. Kenmerkend voor deze manier van werken is dat grote hoeveelheden data uit verschillende bronnen worden verzameld en geanalyseerd. De uitkomsten van deze analyses vormen de bouwstenen voor data- en informatiegestuurde beslissingen. Deze beslissingen worden al dan niet door menselijke tussenkomst geïmplementeerd. Er zijn tal van voorbeelden te geven, denk aan de vele apps en de huidige datacentrische informatiesystemen, waarbij deze manier van werken als succesvol wordt beschouwd omdat zij effectief is gebleken en/of de efficiëntie is verbeterd. Onderzoek en ontwikkeling om data (verder) te benutten vindt inmiddels in vrijwel alle sectoren plaats (zie o.a. Cunha et al., 2010; Harbers, 2019; IMEC, 2019; Netten et al., 2016a; 2016b; 2017; 2018a; 2018b; 2019; Sarkis et al., 2004). De mogelijke toepassingen hiervan kunnen geclassificeerd worden in verschillende 'maturity levels', dat wil zeggen: toepassingen die direct ingebed kunnen worden in onze samenleving, toepassingen die in

de nabije toekomst mogelijk zijn of toepassingen die in de verre toekomst een kans van slagen hebben. Een voorbeeld van de eerste categorie is een systeem om ons huis (en tuin) te beveiligen door te simuleren dat we thuis zijn, terwijl we op vakantie zijn. Dit systeem kan uitgerust worden met allerlei sensoren, die bijhouden wat er in de omgeving gebeurt, bijvoorbeeld wat het geluidsniveau buiten is en hoe licht (of donker) het buiten is. Op basis van de input van de sensoren kan het systeem besluiten om lichten aan te zetten of muziek af te spelen. Een voorbeeld van een toepassing in de nabije toekomst is de zelfrijdende auto en de daarbij behorende infrastructuur. Voorbeelden van toepassingen die in de verre toekomst een kans maken om ingebed te worden in onze samenleving, zijn de wc-pot van de toekomst en 'experience sharing'. De wc-pot van de toekomst zal een soort van laboratorium worden in vergelijking met de wc-pot zoals we die nu kennen. De menselijke uitwerpselen worden geanalyseerd om al dan niet gezondheidsadviezen mee te geven aan de gebruiker, zoals: "Eet vandaag wat meer groente en drink meer water." Het doel van 'experience sharing' is om een gelijksoortig gevoel/ervaring met elkaar te delen, bijvoorbeeld met je partner. Als je je ergens aan stoot en als gevolg daarvan pijn voelt, dan kan je de pijnervaring delen met je partner.

Naast de vele mogelijkheden die het benutten van data ons bieden, brengt het ook vele uitdagingen met zich mee. Al te vaak blijkt dat we nog geen adequaat antwoord hebben op veel van deze uitdagingen. Een van de uitdagingen waar zowel de gemeente Rotterdam als de Belastingdienst tegenaan liep, was hoe om te gaan met profielen en kansen. Een profiel is een conclusie die geldt voor een groep van mensen met een bepaalde kanswaarde. De gemeente Rotterdam maakte gebruik van het Systeem Risico Indicatie (SyRi), dat de kans op bijstandsfraude binnen een huishouden bepaalde. In de wijken Bloemhof en Hillesluis werden data van bewoners over huizen, arbeid, onderwijs, detentie, uitkeringen, schulden, belastingen en ontvangen toeslagen gecombineerd om een risicoanalyse te maken van wie bijstandsfraude zou kunnen plegen. Bij de Belastingdienst werden, onder meer op grond van een dubbele nationaliteit en andere vormen van etnisch profileren, mensen opgespoord die mogelijk fraudeerden met hun kindertoeslag. In deze praktijkgevallen is het onduidelijk hoe de twee organisaties om zijn gegaan met het duiden van de resultaten van de systemen. In deze situaties is het namelijk van cruciaal belang het begrip kans en profiel adequaat te duiden. Immers, ook al is het resultaat van een systeem dat iemand een kans van 98% heeft om fraude te plegen of dat hij/zij volledig voldoet aan een fraudeprofiel, dan mag er vervolgens niet zomaar

geconcludeerd worden dat de fraude ook daadwerkelijk plaats gaat vinden. Een feit bij dergelijke toepassingen is dat van tevoren bekend is dat men er ook naast zal zitten met de analyses. De vraag is dan hoe hiermee omgegaan dient te worden, dat wil zeggen: hoe ga je om met de zogenoemde 'false positives' en 'false negatives' (Choenni et al., 2006; 2018; Franse, 2018). Doordat beide organisaties geen adequaat antwoord op de uitdagingen rondom de duiding van de resultaten van hun systemen hadden geformuleerd, heeft dit geleid tot benadeling, stigmatisering van groepen en maatschappelijke ophef. Als gevolg daarvan is de toepassing zowel bij de gemeente Rotterdam als bij de Belastingdienst stopgezet.

In deze bijdrage bespreken we een aantal uitdagingen – de kloof tussen 'haves' en 'have-nots', de kloof tussen systeemwerkelijkheid en realiteit, duiding van systeemuitkomsten en het fenomeen 'function creep' – waarop adequate antwoorden gevonden dienen te worden voordat we tot de implementatie van een 'full swing' datagedreven samenleving over kunnen gaan en de vruchten van zo'n samenleving kunnen plukken. Tevens zetten wij uiteen waarom we nog niet klaar zijn voor een datagedreven samenleving en dat het nu doordrukken van zo'n samenleving kan leiden tot polarisatie. Tot slot bespreken we hoe onderzoek en onderwijs kunnen bijdragen aan het formuleren van de antwoorden op de uitdagingen van een toekomstig datagedreven maatschappij.

Dit hoofdstuk is verder als volgt opgebouwd. Paragraaf 2 is gewijd aan een aantal belangrijke pijlers, vanuit een technische optiek, waarop een datagedreven samenleving gestoeld is. In paragraaf 3 bespreken we de uitdagingen die van antwoorden voorzien dienen te worden alvorens we de vruchten kunnen plukken van een datagedreven maatschappij. Paragraaf 4 is gewijd aan een discussie hoe het onderwijs kan en moet bijdragen aan de uitdagingen die gepaard gaan met een datagedreven samenleving. In paragraaf 5 schetsen we een viertal toekomstscenario's voor Hogeschool Rotterdam vanuit een datagedreven perspectief. Tot slot, in paragraaf 6 sluiten we dit hoofdstuk af met een aantal conclusies en aanbevelingen.

2. Naar een datagedreven samenleving

De hooggespannen verwachtingen van een datagedreven samenleving zijn te danken aan de vele mogelijke toepassingen die waarde toevoegen in een samenleving en aan ons dagelijks bestaan. Bij deze toepassingen, die veelal mogelijk zijn geworden door de inzet van technologieën ontwikkeld in diverse vakgebieden en de integratie in deze gebieden, speelt de (geavanceerde) verwerking van data een cruciale rol. De komst van internet en later die van de smartphones heeft de ontwikkeling van deze toepassingen een enorme boost gegeven. Zo werden in de luchtvaart, sinds de komst van radars, technieken ontwikkeld om de locatie van vliegtuigen te volgen ('tracking and tracing'). Door de komst van smartphones, uitgerust met sensoren waarmee onder andere onze locatie kan worden bepaald, kan het concept van 'tracken' ook op mensen worden toegepast. Deze sensordata hebben ervoor gezorgd dat we een nieuw type van vragen aan de applicaties op de smartphone kunnen stellen, dat voorheen voorbehouden was aan militaire en luchtvaarttoepassingen. In de bestrijding van het coronavirus is de vraag "Wie is de afgelopen dagen in mijn nabijheid geweest?" een actuele vraag, waar in principe met gegevens over locaties een antwoord op gegeven kan worden. Echter, voordat men een toepassing uit gaat rollen die een dergelijke vraag kan beantwoorden, is het raadzaam om in kaart te brengen wat de (maatschappelijke) gevolgen hiervan kunnen zijn, bijvoorbeeld voor de privacy van de betrokken personen. In het algemeen geldt dat succesvolle concepten ontwikkeld voor en in het technische domein bij transformatie naar het domein van maatschappelijke vraagstukken, de nodige uitdagingen met zich meebrengen. Bij het tracken van vliegtuigen hoefden we ons bijvoorbeeld niet druk te maken over de privacy van een vliegtuig. Zoals eerder opgemerkt, gaan we in paragraaf 3 uitvoerig in op de uitdagingen.

