Juridische aspecten van algoritmen die besluiten nemen

Een verkennend onderzoek

Met casestudy's naar contentmoderatie door online platformen, zelfrijdende auto's, de rechtspraak en overheidsincasso bij verkeersboetes

Auteurs

mr. dr. Stefan Kulk & mr. Stijn van Deursen

Met casestudy's van

mr. dr. Stefan Kulk & Thom Snijders LL.B

mr. dr. Vicky Breemen & mr. Anouk Wouters

mr. Stijn van Deursen & mr. dr. Stefan Philipsen

mr. dr. Martje Boekema & mr. dr. Susanne Heeger

Onder begeleiding van

prof. mr. Janneke Gerards

prof. mr. Eddy Bauw

prof. mr. drs. Madeleine de Cock Buning

prof. dr. mr. Henry Prakken

mr. dr. Nelleke Koffeman

prof. mr. Anna Gerbrandy

Montaigne Centrum voor Rechtsstaat en Rechtspleging, Universiteit Utrecht, 2 juni 2020

© 2020 WODC. Auteursrechten voorbehouden.

Bij voorkeur citeren als:
S. Kulk & S. van Deursen*, Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*,
Den Haag: WODC 2020.

De casestudy's kunnen apart geciteerd worden, bijvoorbeeld:
V.E. Breemen & A.H.H. Wouters, 'Casestudy Zelfrijdende auto's', in: S. Kulk & S. van Deursen*, Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, Den Haag: WODC 2020.

Voorwoord

Het gebruik van algoritmen bij (voorbereiding van) besluitvorming neemt de laatste jaren een grote vlucht. Dat brengt voor publieke waarden en belangen zowel kansen als risico's met zich mee. Het onderwerp mag zich dan ook terecht verheugen in een ruime belangstelling van politiek, wetenschap en samenleving. In opdracht van het WODC en op verzoek van het Ministerie van Justitie en Veiligheid heeft het Montaigne Centrum voor Rechtsstaat en Rechtspleging van de Universiteit Utrecht een verkennend onderzoek gedaan naar de juridische aspecten van algoritmen die besluiten nemen, waarin de publieke waarden en belangen centraal staan. Is het huidige juridisch kader toereikend om de kansen die de inzet van algoritmen om deze waarden te bevorderen te benutten en de risico's voor deze waarden binnen aanvaardbare grenzen te houden?

In het kader van dit onderzoek, in het bijzonder ten behoeve van de casestudy's, is gesproken met een groot aantal personen uit de wetenschap en 'het veld'. De onderzoekers danken alle betrokkenen voor hun bereidheid om aan dit onderzoek mee te werken. De onderzoekers danken voorts de leden van de begeleidingscommissie en het WODC voor de waardevolle begeleiding in de verschillende fasen van het onderzoek en de leden van de klankbordgroep voor het meelezen. Het eindproduct is daar zonder twijfel beter door geworden.

Dit rapport is het product van samenwerking van een omvangrijke groep van onderzoekers. Daarbij heeft een werkverdeling gegolden. Stefan Kulk en Stijn van Deursen zijn de auteurs van de hoofdtekst, terwijl de casestudy's zijn opgesteld door Stefan Kulk en Thom Snijders (contentmoderatie door online platformen), Vicky Breemen en Anouk Wouters (zelfrijdende auto's), Stijn van Deursen en Stefan Philipsen (de rechtspraak), Martje Boekema en Susanne Heeger (overheidsincasso bij verkeersboetes). De begeleiding van het onderzoek was in handen van Janneke Gerards, Eddy Bauw, Madeleine de Cock Buning, Henry Prakken, Nelleke Koffeman en Anna Gerbrandy. Nelleke Koffeman en Anouk Wouters hebben het onderzoek gecoördineerd. Voor het eindresultaat nemen de onderzoekers gezamenlijk de volle verantwoordelijkheid.

Een bijzondere vermelding verdient Anoeska Buijze voor haar inzet en bijdrage aan het onderzoek. Tot slot gaat onze dank uit naar de student-assistenten Claire Stalenhoef, Florianne Peters van Neijenhof, Jesse Balster en Jonas Folkers voor hun diverse en steeds belangrijke werkzaamheden.