De factoren die in de toekomst een rol blijven spelen bij de implementatie van een datagedreven samenleving, zijn de beschikbaarheid van de hardware voor diverse devices, de connectiviteit tussen devices en de systeem- en applicatiesoftware op de devices. Laatstgenoemde type software is van cruciaal belang voor het ontwikkelen van nieuwe en eigentijdse toepassingen. Deze toepassingen zullen afhankelijk zijn van de voortgang die geboekt wordt op de gebieden van informatiesystemen, mens-machine-interactie en kunstmatige intelligentie. In het bijzonder zijn de deelgebieden conventionele databases, real-time gegevensverwerking, multimediasystemen en natuurlijke taalverwerking van belang. Het vakgebied van conventionele databases richt zich hoofdzakelijk op het efficiënt en effectief opslaan en bevragen van grote

hoeveelheden gestructureerde data (bijvoorbeeld administratieve data, zoals naam en adres) en semi-gestructureerde data (bijvoorbeeld documenten met een structuur, zoals een boek en artikel) (De Haan et al., 2011). Vragen waarvoor het noodzakelijk is om data direct te verwerken met het doel om een antwoord te kunnen formuleren op die vragen, is het terrein van real-time gegevensverwerking. Een voorbeeld van zo'n vraag kan zijn: "Waar bevindt het vliegtuig zich in het luchtruim op dit moment?" Multimediasystemen houden zich hoofdzakelijk bezig met het opslaan en verwerken van audio- en videodata. Ontwikkelingen op dit vakgebied maken het mogelijk (of zullen het mogelijk maken) om vragen te beantwoorden zoals: "Wat was het winnende doelpunt op het WK van 1978?"

Thans is het gebied van natuurlijke taalverwerking in opkomst en wij verwachten dat de ontwikkelingen op dit gebied een enorme boost kunnen geven aan een datagedreven samenleving. Het vakgebied van natuurlijke taalverwerking houdt zich bezig met het opslaan en analyseren van ongestructureerde data, veelal teksten in verschillende talen. De analyses op die teksten kunnen variëren van syntactische en semantische analyses tot het duiden van subjectiviteit, bijvoorbeeld het filteren op bias en/of discriminerende opmerkingen. Of een opmerking als discriminerend of als bias wordt beoordeeld, is afhankelijk van de tijdgeest en de taal waarin een opmerking is gemaakt. Verder zien we ook dat er onderzoek gaande is naar de overeenkomsten en verschillen tussen verschillende talen. Het zijn hoofdzakelijk de bekende talen, zoals het Engels, of de talen die door veel mensen gesproken worden, zoals Hindi en Chinees, die onderwerp van onderzoek zijn. Bevindingen over en inzichten in de overeenkomsten en verschillen tussen talen kunnen ertoe leiden dat ook de verwerking van 'kleine talen' een boost kan krijgen. Dit kan leiden tot toepassingen waarmee mensen in hun eigen taal met elkaar kunnen communiceren terwijl ze elkaar toch kunnen verstaan. Ook kan er een verschuiving van importantie plaatsvinden van geschreven naar gesproken taal. Hierdoor zouden mensen die de officiële taal in woord en/of geschrift niet machtig zijn, toch meer betrokken kunnen worden bij de samenleving.

In een stad als Rotterdam waar vele nationaliteiten naast elkaar wonen en niet iedereen het Nederlands in woord en/of geschrift voldoende beheerst, kunnen deze toepassingen een uitkomst bieden voor de communicatie tussen overheid en verschillende doelgroepen. Deze toepassingen zouden ingezet kunnen worden om elkaar beter te begrijpen en mogelijk polarisatie tegen te gaan.

3. Uitdagingen

In een samenleving waar data (als onderdeel van technologische toepassingen) een cruciale rol spelen in het faciliteren van het dagelijks leven en het oplossen van maatschappelijke opgaves, noemen we een datagedreven samenleving. Het (her)inrichten van zo'n maatschappij brengt de nodige uitdagingen met zich mee. In deze paragraaf bespreken we een aantal van deze uitdagingen. Binnen en buiten Hogeschool Rotterdam (HR) is er een groot bewustzijn van de uitdagingen die zich voordoen op het terrein van privacy, security en ethiek. Binnen HR is er sinds 1 december 2017 een lectoraat op het gebied van privacy en security en er is recentelijk een programma gelanceerd op het terrein van AI en ethiek. Het lectoraat Privacy & security timmert flink aan de weg (zie o.a. Bargh et al., 2016a; 2016b; 2019; Bargh, 2019; Choenni et al., 2016; Harbers, 2018). De uitdagingen die wij hier de revue laten passeren, zijn thans minder manifest maar zeker even belangrijk voor het welslagen van een datagedreven samenleving. In vier opeenvolgende subparagrafen bespreken we de kloof tussen de 'haves' en 'have-nots', de kloof tussen systeemwerkelijkheid en realiteit, de duiding van het kansbegrip en het gevaar van het fenomeen 'function creep' in een datagedreven samenleving.

3.1 'Haves' versus 'have-nots'

In de literatuur wordt de uitdrukking 'haves' en 'have-nots' gebruikt om de tegenstellingen tussen verschillende (inkomens)groepen duidelijk te maken met betrekking tot maatschappelijke mogelijkheden en werkgelegenheid. De 'haves' bezitten vrij veel tot alles terwijl de 'have-nots' weinig tot vrijwel niets bezitten (Wresch, 1996). De maatschappelijke mogelijkheden voor de laatstgenoemde groep is vele malen kleiner dan voor de eerstgenoemde groep. In deze bijdrage gaan we er tevens van uit dat de 'haves' digitaal geletterd zijn en begrijpen hoe eigentijdse systemen werken.

Er is altijd al een kloof geweest tussen de 'haves' en 'have-nots'. De indruk bestaat echter dat deze kloof groter is geworden door de automatisering van allerlei diensten en processen. Er wordt van burgers min of meer verwacht om mee te gaan in de ontwikkeling, ook al hebben ze geen baat bij de ontwikkeling. Zo zien we dat banken het bankieren middels internet goedkoper hebben gemaakt dan het conventionele bankieren. Bovendien ontmoedigen de banken laatstgenoemde vorm van bankieren. Echter, als men geen computer of internet tot zijn beschikking heeft ofwel niet kan lezen en schrijven, is het

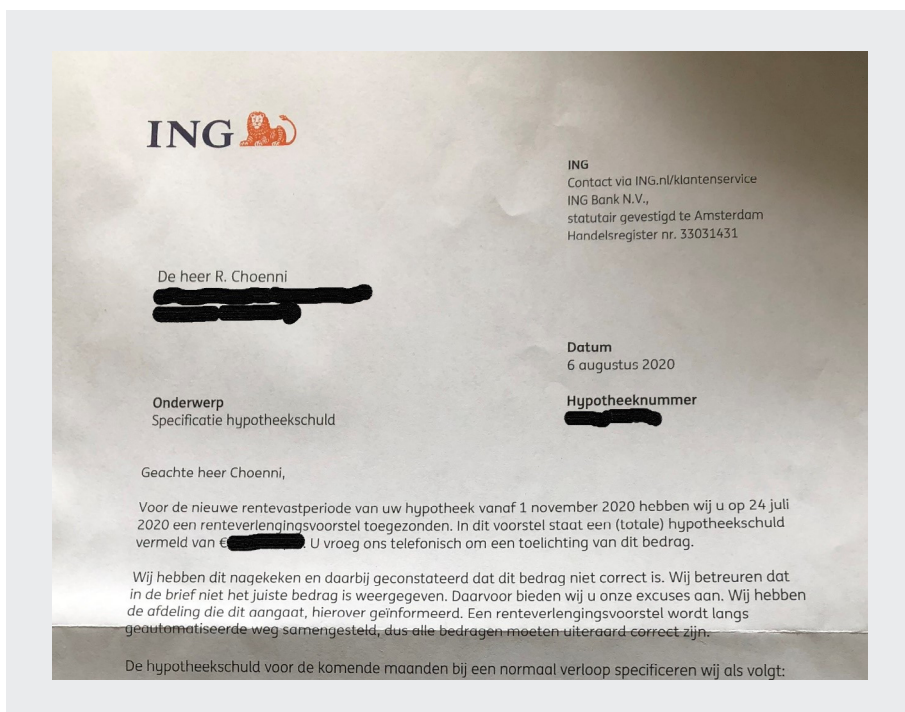
onmogelijk om de voordelen van het internetbankieren te benutten.

Een ander fenomeen dat zich subtiel voordoet, is dat door de automatisering veel professionals niet meer weten hoe de processen en de uitkomsten van geautomatiseerde systemen tot stand komen. Hierdoor hebben velen een blind vertrouwen in een uitkomst afkomstig van een geautomatiseerd systeem en verwachten deze professionals ook van anderen dat zij deze uitkomst als juist beschouwen. Deze professionals hebben moeite met het feit dat er getwijfeld wordt aan de uitkomst van een geautomatiseerd systeem. We lichten dit toe aan de hand van een recent praktijkvoorbeeld.

Van een van de auteurs liep de rentevaste periode van zijn hypotheek af. De bank heeft hem een nieuw rentevoorstel opgestuurd. Een nieuw rentevoorstel komt al heel lang automatisch tot stand. In het rentevoorstel dat de auteur ontvangen had, bleken de bedragen niet te kloppen. De auteur heeft dit in diverse telefoongesprekken aan de orde gebracht bij de bank. Elke keer probeerde de bankmedewerker met argumenten en verklaringen de auteur te overtuigen dat de bedragen correct waren. De auteur heeft iedere keer de argumenten en verklaringen tijdens de telefoongesprekken ontkracht. Desalniettemin heeft het hem de nodige overredingskracht gekost om de bank ervan te overtuigen om toch eens naar haar informatiesysteem te kijken. Dit heeft geresulteerd in een excuusbrief van de bank waarin de fout wordt toegegeven, zie figuur 1.

In deze casus heeft de bank in eerste instantie geprobeerd te rechtvaardigen dat het rentevoorstel met bijbehorende bedragen klopte. Dat de auteur absoluut zeker was van zijn berekeningen en over de nodige doorzettingsvermogen beschikte, heeft geleid tot de excuusbrief. Een 'have-not' zou het in deze casus waarschijnlijk al lang hebben opgegeven en derhalve zijn benadeeld. Overigens is het opvallend dat de bank in de brief opmerkt: "Een renteverlengingsvoorstel wordt langs geautomatiseerde weg samengesteld ..." De vragen die meteen bij ons opkomen zijn: is deze fout incidenteel of structureel en wat betekent dit voor de consument?

Indien bij de inrichting van een datagedreven samenleving niet adequaat wordt omgegaan met de kloof tussen de 'haves' en 'have-nots', wordt deze kloof alleen maar groter. We zullen aan de hand van een hypothetisch voorbeeld toelichten hoe subtiel dit kan gaan. Stel dat zowel de 'haves' als de 'have-nots' in het bezit zijn van fietsen om zich door het stadscentrum te



Figuur 1. Excuusbrief ING Bank

verplaatsen. Omdat diefstal van fietsen een probleem is in de stad, hebben velen van de 'haves' besloten een tracking device in hun fiets te monteren. Het tracking device houdt bij waar een fiets zich op een bepaald moment bevindt. Het opsporen van een fiets met een tracking device is vele malen eenvoudiger voor de politie dan een fiets zonder zo'n device. Omdat de 'haves' de middelen (en kennis) hebben om zo'n device te monteren in hun fiets en de 'have-nots' niet, is het aantrekkelijk voor de politie om de 'haves' te bevoordelen. Immers, het opsporen van gestolen fietsen met een device zorgt voor een hoger ophelderingspercentage. Het gevaar hiervan is dat de politie zich richt op het opsporen van gestolen fietsen van de 'haves', waardoor de 'have-nots' minder gebruik maken van de politie en haar middelen. Hierdoor zou het beeld kunnen ontstaan dat de politie hoofdzakelijk ten dienste staat van de 'haves'. Als de politie haar capaciteit (verhoudingsgewijs) eerlijk verdeelt tussen de twee groepen, bestaat de kans dat aangiftes van gestolen fietsen van de 'haves' blijven liggen. Dit zal tot verontwaardiging leiden bij de 'haves' omdat dit de vraag zal oproepen waarom de politie niets met de aangiftes doet terwijl ze weet waar de gestolen fietsen zich bevinden.

Bij de inrichting van een datagedreven samenleving moet een evenwichtige strategie gevonden worden om met dergelijke dilemma's om te gaan. Deze strategie moet breed gedragen worden, anders ligt polarisatie op de loer ten gevolge van een groter wordende kloof tussen 'haves' en 'have-nots'. Vooral nog lijkt het erop dat bijvoorbeeld het gebruik van socialmediaplatforms polarisatie in de hand werkt. Groepen gebruiken deze platforms om hun eigen denkbeelden te versterken en sluiten zich af van andere groepen, waardoor het vertrouwen tussen hen afneemt. Dit heeft als gevolg dat groepen uit elkaar groeien en zich harder tegen elkaar opstellen. Deze trend zal gekeerd moeten worden door de platforms in te zetten om mensen op een positieve manier met elkaar te verbinden.

3.2 Systeemwerkelijkheid versus realiteit

Een uitkomst die afgeleid kan worden uit een verzameling gegevens die zich in een of meerdere (informatie)systemen bevindt, noemen we een 'systeemwerkelijkheid': een nieuwe realiteit die tot stand is gekomen op basis van data uit informatiesystemen en daarmee slechts een representatie of interpretatie is van verbanden in de echte wereld (Choenni et al., 2018). Aangezien een systeemwerkelijkheid dikwijls gebaseerd is op onzekere en onvolledige data, komt deze niet altijd overeen met de echte werkelijkheid oftewel realiteit. Ter illustratie: de eerste alinea in de brief van figuur 1 doet voorkomen alsof er een verzoek is geweest om een toelichting, die zonder slag of stoot is gehonoreerd. In de realiteit zijn er diverse stevige telefoongesprekken geweest voordat het verzoek serieus werd genomen door de bank.

Als een systeemwerkelijkheid is gebaseerd op verouderde data (bijvoorbeeld uit legacysystemen), dan is deze vaak niet (meer) van toepassing op het 'hier en nu'. Dit omdat de gevonden uitkomst is gebaseerd op waarnemingen uit het verleden. Bij het interpreteren van analysesresultaten moet daarom goed in de gaten gehouden worden op welke periode de gevonden systeemwerkelijkheid betrekking heeft. Steeds moet beoordeeld worden of de verzamelde gegevens nog steeds representatief zijn voor het heden. In domeinen waarin de omstandigheden vaak en veel veranderen, is dit van zeer groot belang. Als de systeemwerkelijkheid te veel afwijkt van de realiteit, dan is de bruikbaarheid ervan in het geding. Het is dan lastig om op basis daarvan uitspraken te doen over de echte wereld.