Utrecht, 2 juni 2020
De onderzoekers

Samenvatting

1. Aanleiding en onderzoeksvraag

Iedereen heeft in het dagelijks leven te maken met beslissingen die worden genomen door of met behulp van algoritmen. De inzet van algoritmen kan kansen opleveren voor het verwezenlijken van publieke waarden en belangen. Zo kunnen algoritmen besluitvormingsprocessen efficiënter maken en bijdragen aan het vinden van oplossingen voor verschillende soorten maatschappelijke uitdagingen. Tegelijkertijd kan de inzet van algoritmen risico's met zich brengen en vragen oproepen over de bestendigheid van de juridische kaders die beschikbaar zijn om publieke waarden en belangen te beschermen. De onderzoeksvraag van dit onderzoek is in dit verband als volgt gedefinieerd:

Welke kansen en risico's doen zich voor bij algoritmische besluitvorming met betrekking tot de bescherming en realisering van publieke waarden en belangen, en zijn de bestaande juridische kaders voldoende bestendig om kansen te verwezenlijken en het intreden van geïdentificeerde risico's te voorkomen of de gevolgen daarvan te mitigeren?

Centraal in het onderzoek staan de huidige toepassingen van algoritmen in besluitvormings-processen en de ontwikkelingen die in de komende vijf tot tien jaar op dat gebied te verwachten zijn. Voor de beantwoording van de onderzoeksvraag is onder meer gebruikgemaakt van casestudy's naar de inzet van algoritmen in vier, door het WODC en de Directie Wetgeving en Juridische Zaken van het Ministerie van Justitie en Veiligheid geselecteerde domeinen: contentmoderatie, zelfrijdende auto's, rechtspraak en overheidsincasso bij verkeersboetes.

2. Algoritmische besluitvorming en publieke belangen

Algoritmische besluitvorming in de zin van dit onderzoek omvat alle processen waarin een algoritme wordt ingezet om beslissingen te nemen die raken aan de rechtspositie van rechtssubjecten of die hen anderszins in hun belangen treffen. Het kan daarbij gaan om gevallen waarin een algoritme zelf een besluit neemt, of gevallen waarin de uitvoer van een algoritme wordt meegenomen in een menselijk besluitvormingsproces. Vormen van algoritmische besluitvorming kunnen verschillen afhankelijk van het doel waarmee algoritmen worden ingezet. Ook kan er op technisch vlak een onderscheid worden gemaakt tussen regelgebaseerde algoritmen en zelflerende algoritmen. Tot slot is ook de organisatorische en maatschappelijke context waarin algoritmen worden ingezet van belang.

Zowel voor het identificeren van kansen en risico's, als voor de beoordeling van de bestendigheid van de juridisch kaders, fungeren publieke waarden en belangen als normatief kader. De publieke waarden en belangen zijn vanwege de juridische invalshoek van het onderzoek geconcretiseerd

aan de hand van drie grondrechten die vrijwel steeds aan de orde zijn bij de verschillende toepassingen van algoritmen: het recht op gegevensbescherming, het recht op non-discriminatie en het recht op rechtsbescherming. In de verschillende casestudy's spelen daarnaast meer specifieke waarden als duurzaamheid of de vrijheid van meningsuiting een rol.

3. Kansen en risico's van algoritmische besluitvorming

Kansen zijn in dit onderzoek gedefinieerd als de mogelijkheid om publieke waarden en belangen te verwezenlijken, terwijl het bij risico's gaat om de mogelijkheid dat publieke waarden en belangen niet verwezenlijkt of zelfs geschaad worden. Bij de bestendigheid van het juridisch kader gaat het om de vraag in hoeverre het juridisch kader het mogelijk maakt dat kansen worden verwezenlijkt en dat het intreden van risico's wordt vermeden of gemitigeerd.

Kansen en risico's voor publieke waarden en belangen hangen samen met het type algoritme dat wordt gebruikt, maar voor een belangrijk deel ook met het domein en de organisatorische en maatschappelijke context waarin algoritmen worden ingezet. Uiteraard geldt dat bij de inzet van een algoritme tegelijkertijd kansen en risico's kunnen bestaan. De inzet van algoritmen in een besluitvormingsproces vergt dan ook altijd een afweging van de behoeften in het betreffende domein en de waarden en belangen die daarin gelden. De juridische kaders geven de grenzen aan waarbinnen dergelijke afwegingen dienen plaats te vinden.