Stel dat een database informatie bevat over alle klachten van klanten en de afhandeling ervan, in de afgelopen vijftig jaar. Deze database kan dan geanalyseerd worden om te kijken welk type klanten het meest succesvol is bij het indienen van klachten. Stel dat op basis van een datamining-algoritme gevonden wordt dat goed opgeleide mannen uit stedelijke gebieden de grootste kans hebben om succesvol een klacht in te dienen. Dit is dan de gevonden systeemwerkelijkheid. Hoewel dit misschien gold voor de situatie in het verleden, is het zeer goed mogelijk dat deze regel vandaag de dag niet meer opgaat. Een van de redenen daarvan kan zijn dat vrouwen dertig tot vijftig jaar geleden minder vaak hoger onderwijs genoten en (daardoor) ook minder vaak (zelf) klachten indienden. Vroeger werden vooral door hogeropgeleide mannen klachten ingediend en dat vertekent de resultaten. Tegenwoordig zijn vrouwen ook goed opgeleid en dienen ze ook vaker klachten in. Het gevonden dataminingresultaat is daarom niet noodzakelijkerwijs geldig en bruikbaar voor de wereld van vandaag. Voor het verleden is de bevinding misschien juist wel (nog steeds) bruikbaar.

Gegeven het feit dat een systeemwerkelijkheid hooguit een afbeelding is van de realiteit en nooit de gehele realiteit in al haar facetten kan vangen, zal er altijd een kloof zijn tussen een systeemwerkelijkheid en de realiteit. De kloof mag niet zo groot zijn dat we de realiteit niet meer kunnen herkennen in de systeemwerkelijkheid, met name in het geval dat we deze werkelijkheid gebruiken om de realiteit (aan) te sturen. Hoewel de uitkomsten uit een systeemwerkelijkheid kritisch bekeken dienen te worden, zien we tegenwoordig dat men geneigd is de systeemwerkelijkheid eerder te vertrouwen dan de realiteit. Het kritisch beoordelen van de uitkomsten van een systeemwerkelijkheid in het licht van de realiteit wordt voor veel professionals steeds lastiger. De reden hiervoor is dat de wijze waarop de uitkomsten uit een geautomatiseerd systeem verkregen zijn, voor professionals steeds meer als 'black-box' ervaren wordt. Er begint steeds meer draagvlak voor te ontstaan om deze trend te keren en de kwaliteit van de data waar de uitkomsten op gebaseerd zijn, beter in kaart te brengen.

3.3 Omgaan met kansen

Het begrip kans lijkt ogenschijnlijk goed gedefinieerd te zijn en velen van ons zijn ermee bekend, maar het begrip blijkt in de praktijk moeilijk hanteerbaar te zijn met name in de context van data science (Choenni et al., 2006; 2018; Franse, 2018). In deze subparagraaf beperken we ons tot de definities van

de twee meest gangbare vormen van kans: de frequentistische kans en de subjectieve kans.

De frequentistische kans is gedefinieerd als de quotiënt van het aantal gunstige uitkomsten en het aantal mogelijke uitkomsten. Aan deze definitie wordt de volgende interpretatie aan toegekend: bij een groot aantal waarnemingen/ experimenten convergeren de quotiënten naar een verdelingsfunctie van kansen met betrekking tot iedere mogelijke uitkomst. Bijvoorbeeld bij een (heel) groot aantal worpen met een zuivere dobbelsteen vinden we dat de kans op het gooien van elk van de ogen van de dobbelsteen 1 op 6 is (dus een uniforme verdelingsfunctie). Echter, in veel praktijksituaties zijn experimenten zoals deze niet uitvoerbaar. Stel dat we de kans willen bepalen dat een vliegtuig neerstort als een van de motoren stuk is, dan is het niet praktisch om heel veel vliegtuigen met een kapotte motor de lucht in te sturen en te turven hoeveel van de vliegtuigen neerstorten.

In de praktijk wordt daarom veel met het subjectieve kansbegrip gewerkt. Een nadeel van dit kansbegrip is dat er geen exacte definitie van bestaat. Vaak wordt een subjectieve kans ingeschat op basis van ervaring en kennis en deze kan opgevat worden als een gekwantificeerd oordeel, aangeduid als een 'a priori kans'. Als we eenmaal de a priori kansen kennen in een kansmodel dan kunnen we de rekenregels van de kansrekening toepassen.

Datagedreven toepassingen leiden vaak tot uitkomsten die gepaard gaan met een zekere mate van onzekerheid. Bijvoorbeeld, een uitkomst kan zijn: er bestaat een kans van 80% dat Piet fraudeert. De vraag die rijst, is: hoe moeten we de 80% interpreteren? Een mogelijke interpretatie in lijn met het frequentistische kansbegrip is dat we heel veel klonen van Piet produceren en die laten participeren in de samenleving, dan zal 80% van deze klonen fraude plegen. Maar zo'n experiment is niet uitvoerbaar. Een ander interpretatie is dat we alle mogelijke handelingen onderverdelen in fraude en niet-fraude. We laten Piet heel veel handelingen plegen, dan zullen we zien dat van alle handelingen die Piet verricht 80% een frauduleuze handeling is. Buiten het feit dat ook dit experiment niet uitvoerbaar is, zal Piet leren van zijn handelingen en voortaan verkeerde handelingen nalaten. Kortom, het frequentistische kansbegrip helpt ons niet bij het duiden van resultaten.

Als we de 80% als een subjectieve kans opvatten, dan kunnen we onderscheid maken tussen een entiteit die de kans genereert, bijvoorbeeld een algoritme,

en een subject waarop die kans betrekking heeft, bijvoorbeeld een persoon. De entiteit die de kans genereert, doet dit op basis van domeinkennis en ervaring, bijvoorbeeld: het algoritme heeft veel data over mensen zoals Piet verzameld en deze gecombineerd met data over het verleden van Piet. Over de persoon, in dit geval Piet, wordt alleen een oordeel gevormd, namelijk dat hij met een kans van 80% fraude zal plegen. Echter, het blijft onduidelijk of Piet daadwerkelijk fraude zal plegen of niet. Kortom, voor een entiteit die de kans genereert, is het subjectieve kansbegrip zinvol. Immers als de a priori kansen goed worden ingeschat, dan zal in het geval van heel veel voorspellingen de verdelingsfunctie van kansen convergeren naar de geschatte kansen, met ander woorden: als het algoritme voorspellingen doet bij heel veel mensen die op Piet lijken, dan zal 80% van die mensen fraude plegen en 20% niet. Voor de betreffende persoon is de kans echter betekenisloos. Immers voor de persoon geldt: hij heeft fraude gepleegd of niet, ongeacht wat de kans is. Als het subject geen fraude heeft gepleegd maar wel beticht wordt van fraude, dat wil zeggen: hij valt in de klasse van false positives, is het plausibel dat hij zich tegen de entiteit zal keren die de fraude voorspelt omdat hij mogelijk wordt benadeeld door de voorspelling. Derhalve zullen er strategieën ontwikkeld moeten worden die aangeven hoe om te gaan met false positives (en false negatives). Er zal bijvoorbeeld een procedure klaar moeten liggen met betrekking tot hoe en wat te communiceren naar mensen die ten onrechte in een bepaalde groep worden ingedeeld en hoe hen te compenseren als ze zijn benadeeld. Hiervoor kan gekeken worden naar domeinen waar het subjectieve kansbegrip succesvol wordt toegepast, zoals in het medische domein. Best practices kunnen dan ingebouwd worden in datagedreven systemen. Want één ding weten we zeker bij voorspellingen die gebaseerd zijn op kansen, namelijk dat er voorspellingen zijn die ook niet zullen uitkomen.

3.4 Function creep

Het fenomeen function creep kan beschouwd worden als een concept waarbij applicaties maar ook data geleidelijk aan anders gebruikt worden dan oorspronkelijk is bedoeld (Prins, 2011). Zo was Facebook ooit bedoeld om mensen met elkaar te verbinden, maar het is nu ook een tool voor organisaties om zoveel mogelijk data te verzamelen. Vaak begint het met het verzamelen van microdata voor het verlenen van diensten aan gebruikers. Microdata betreffen gegevens van een gebruiker, die veelal door de gebruiker zelf worden ingevoerd. Waar de gebruiker minder grip op heeft, zijn de bewerkingen die in de

loop van de tijd op de data uitgevoerd worden. Zo kunnen de data met andere data gecombineerd worden, vervolgens geaggregeerd worden en voor (direct)marketingdoeleinden worden gebruikt. Ook kunnen geaggregeerde data benut worden voor het formuleren van beleid. De microdata kunnen in combinatie met andere data ook voor surveillance-doeleinden dienen. We lichten dit toe aan de hand van een voorbeeld.

In veel Nederlandse straten zien we ondergrondse vuilcontainers. Deze containers hebben als voordeel dat er op elk tijdstip van de dag afval in gegooid kan worden en derhalve niet gewacht hoeft te worden tot vaste dagen waarop vuilnis wordt opgehaald. Steeds vaker zien we dat bewoners een scanpas krijgen om de container te openen. Zo'n scanpas levert data op die in de toekomst op verschillende manieren gebruikt kunnen worden. Er kan bijgehouden worden hoe vaak en op welke tijdstippen men afval stort in de containers. Niet ondenkbaar is dat er ooit beleid wordt ontwikkeld volgens het principe 'de vervuiler betaalt'. Mensen die vaker gebruik maken van de containers, zouden dan meer belasting kunnen gaan betalen. Voor marketeers zou het interessant zijn om te weten wat de vuilniszakken bevatten. Dit zou hen inzicht kunnen verschaffen in wat voor producten in een bepaalde wijk/postcodegebied populair zijn. Marketeers kunnen dan in hun reclamecampagnes hierop inspelen. In het licht van gescheiden afval, zou de lokale overheid op basis van de scangegevens gemakkelijker bewoners kunnen opsporen die hun afval niet op een juiste manier scheiden. Kortom, de vuilcontainer die in eerste instantie is bedoeld om het wonen aangenamer te maken voor de bewoners, kan eenvoudig worden omgetoverd tot een middel dat toezicht- en marketingdoeleinden faciliteert.

Function creep vindt veelal geleidelijk plaats en buiten het zicht van belanghebbenden. Dit heeft als gevolg dat het een poos duurt alvorens belanghebbenden in de gaten krijgen dat er feitelijk iets anders wordt geleverd of gedaan dan initieel afgesproken was en waarvoor zij de data leveren. Als function creep zich veelvuldig voor gaat doen in een datagedreven samenleving, dan kan het beeld ontstaan dat function creep een strategie is om belanghebbenden voor voldongen feiten te plaatsen, waarmee het draagvlak voor die samenleving zal afnemen. Derhalve zal er gezocht moeten worden naar manieren om met function creep om te gaan en scherp toezicht te houden op function creep. In het bijzonder, dient te worden voorkomen dat ten gevolge van function creep data voor opsporingsdoeleinden worden gebruikt. We zien nu al dat burgers moeite hebben met het in gebruik nemen van de app die moet helpen om de verspreiding van COVID-19 in te dammen.

Tot slot gaat function creep vaak gepaard met schending van de privacy van burgers en stigmatisering. Derhalve zijn dit ook aandachtspunten.

4. Onderwijs

De digitalisering van onze samenleving zal verder gaan toenemen en daarmee zal ook de datagedreven samenleving verder vorm krijgen. Dit heeft al zijn weerslag op en in het onderwijs en dat zal alleen maar toenemen. Om kennis op te doen is de student niet meer afhankelijk van een docent op een fysieke locatie. Tegenwoordig zijn er verschillende bronnen die studenten kunnen benutten om kennis op te doen, zoals online presentaties op YouTube, Wikipedia, kennis op verschillende socialmediawebsites en data portals. Ook hier kunnen we een 'dark side' herkennen, namelijk de kans op 'fake' data en kennis. De student kan er niet meer blindelings van uitgaan dat de data en kennis die worden gepresenteerd op de kanalen, voldoen aan de wetenschappelijke vereisten en of ze überhaupt een toegevoegde waarde hebben. Immers met een beetje inspanning kan iedereen data en kennis publiceren op deze eigentijdse kanalen. Derhalve moeten studenten in het onderwijs getraind worden in het herkennen van fake data en kennis, hoe ermee om te gaan en hoe data en kennis te vinden die aan de wetenschappelijke vereisten voldoen en daadwerkelijk een toegevoegde waarde hebben in het oplossen van (maatschappelijke) opgaven. Dit vraagt om het op een andere manier organiseren van onderwijs teneinde de vele kennis- en databronnen adequaat te benutten. Hiervoor dient de student, ongeacht studierichting, haar/zijn digitale vaardigheden uit te bouwen. COVID-19 heeft ervoor gezorgd dat hier een enorme aanzet toe is gegeven. In het onderwijs dient ook overwogen te worden om eigentijdse accenten in verschillende curricula aan te brengen. We bespreken een aantal van deze accenten die samenhangen met een datagedreven samenleving.

Er is altijd al een neiging geweest om succesvolle toepassingen, concepten en technologieën in het technische domein in te zetten om er maatschappelijke problemen mee op te lossen. Zoals betoogd in de voorgaande paragrafen, gaat dit vrijwel altijd gepaard met (nieuwe) additionele uitdagingen. Om te voorkomen dat men voor verrassingen komt te staan, is het aan te bevelen om de uitdagingen die deze toepassingen, concepten en technologieën met zich meebrengen, systematisch in kaart te brengen, nog voordat ze ingezet worden om maatschappelijke opgaven mee op te lossen. Studenten moeten getraind worden in het herkennen en het in kaart brengen van deze

uitdagingen. Hiervoor zullen in het onderwijs de juiste 'tools' aangereikt moeten worden en dienen de juiste vaardigheden aangeleerd te worden.

Er kan overwogen worden om het onderwijs meer vorm te geven langs de uitdagingen opgesomd in paragraaf 3. Per sector zou gekeken moeten worden in hoeverre een uitdaging bepalend is voor de toekomst van die sector. Afhankelijk hiervan zou de uitdaging al dan niet terug te moeten vinden zijn in de curricula van de opleidingen die betrekking hebben op die sector.

In het onderwijs moet men ervoor zorgen waken dat de beperkingen en uitkomsten van (computer)systemen niet klakkeloos worden overgenomen en gaan domineren. Vaak worden de beperkingen als een gegeven beschouwd, worden oplossingen bedacht die voldoen aan de beperkingen en vervolgens worden deze oplossingen voor waar aangenomen. Kijk naar een programmeertaal die alleen kan optellen en de uitkomsten afrondt op gehele getallen, waarbij kleiner dan 0,3 naar beneden wordt afgerond en anders naar boven. Stel dat we $(4 \times 1,6)$ willen uitrekenen, dan wordt dit herschreven tot $1,6 + 1,6 + 1,6 + 1,6$ en krijgen we als antwoord 7. In dit geval zouden studenten de vragen moeten kunnen opwerpen waarom er geen programmeertaal is waar een vermenigvuldiging-operator is ingebouwd en waarom 6,4 naar 7 wordt afgerond. Omdat de huidige generatie informatiesystemen uit veel componenten bestaat die met elkaar verbonden zijn, wordt het doorgronden van deze informatiesystemen steeds complexer, waardoor het steeds meer inspanningen vergt om ermee om te gaan en zich deze eigen te maken. Het gevolg is dat er steeds minder aandacht wordt besteed aan onderliggende concepten waarop een informatiesysteem gestoeld is. Deze concepten zijn van belang voor het begrijpen van de beperkingen en verklaren van de uitkomsten. Met name in het informatica-onderwijs moet ervoor gezorgd worden dat er voldoende tijd en ruimte is om op de eigentijdse informatiesystemen, hun beperkingen en hun uitkomsten te reflecteren. Echter, gelet op de ontwikkelingen in de samenleving is het raadzaam om ook in andere opleidingen aan de genoemde elementen in aangepaste vorm (lees: hoger abstractieniveau) aandacht te besteden. Tot op zekere hoogte moet men in alle opleidingen in staat zijn geautomatiseerde informatiesystemen te doorgronden, inzicht hebben in de interactie tussen mensen en deze systemen en de impact van deze systemen in een datagedreven samenleving kunnen overzien.