3.1 Kansen van algoritmische besluitvorming

De inzet van algoritmen kan efficiëntiewinst opleveren doordat zij in staat kunnen zijn om (besluitvormings)processen sneller, beter of nauwkeuriger te doorlopen. Zelflerende algoritmen in het bijzonder kunnen verbanden ontdekken in grote hoeveelheden gegevens. De efficiëntiewinst die op die manier geboekt kan worden, heeft veelal (bedrijfs)economische waarde, maar kan ook een belangrijke bijdrage leveren aan het verwezenlijken van publieke waarden.

Meer in het bijzonder creëert de inzet van algoritmen concrete kansen ten aanzien van de waarden van rechtsbescherming en non-discriminatie**. De kansen voor rechtsbescherming** hangen met name samen met de efficiëntiewinst die door de inzet van algoritmen geboekt kan worden, zoals vooral blijkt uit de casestudy naar de rechtspraak.

De inzet van algoritmen creëert ook **kansen voor het recht op non-discriminatie.** Doordat algoritmen in staat zijn om veel informatie te verwerken kunnen ze veel individuele kenmerken van personen meenemen in besluitvormingsprocessen. Algoritmen kunnen daarnaast helpen om besluiten beter af te stemmen op betrokken personen, waardoor de inzet van algoritmen kan bijdragen aan het realiseren van materiële gelijkheid. Ook kunnen algoritmen bijdragen aan gelijkheid in de vorm van consistentie van besluitvorming. Bovendien zijn goed-geprogrammeerde

en gevalideerde algoritmen in beginsel beter dan mensen in staat om zonder aanziens des persoons een besluit te nemen. Tot slot kunnen algoritmen juist ook worden ingezet om discriminatie in besluitvormingsprocessen te detecteren.

Ten aanzien van de **bescherming van persoonsgegevens** zijn, in het kader van dit onderzoek, **geen mogelijke kansen** vastgesteld.

3.2 Risico's van algoritmische besluitvorming

Risico's ten gevolge van de inzet van algoritmische besluitvorming kunnen bestaan voor alle drie de algemene publieke waarden en belangen. De **bescherming van persoonsgegevens** komt in het gedrang als algoritmen worden ingezet om op grote schaal persoonsgegevens te verzamelen of anderszins te verwerken. Daarnaast maakt de inzet van algoritmen het mogelijk om in bestaande informatie verbanden aan te brengen en zo (nog) meer te weten te komen over personen. Daardoor verliezen individuen niet alleen controle op hun persoonsgegevens, maar vervagen ook de grenzen tussen wat persoonsgegevens zijn en wat niet. Als algoritmen verbanden leggen die niet kloppen, kan dat bovendien raken aan de identiteit en reputatie van individuen.

Risico's in verband met het **recht op non-discriminatie** zijn er als algoritmen ten onrechte onderscheidingen maken, of die onderscheiding ten onrechte niet maken. Dat kan zich voordoen als er bij het toepassen van regels of het leggen van verbanden sprake is van over- of onderinclusiviteit van bepaalde categorieën. Discriminatierisico's kunnen zich ook concreet voordoen als vooroordelen of onaanvaardbare stereotypen via de programmeur(s) van een algoritme of via anderen die betrokken zijn bij de ontwikkeling van het algoritme (bewust of onbewust) een weerslag krijgen op het algoritme. Als zelflerende algoritmen worden ingezet bestaat verder het gevaar dat de data waarmee het algoritme wordt getraind, gevalideerd, of getest niet voldoende representatief zijn voor de groep mensen waarover wordt beslist. Een vergelijkbaar risico van discriminatie bestaat wanneer de gebruikte data een reflectie vormen van problematisch geachte maatschappelijke stigmatisering, stereotypering of vooroordelen. Als er bij de toepassing van het algoritme nieuwe data wordt verzameld die wordt gebruikt om het algoritme te trainen kan er bovendien een *feedback loop* ontstaan waarin het discriminerende effect wordt versterkt.