In paragraaf 2 hebben we betoogd dat een aantal gebieden binnen de informatica beschouwd kunnen worden als hoekstenen van een datagedreven

samenleving. Ontwikkelingen in deze (deel)gebieden waarvan voorzien kan worden dat ze tot toepassingen leiden die een toegevoegde waarde hebben in onze maatschappij, dienen ingebed te worden in het curriculum van de juiste studierichtingen. In paragraaf 2 hebben we gemotiveerd dat de ontwikkelingen op het gebied van natuurlijke taalverwerking tot zulke toepassingen kunnen leiden.

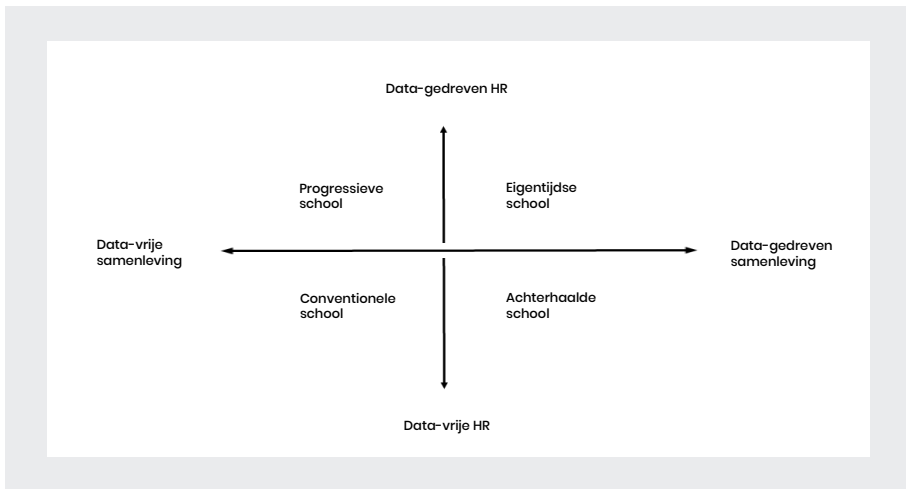
Ten slotte merken we op dat ook docenten hun kennis met betrekking tot een datagedreven samenleving up-to-date moeten houden. Daarom is een nauwe betrokkenheid van docenten bij de onderzoeksactiviteiten die betrekking hebben op het gebied van een datagedreven samenleving, onontbeerlijk. Zo ontvangen studenten state-of-the-art onderwijs en ontstaat er een versterkende cirkel tussen onderwijs, onderzoek en maatschappij.

5. Scenario's Hogeschool Rotterdam

In deze paragraaf schetsen we vier toekomstscenario's voor HR vanuit een datagedreven perspectief. Dit doen we aan de hand van twee dimensies. Deze dimensies betreffen de mate waarin datagedreven ontwikkelingen omarmd worden door de maatschappij respectievelijk HR. Indien er geen adequate antwoorden worden gevonden op de eerder beschreven uitdagingen die een datagedreven samenleving met zich meebrengt, leidt dat tot weerzin tegen zo'n maatschappij. De datagedreven maatschappij wordt verworpen en men streeft naar een maatschappij waar het gebruik van data bij besluitvormingen en apps nauwelijks een rol meer speelt, de zogenoemde datavrije samenleving. Het andere uiterste is dat de ontwikkelingen op het terrein van een datagedreven samenleving worden aangemoedigd. De ontwikkelingen beschreven in paragraaf 2 leiden tot nieuwe toepassingen en tot veel meer apps op onze (mobiele) devices.

In figuur 2 hebben we de dimensies tegen elkaar uitgezet en komen we tot vier kwadranten die elk een scenario representeren, de zogenoemde conventionele school, de achterhaalde school, de progressieve school en de eigentijdse school.

In het scenario van de conventionele school is er onvrede over de wijze waarop wordt omgegaan met de uitdagingen die het gebruik van data en internetfaciliteiten met zich meebrengt, bijvoorbeeld hoe fake kennis van echte kennis te onderscheiden. Men gaat terug naar het verleden toen het



Figuur 2. Toekomstscenario's HR

gebruik van data en internetfaciliteiten minimaal was. Er wordt klassikaal lesgegeven, de docent voor de klas is een autoriteit en er wordt niet getwijfeld aan zijn/haar kennis. Er wordt alleen op zoek gegaan naar data om van tevoren bedachte theorieën te verifiëren. Besluiten worden hoofdzakelijk genomen op basis van intuïtie en 'gut feeling'.

In het scenario van de achterhaalde school maakt de samenleving volop gebruik van de mogelijkheden van internetfaciliteiten en datagedreven toepassingen, in tegenstelling tot de school zelf. In de samenleving worden volop nieuwe apps gedownload, wordt gebruik gemaakt van kennis aanwezig op diverse kanalen/websites en open data, terwijl de school nog gelooft dat de docent de autoriteit is waar 'echte' kennis valt te halen. In dit scenario is er een grote discrepantie tussen de behoefte van de samenleving en wat de school te bieden heeft.

Bij de progressieve school worden niet alleen in het onderwijs de mogelijkheden van datagedreven ontwikkelingen en internetfaciliteiten volledig benut, maar ook de organisatie van de school maakt volop gebruik van deze mogelijkheden en faciliteiten. De organisatie combineert, integreert en analyseert data die betrekking hebben op studenten, zoals leefmilieu, medische indicaties, aanleg en cognitieve vaardigheden, en voorspelt op basis hiervan de studies waarin de student succesvol zal zijn en zichzelf zal overstijgen. De organisatie stelt een uitvoerbaar opleidingsplan op – met bijbehorende benodigdheden, zoals huisvesting, en dagindeling – voor de

student, om dit succes te bereiken. De student leert in de opleiding relevante kennis en data te halen uit verschillende bronnen, deze op waarde te schatten en toe te passen voor het oplossen van opgaven. De school is een hub in een breed netwerk van data- en kennisintensieve organisaties. De rol van de docent bestaat niet langer uit het doceren maar hij/zij is een respectabele 'peer' bij het oplossen van opgaven. Omdat de samenleving zich meer in de richting van een data-vrije samenleving bevindt, zal de progressieve school nauwelijks in verbinding staan met de maatschappij en in zich zelf gekeerd zijn. Hierdoor zal er een grote kloof zijn tussen school en maatschappij. Bij een eigentijdse school speelt dit fenomeen geen rol, omdat niet alleen de school de mogelijkheden van datagedreven toepassingen en internetfaciliteiten volledig benut, maar tegelijkertijd ook de samenleving. De eigentijdse school staat in verbinding met de samenleving, herkent de uitdagingen van de datagedreven samenleving en voorziet deze uitdagingen van adequate antwoorden. Naar de samenleving toe vervult de school de functie van het vroegtijdig signaleren van nieuwe ontwikkelingen.

Hoewel HR als organisatie nog niet volledig de mogelijkheden van de grote hoeveelheden beschikbare data benut, heeft zij de potentie om zich te ontwikkelen tot een eigentijdse school in een datagedreven samenleving. HR heeft aangetoond een adequate invulling te kunnen geven aan het vroeg signaleren van nieuwe ontwikkelingen. Enkele voorbeelden zijn de ontwikkelingen rondom open data, datagedreven profilering, privacy en cybersecurity. Zo is HR al in een heel vroeg stadium het project Rotterdam Open Data gestart. De resultaten van dit project zijn niet alleen in de wetenschappelijke wereld gecommuniceerd maar ook naar lokale overheden en de rijksoverheid. Inmiddels worden open data op verschillende ministeries en lokale overheden gestimuleerd. Op het moment dat veel lokale overheden en de rijksoverheid toepassingen van datagedreven profilering aan het verkennen waren, gaf HR al in een vroeg stadium het belang van transparantie en verantwoord gebruik van profielen aan in situaties met een grote impact op mens en maatschappij. Resultaten verkregen uit een datagedreven profileringsproject uitgevoerd bij de gemeente Rotterdam, zijn wetenschappelijk gecommuniceerd en worden in vergelijkbare profileringsprojecten toegepast. Er lopen ook nieuwe ontwikkelingen binnen HR om het model van transparantie bij datagedreven profilering nog verder te verbeteren. HR heeft als een van de eersten geopperd om privacy-engineeringstechnieken te combineren met het 'privacy by design'-paradigma. Inmiddels lopen er bij het ministerie van Justitie en Veiligheid diverse projecten om dit te realiseren.

Tot slot bood HR als eerste de minoren Data Science en Cybersecurity aan. HR signaleerde vrij vroeg de relevantie van deze onderwerpen voor de samenleving.

6. Conclusies

In dit hoofdstuk hebben we betoogd dat we steeds meer naar een datagedreven samenleving bewegen, gezien de vele mogelijkheden die dit met zich meebrengt. We bespraken ook dat een datagedreven samenleving de nodige uitdagingen met zich meebrengt. Indien er geen dequaat antwoord wordt gevonden op de uitdagingen, kan dit vergaande consequenties hebben voor een maatschappij, zoals verlies van draagvlak voor een datagedreven samenleving, verdere polarisatie tussen de 'haves' en 'have-nots' en een systeemwerkelijkheid die vervreemd is van de realiteit. Derhalve pleiten we voor meer bewustzijn voor deze uitdagingen, om te beginnen binnen het onderwijs. We hebben een eerste voorstel gedaan waar het onderwijs zich op zou kunnen richten, teneinde in de toekomst een eigentijdse school te blijven.

Literatuur

- Bargh, M.S., Meijer, R., Vink, M., van den Braak, S., Schirm, W., & Choenni, S. (2019, June). Opening Privacy Sensitive Microdata Sets in Light of GDPR. In *Proceedings of the 20th Annual International Conference on Digital Government Research* (pp. 314-323).
- Bargh, M.S. (2019). Realizing Secure and Privacy-Protecting Information Systems: Bridging the Gaps. Inauguration Lecture at Rotterdam University of Applied Sciences (RUAS), The Netherlands, ISBN: 9789493012080, RUAS Press.
- Bargh, M. S., Choenni, S., & Meijer, R. (2016a). On design and deployment of two privacy-preserving procedures for judicial-data dissemination. *Government Information Quarterly*, 33(3), 481-493.
- Bargh, M. S., Choenni, S., & Meijer, R. (2016b). On design and deployment of two privacy-preserving procedures for judicial-data dissemination. *Government Information Quarterly*, 33(3), 481-493.
- Choenni, S., Netten, N., Shoaie-Bargh, M., & Choenni, R. (2018, December). On the Usability of Big (Social) Data. In *2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom)* (pp. 1167-1174). IEEE.
- Choenni, S., Bargh, M. S., Roepan, C., & Meijer, R. F. (2016). Privacy and security in smart data collection by citizens. In *Smarter as the New Urban Agenda* (pp. 349-366). Springer, Cham.

- Choenni, S., Blok, H. E., & Leertouwer, E. (2006, April). Handling uncertainty and ignorance in databases: *A rule to combine dependent data*. In *International Conference on Database Systems for Advanced Applications* (pp. 310-324). Springer, Berlin, Heidelberg.
- Cunha, M. M. C., Tavares, A. J., & Simoes, R. (2010). *Handbook of research on developments in e-health and telemedicine: technological and social perspectives*. Information Science Reference.
- De Haan, G., Choenni, S., Mulder, I., Kalidien, S., & van Waart, P. (2011). Bringing the research lab into everyday life: exploiting sensitive environments to acquire data for social research. *The Handbook of Emergent Technologies in Social Research*, 522-541.
- Franse, A. (2018) "Je moet weten wat de foutmarge in een model is (You need to know the margin of error in a model)." Big Data JenV, Ministerie van Justitie en Veiligheid, Minisymposium Datakwaliteit, Vol. 1, Nr. 2, pp 5-6.
- Harbers, M., Bargh, M., Pool, R., Van Berkel, J., Van den Braak, S., & Choenni, S. (2018, January). A conceptual framework for addressing IoT threats: challenges in meeting challenges. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.
- Harbers, M., Bargh, M. S., Cramer, F., Choenni, S., Nijkamp, J., & Nigten, A. (2019, June). Crafting Privacy: Two Case Studies Integrating Cross-Disciplinary Perspectives on Privacy in Design. In *EICS Workshops* (pp. 88-94).
- IMEC, (2019). <https://www.imec-int.com/en/the-netherlands/smart-toilet>
- Netten, N., Bargh, M. S., & Choenni, S. (2018a, April). Exploiting data analytics for social services: on searching for profiles of unlawful use of social benefits. In *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance* (pp. 550-559).
- Netten, N., Bargh, M. S., Choenni, S., & Meijer, R. (2017, March). Exploiting Big Data for Evaluation Studies. In *Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance* (pp. 228-231).
- Netten, N., Bargh, M. S., van den Braak, S., Choenni, S., & Leeuw, F. (2016a, June). On enabling smart government: A legal logistics framework for future criminal justice systems. In *Proceedings of the 17th International Digital Government Research Conference on Digital Government Research* (pp. 293-302).
- Netten, N., Choenni, S., & Bargh, M. S. (2019, September). The Praxis of HR Analytics. In *Conference on e-Business, e-Services and e-Society* (pp. 190-202). Springer, Cham.
- Netten, N., van den Braak, S. W., Bargh, M. S., Choenni, S., & Leeuw, F. L. (2018b). *Legal logistics: a Framework to unify data centric services for smart and open justice*. *International Journal of E-Planning Research (IJEPR)*, 7(2), 51-69.

- Netten, N., van den Braak, S., Choenni, S., & van Someren, M. (2016b, March). A Big Data Approach to Support Information Distribution in Crisis Response. *In Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance* (pp. 266-275).
- Prins, J. E. J. (2011). Function creep: over het wegen van risico's en kansen. *Justitiële Verkenningen*, 37(8), 9-21.
- Sarkis, J., Meade, L. M., & Talluri, S. (2004). E-logistics and the natural environment. *Supply Chain Management: An International Journal*.
- Wresch, W. (1996). *Disconnected: Haves and have-nots in the information age*. Rutgers University Press.

Auteurs

Sunil Choenni

Lector Future Information and Communication
Technology, Kenniscentrum Creating010



<https://www.hogeschoolrotterdam.nl/onderzoek/lectoren/creating-010/lectoren/dr.-ir.-sunil-choenni/>

Sunil is afgestudeerd in theoretische informatica aan de TU Delft en gepromoveerd in databasetechnologie aan de Universiteit Twente. Na zijn promotie is hij werkzaam geweest bij verschillende onderzoeksinstituten en universiteiten op het gebied van informatica en toepassing van informatica. Sinds 2008 is hij verbonden als lector aan de HR. Tevens is hij hoofd van de afdeling Statistische Informatievoorziening en Beleidsanalyse (SIBa) bij het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) binnen het ministerie van Justitie en Veiligheid. Hij publiceert regelmatig (met leden van zijn onderzoeksgroep) op het gebied van forecasting, e-government, cybersecurity en informatiemanagement en is hij lid van tal van programmacommissies en editorial boards.

Recente publicaties

2018-2019

Bargh, M.S., and Choenni, S. (2019). Towards applying design-thinking for designing privacy protecting information systems. *In Proc. IEEE Int. Conf. on Trust, Privacy, and Security in Intelligent Systems and Applications (IEEE TPS)*, December 12-14, Los Angeles, USA, IEEE Press.

Bargh, M.S., Choenni, S., Meijer, R., and Choenni, S. (2019). A method for assessing the openness of semi-open data initiatives: Applied to the Justice Domain. *In Proc. of the 8th Int. Conf. on e-Democracy*, Athens, Greece, December 12-13, 2019, CCIS 1111, Springer, Germany, pp.110-125.

Bargh, M.S., Vink, M., and Choenni, S. (2018). On using obligations for usage control in joining of datasets. *In Proc. of the 3rd Int. Conf. Information Systems Security and Privacy (ICISSP'17)*, Porto, Portugal, February 19-21, 2017, Revised Selected Papers, P. Mori, S. Furnell, O. Camp (eds.), Springer, pp. 173-196.