Het recht op **rechtsbescherming** kan in het gedrang komen doordat de werking van algoritmen vaak lastig uitlegbaar is, wat in de weg kan staan aan de inzichtelijkheid van het besluitvormingsproces en daarmee de mogelijkheid om met argumenten te ageren tegen een besluit. Het probleem van uitlegbaarheid kan komen door de koppeling van verschillende (regelgebaseerde) algoritmische systemen, maar kan ook het gevolg zijn van het gebruik van zelflerende systemen, die inherent lastig uitlegbaar zijn. Van belang is verder dat de verantwoordelijkheidsvragen toenemen en complexer te beantwoorden zijn naarmate de

samenleving verder gedigitaliseerd raakt en systemen in toenemende mate met elkaar interacteren. De daaruit resulterende onduidelijkheid over de verantwoordelijkheid in (de keten van) algoritmische besluitvorming, levert eveneens een risico op voor effectieve rechts-bescherming.

4. Bestendigheid juridisch kader

Omdat kansen en risico's van de inzet van algoritmen sterk samenhangen met het domein waarin zij worden ingezet, dient de bestendigheid van de juridische kaders ook plaats te vinden in het licht van de betreffende domeinen. Daarbij geldt dat een concrete toepassing altijd ten dele wordt gereguleerd door domeinspecifieke juridische kaders en ten dele door algemene juridische kaders.

Uit de casestudy's volgt dat de daarin bestudeerde specifieke juridische kaders niet direct in de weg lijken te staan aan het realiseren van kansen ten aanzien van de in dit onderzoek betrokken publieke waarden en belangen. Wel kan het zo zijn dat de juridische kaders onvoldoende voedingsbodem bieden om te kunnen profiteren van de voordelen van de inzet van algoritmen.

Uit de casestudy's volgt daarnaast dat de bestudeerde specifieke juridische kaders veelal voldoende ruimte bieden om geïdentificeerde risico's van algoritmische besluitvorming te vermijden of te mitigeren. Daarvoor is wel vaak vereist dat interpretaties van de bestaande ruim geformuleerde normen worden toegesneden op de specifieke inzet van het algoritme en de daarbij geldende waarden en belangen. Uit het onderzoek blijkt echter ook dat (gedeelten) van specifieke juridische kaders in sommige gevallen tekort te schieten of risico's te voorkomen of te mitigeren.

Ten aanzien van de algemene juridische kaders, zoals de AVG, de Awb en het aansprakelijkheidsrecht, kan op basis van het onderzoek worden geconcludeerd dat zij een groot absorberend vermogen hebben ten aanzien van nieuwe technologische ontwikkelingen ten aanzien van algoritmische besluitvorming. De algemene juridische kaders kunnen daardoor ten aanzien van een zich ontwikkelende technologie geleidelijk en flexibel verder vormgegeven worden. Daarvoor is wel vereist dat de nodige rechtsontwikkeling plaatsheeft waarin bijvoorbeeld rechters en toezichthouders een op algoritmen toegesneden uitleg of interpretatie geven van de algemene kaders. De eerste voorzichtige en belangrijke stappen daartoe zijn in de rechtspraak en het toezicht reeds gezet.

Hoewel er geen structurele knelpunten zijn aan te wijzen, zijn er wel verschillende knelpunten geïdentificeerd die raken aan concrete normen in de algemene kaders. Zo kunnen de strenge normen met betrekking tot het gebruik van bijzondere persoonsgegevens, zoals gegevens over etnische afkomst en seksualiteit, juist in de weg staan aan het detecteren van

discriminerende effecten in algoritmen. Ook is ten aanzien van art. 22 AVG niet duidelijk welke mate van het ontbreken van menselijke betrokkenheid nodig is om beschermd te zijn tegen volledig geautomatiseerde individuele besluitvorming en om een beroep te kunnen doen op bijbehorende informatierechten.

5. Conclusie

De eventuele regulering van algoritmen vraagt om een integrale (beleids)afweging, waarbij wordt geïdentificeerd welke van de relevante kansen en risico's voor publieke waarden en belangen zich voordoen en of het – in het licht van de behoeften, normen, waarden, belangen en context in een specifiek domein – mogelijk is om de risico's te mitigeren of te vermijden, terwijl de kansen wel kunnen worden gerealiseerd. De weging van kansen en risico's en het vinden van de balans daartussen is uiteindelijk een politiek en beleidsmatig proces. Het juridisch kader dat daarvan het resultaat is normeert en stuurt of en hoe algoritmen worden ingezet en heeft zo ook invloed op de mate waarin kansen en risico's worden gerealiseerd of vermeden.

Het is daarom van belang dat regelgeving zo wordt geformuleerd dat verantwoorde innovatie op het gebied van algoritmische besluitvorming mogelijk is. Voorkomen moet in het bijzonder worden dat normen te veel worden toegespitst op reeds bestaande technologieën en geen ruimte laten voor nieuwe ontwikkelingen. Dergelijke regelgeving biedt namelijk enkel rechtszekerheid en bescherming zolang de specifieke technologie ook daadwerkelijk gereguleerd wordt door de opgestelde regels. Als de technologieën zich ontwikkelen bestaat de mogelijkheid dat de die regelgeving niet meer actueel is, waardoor deze geen houvast meer biedt als nieuwe technologische ontwikkelingen zich voordoen. Bovendien kan het gebruik van technologiespecifieke regelgeving het zicht op de onderliggende uitgangspunten en beginselen ontnemen, wat uiteindelijk de rechtsontwikkeling niet ten goede komt. De publieke waarden die de wetgever probeerde te waarborgen, komen dan door de snelle ontwikkeling van de technologie weer op het spel te staan.

Tot slot lijkt het in het licht van de bevindingen van dit onderzoek niet zinvol om het juridisch kader in te richten op algoritmische besluitvorming in algemene zin. De algemene kaders zoals die momenteel beschikbaar zijn, vertonen geen grote tekortkomingen of structurele problemen. Integendeel: hiervoor werd al vermeld dat ze een aanzienlijk absorberend vermogen hebben. Bij de ontwikkeling en inzet van nieuwe technologieën kunnen deze algemene kaders in belangrijke mate richting bieden. De casestudy's laten bovendien zien dat de kansen en risico's voor alle onderzochte publieke waarden en belangen sterk afhankelijk zijn van het domein en de organisatorische context waarin een algoritme wordt ingezet. Nadere algemene regelgeving heeft voor het bestrijden van risico's voor de publieke waarden dan ook nauwelijks meerwaarde. De kansen en risico's voor publieke waarden en belangen moeten vooral in het licht van de dynamiek

in een domein en de verhoudingen tussen betrokken partijen in kaart worden gebracht en gewogen. Het voorgaande pleit er dan ook voor om, daar waar knelpunten worden ervaren, deze zoveel mogelijk domeinspecifiek aan te pakken. Alleen op die manier kan voldoende recht worden gedaan aan de specifieke kansen en risico's voor de specifieke publieke waarden en belangen die op het spel staan.

Inhoudsopgave

Lijst van afkortingen **15**

Hoofdstuk 1. Inleiding **17**
1.1 Algemene vraagstelling 19
1.2 Afbakening 19
1.2.1 Algoritmische besluitvorming 20
1.2.2 Kansen en risico's voor publieke waarden en belangen 21
1.2.3 Bestendigheid juridisch kader 22
1.3 Methodologie 23
1.3.1 Casestudy's 24
1.3.2 Literatuuronderzoek 25
1.3.3 Interviews 26
1.3.4 Expertmeeting 27
1.4 Leeswijzer 27

Hoofdstuk 2. Algoritmen: een introductie **29**
2.1 Algoritmen 29
2.1.1 Regelgebaseerde algoritmen 30
2.1.2 Zelflerende algoritmen 32
2.2 Organisatorische en maatschappelijke context 35
2.3 Afsluitende opmerkingen 37

Hoofdstuk 3. Publieke waarden en belangen **39**
3.1 Bescherming van persoonsgegevens 39
3.2 Non-discriminatie 42
3.3 Rechtsbescherming 43
3.4 Conclusie 44

Hoofdstuk 4. Casestudy Contentmoderatie door online platformen **45**
Stefan Kulk & Thom Snijders
4.1 Introductie 45
4.1.1 Methodologie 45
4.1.2 Opzet van de casestudy 46
4.1.3. Contentmoderatie door algoritmen 46
4.1.3.1 Onrechtmatige en onwenselijke online content 47
4.1.3.2 Contentmoderatie 48
4.1.3.3 Inzet van algoritmen 49
4.2 De aanpak van *hate speech* door online platformen 51
4.2.1 Hate speech 51
4.2.2 De werking van contentmodereeralgoritmen 53