Choenni, S., Netten, N., Bargh, M.S., and Choenni, R. (2018). On the usability of big (social) data. *In Proc. of IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications*, Melbourne,

Australia, December 11-13, IEEE Press, USA, pp. 1167-1174.

Netten, N., Choenni, S., and Bargh, M.S. (2019). The praxis of HR analytics. *In Proc. of the 18th IFIP Conf. on e-Business, e-Services and e-Society (I3E'19)*, Trondheim, Norway, September 18-20, LNCS 11701, Springer-Verlag, Germany, pp. 190-202.

Netten, N., Bargh, M.S., and Choenni, S. (2018). Exploiting data analytics for social services: On searching for profiles of unlawful use of social benefits. *In Proc. of the 11th Int. Conf. on Theory and Practice of Electronic Government (ICEGOV)*, Galway, Ireland, April 4-6, ACM Press, USA, pp. 550-559.

Dr. Niels Netten

Docent-onderzoeker,
Kenniscentrum Creating010



Niels is sinds september 2016 actief binnen het Kenniscentrum Creating010. Hij is actief binnen het project Data Driven Society. Daarbinnen voert Niels onderzoek uit naar transparantie van algoritmen, profiling en (big) data analytics. Hij begeleidt daarnaast ook promovendi.

<https://www.hogeschoolrotterdam.nl/onderzoek/lectoren/creating-010/medewerkers/netten-niels/>

Niels is afgestudeerd in informatica aan de Universiteit Leiden (2004) en gepromoveerd in data mining / machine learning aan de Universiteit van Amsterdam (2015). Tijdens zijn promotie heeft Niels ook gewerkt in het bedrijfsleven als business- en software engineer. Naast zijn huidige werkzaamheden als onderzoeker bij Creating 010, werkt hij als wetenschappelijk medewerker van de afdeling Statistische Informatievoorziening en Beleidsanalyse (SIBa) bij het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) binnen het ministerie van Justitie en Veiligheid. Hij publiceert en geeft regelmatig presentaties op het gebied van e-government en (big) data analytics, en hij is lid van de programmacommissie van ICEGOV. Hij is sinds juli 2018 ook vennoot in een bedrijf dat trainingen geeft op het gebied van (big) data analytics.

Recente publicaties

2018-2019

Bargh, M.S., Choenni, S., and Netten, N. (2019). Open justice in the Netherlands: An overview. A book chapter in *Open Justice: An Innovation-Driven Agenda for Inclusive Societies*,

- coordinator: S. Elena, Ciudad Autónoma de Buenos Aires: Ediciones SAIJ, pp. 259-271, may 2019. Available: <http://www.bibliotecadigital.gob.ar/items/show/2569>.
- Choenni, S., Netten, N., Bargh, M.S., and van den Braak, S. (to appear). Exploiting big data for smart government: Facing the challenges. A chapter in the *Handbook on Smart Cities*, ed. J. Carlos Augusto, Springer.
- Choenni, S., Bargh, M.S., Netten, N., and van den Braak, S. (2019). Using data analytics results in practice: Challenges and solution directions. A chapter of the IOS book *Digital Social Innovation for the European Welfare System: A Multi-Disciplinary Reflection and Future Perspectives from Internet Science, Human-Computer Interaction and Socio-Economics*, eds. F. Davide, G. Misuraca and G. Misuraca, IOS PRESS.
- Netten, N., Choenni, S., and Bargh, M.S. (2019). The praxis of HR analytics. In *Proc. of Conference on e-Business, e-Services and e-Society*, pp. 190-202. Springer, Cham, (In: Pappas I., Mikalef P., Dwivedi Y., Jaccheri L., Krogstie J., Mäntymäki M. (eds) *Digital Transformation for a Sustainable Society in the 21st Century*. I3E. Lecture Notes in Computer Science, vol 11701. Springer, Cham).
- Vink, M., Netten, N., Choenni, S., Bargh, M.S., and van den Braak, S. (2020). Mapping crime descriptions to law articles using deep learning. In *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance (ICEGOV)*, 22-25 September, Online, (Best Paper award).
- Weltevreden, J., Effing, R., Ectors, I., Raessens, B., Bruins, R., Brengman, M., Trampe, D., Veenstra, M., Vos, I., Willems, K., van de Sanden, S., Netten, C.P.M., and Steeman, M. (2020). Data-driven retail: Welke bijdrage aan de kennisontwikkeling binnen dit thema willen we leveren aan de retailsector? *Retail Innovation Platform*, Position paper, maart.

Dr.ir. Mortaza S. Bargh (Mortaza Shoaie Bargh)

Lector Privacy and Cybersecurity,
Kenniscentrum Creating010

<https://www.hogeschoolrotterdam.nl/onderzoek/lectoren/creating-010/lectoren/dr.-ir.-Shoaie-Bargh-Mortaza/>



Mortaza is afgestudeerd in telecommunicatie- en informatietheorie (1995) en gepromoveerd in informatietheorie (1999), beide aan de TU Eindhoven. Na zijn promotie heeft hij gewerkt bij het voormalige Novay of Telematica Instituut – in Enschede (van 1999 – 2011). Hij deed hier toegepast onderzoek op het gebied van veilige ICT- en internettoepassingen. Tussen 2012 en 2015 was hij reeds gastlector Cybersecurity en Privacy bij Kenniscentrum Creating 010 van Hogeschool Rotterdam. Sinds 2013 is Mortaza

als wetenschappelijke onderzoeker verbonden aan het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) van het Ministerie van Justitie en Veiligheid. Mortaza is sinds 1 december 2017 actief binnen het Kenniscentrum Creating010 als Lector Privacy en Cybersecurity. Zijn interessegebieden zijn: Algoritmen (van machine learning, data mining en signal processing); modelleren van onzekerheid; privacy engineering, security en privacy by design; gedistribueerde systemen; risicomangement (toegepast in domeinen cybersecurity en/of smart/e-government).

Recente publicaties

2019-2020

- Amighi, A., Bargh, M.S., Choenni, S., Latenko, A., and Meijer, R. (2020). On protecting microdata in open data settings from a data utility perspective. *In Proceedings of the Fourteenth International Conference on Digital Society (ICDS), special track Protecting Privacy in Open (& big) Data Settings (PPODS)*, Valencia, Spain, November 21-25, (Top Paper award).
- Bargh, M.S., Meijer, R., Vink, M., Schirm, W., van den Braak, S., and Choenni, S. (2019). Opening privacy sensitive microdata sets in light of GDPR: The case of opening criminal justice domain microdata. *In Proceedings of the 20th Annual International Conference on Digital Government Research (dg.o)*, June 18-20, Dubai, UAE.
- Bargh, M.S., and Choenni, S. (2019). Towards applying design-thinking for designing privacy-protecting information systems. *In Proceedings of the 1st IEEE International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications (IEEE TPS)*, Los Angeles, California, USA (Co-located with IEEE CIC 2019 & IEEE CogMI 2019), December 12 - 14.
- Bargh, M.S., Choenni, S., Meijer, R., and Choenni, A. (2019). A method for assessing the openness of semi-open data initiatives: Applied to the justice domain. *In Proceedings of the 8th International Conference on e-Democracy (e-Democracy)*, 12-13 December, Athens, Greece.
- Choenni, S., Netten, N., Bargh, M.S., and van den Braak, S. (to appear). Exploiting big data for smart government: Facing the challenges. A chapter in the *Handbook on Smart Cities*, ed. J. Carlos Augusto, Springer.
- Vink, M., Netten, N., Choenni, S., Bargh, M.S., and van den Braak, S. (2020). Mapping crime descriptions to law articles using deep learning. *In Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance (ICEGOV)*, 22-25 September, Online, (Best Paper award).

Dit artikel is onderdeel van de bundel:

Gijsbertse, D. P., Van Klink, H. A., Machielse, C., & Timmermans, J. H. (Red.). (2020). *Hoger beroepsonderwijs in 2030: Toekomstverkenningen en scenario's vanuit Hogeschool Rotterdam*. Hogeschool Rotterdam Uitgeverij.

De volledige bundel is te vinden op: <https://hr.nl/hbo2030>