4.2.3 Blik op de toekomst 56
4.3 Vrijheid van meningsuiting en informatie 57
4.3.1 Vrijheid van meningsuiting en contentmoderatie 57
4.3.2 Kansen, risico's en bestendigheid juridisch kader 60
4.3.3 Tussenconclusie 62
4.4 Bescherming van persoonsgegevens 62
4.4.1 Bescherming van persoonsgegevens en contentmoderatie 62
4.4.2 Kansen, risico's en bestendigheid juridisch kader 62
4.4.3 Tussenconclusie 63
4.5 Non-discriminatie 63
4.5.1 Non-discriminatie en contentmoderatie 63
4.5.2 Kansen, risico's en bestendigheid juridisch kader 64
4.5.3 Tussenconclusie 65
4.6 Rechtsbescherming 65
4.6.1 Rechtsbescherming en contentmoderatie 65
4.6.2 Kansen, risico's en bestendigheid juridisch kader 65
4.6.3 Tussenconclusie 67
4.7 Conclusie 68

Hoofdstuk 5. Casestudy Zelfrijdende auto's 71
Vicky Breemen & Anouk Wouters
5.1 Introductie 71
5.2 Beslissingsalgoritmen, sensoriek en communicatietechnologie 74
5.2.1 De werking van beslissingsalgoritmen 74
5.2.2 Huidige toepassing 75
5.2.3 Blik op de toekomst 76
5.2.4 Tussenconclusie 78
5.3 Bescherming van persoonsgegevens 79
5.3.1 Bescherming van persoonsgegevens en (deels) zelfrijdende auto's 79
5.3.2 Kansen, risico's en bestendigheid juridisch kader 80
5.3.3 Tussenconclusie 82
5.4 Non-discriminatie 82
5.4.1 Non-discriminatie en (deels) zelfrijdende auto's 82
5.4.2 Kansen, risico's en bestendigheid juridisch kader 84
5.4.3 Tussenconclusie 87
5.5 Rechtsbescherming 87
5.5.1 Rechtsbescherming en (deels) zelfrijdende auto's 87
5.5.2 Kansen, risico's en bestendigheid juridisch kader 88
5.5.3 Tussenconclusie 94
5.6 Duurzaamheid 94
5.6.1 Duurzaamheid en (deels) zelfrijdende auto's 95
5.6.2 Kansen, risico's en bestendigheid juridisch kader 96

5.6.3 Tussenconclusie 97
5.7 Verkeersveiligheid 98
5.7.1 Verkeersveiligheid en (deels) zelfrijdende auto's 98
5.7.2 Kansen, risico's en bestendigheid juridisch kader 98
5.7.3 Tussenconclusie 101
5.8 Conclusie 101

Hoofdstuk 6. Casestudy De rechtspraak 105
Stijn van Deursen & Stefan Philipsen
6.1 Introductie 105
6.1.1 Verantwoording en aanpak 106
6.2 De inzet van algoritmen in de rechtspraak 107
6.2.1 Organisatie, bedrijfsvoering en management 108
6.2.2 Rechterlijke oordeelsvorming 109
6.2.2.1 Voorbereiding van een rechterlijke beslissing 110
6.2.2.2 Een blik op de toekomst: 'De Robotrechter'? 111
6.2.3 Ontwikkelingen buiten de rechtspraak en implicaties voor de rechtspraak 113
6.2.4 Tussenconclusie 114
6.3 Bescherming van persoonsgegevens 114
6.3.1 Bescherming van persoonsgegevens en algoritmen in de rechtspraak 114
6.3.2 Kansen, risico's en bestendigheid juridisch kader 116
6.3.3 Tussenconclusie 118
6.4 Non-discriminatie 119
6.4.1 Non-discriminatie en algoritmen in de rechtspraak 119
6.4.2 Kansen, risico's en bestendigheid juridisch kader 119
6.4.3 Tussenconclusie 122
6.5 Rechtsbescherming 123
6.5.1 Rechtsbescherming en algoritmen in de rechtspraak 123
6.5.1.1 Eisen aan de rechterlijke procedure als zodanig 123
6.5.1.2 Eisen aan de rechterlijke beslissing 124
6.5.1.3 Eisen aan de rechter 124
6.5.2 Kansen, risico's en bestendigheid juridisch kader 125
6.5.2.1 Het gebruik van algoritmen in de bedrijfsvoering en organisatie 125
6.5.2.2 Het gebruik van beslisondersteuningsalgoritmen 128
6.5.2.3 Het gebruik van algoritmen als robotrechter 130
6.5.3 Tussenconclusie 133
6.6 Conclusie 135

Hoofdstuk 7. Casestudy Overheidsincasso bij verkeersboetes **141**
Martje Boekema & Susanne Heeger
7.1 Introductie 141
7.1.1 WRR-rapport ‘Weten is nog geen doen’ (2017) 142
7.1.2 Kabinetsreactie op het WRR-rapport en probleemstelling 143
7.1.3 Verantwoording keuze casestudy 145
7.1.4 Methodologie 147
7.2 Pilot Telefonisch Innen bij verkeersboetes 148
7.2.1 De pilot 148
7.2.2 Werking van het algoritme in de pilot 150
7.2.3 Blik op de toekomst 152
7.2.4 Tussenconclusie 152
7.3 Rechtsbescherming 153
7.3.1 Rechtsbescherming en algoritmen in overheidsincasso 153
7.3.2 Kansen en risico’s in relatie tot het juridisch kader 156
7.3.3 Tussenconclusie 157
7.4 Non-discriminatie 157
7.4.1 Non-discriminatie en algoritmen in overheidsincasso 157
7.4.2 Kansen en risico’s in relatie tot het juridisch kader 158
7.4.3 Tussenconclusie 160
7.5 Bescherming van persoonsgegevens 161
7.5.1 De bescherming van persoonsgegevens en algoritmen in overheidsincasso 161
7.5.2 Kansen en risico’s in relatie tot het juridisch kader 162
7.5.3 Tussenconclusie 164
7.6 Conclusie 164

Hoofdstuk 8. Kansen en risico’s **169**
8.1 Kansen 169
8.1.1 Bescherming van persoonsgegevens 170
8.1.2 Non-discriminatie 171
8.1.3 Rechtsbescherming 172
8.2 Risico’s 172
8.2.1 Bescherming van persoonsgegevens 172
8.2.2 Non-discriminatie 173
8.2.3 Rechtsbescherming 176
8.3 Organisatorische en maatschappelijk en context 177
8.4 Conclusie 179

Hoofdstuk 9. Bestendigheid van de juridische kaders **181**
9.1 Algemene en specifieke juridische kaders 181
9.2 Bestendigheid specifieke juridische kaders 181
9.3 Bestendigheid algemene juridische kaders 182
9.2.1 Bescherming van persoonsgegevens 182
9.2.2 Non-discriminatie 185
9.2.3 Rechtsbescherming 186
9.3 Conclusie 189

Hoofdstuk 10. Conclusie **191**
10.1 Kansen en risico's 191
10.2 Bestendigheid juridisch kader 193
10.3 Slot 194

Bijlage 1. Begeleidingscommissie **197**

Bijlage 2. Lijst van deelnemers Expertmeeting 27 november 2019 **199**

Bijlage 3. Geïnterviewde personen **203**

Bijlage 4. Vragenlijst semi-gestructureerde interviews **205**

Bibliografie **213**

Over het onderzoek **253**

Lijst van afkortingen

AA	Ars Aequi
AARvS	Afdeling advisering van de Raad van State
ABRvS	Afdeling bestuursrechtspraak van de Raad van State
AI	Artificial Intelligence (ook wel: KI)
AP	Autoriteit Persoonsgegevens
Art.	Artikel
AV&S	Aansprakelijkheid, Verzekering & Schade
AVG	Algemene Verordening Gegevensbescherming
Awb	Algemene wet bestuursrecht
AWGB	Algemene Wet Gelijke Behandeling
BW	Burgerlijk Wetboek
CBb	College van Beroep voor het bedrijfsleven
CBR	Centraal Bureau Rijvaardigheidsbewijzen
CJIB	Centraal Justitieel Incassobureau
COMPAS	Correctional Offender Management Profiling for Alternative Sanctions
CRM	College voor de Rechten van de Mens
CRT	Canadese Civil Resolutions Tribunal
CRvB	Centrale Raad van Beroep
DPIA	Data Protection Impact Assessment (gegevensbeschermingseffectbeoordeling)
DWJZ	Directie Wetgeving en Juridische Zaken
EHRM	Europees Hof voor de Rechten van de Mens
EU	Europese Unie
EVRM	Europees Verdrag tot bescherming van de rechten van de mens en de fundamentele vrijheden
Gw	Grondwet
Handvest	Handvest voor de Grondrechten van de Europese Unie
AI HLEG	High-Level Expert Group on Artificial Intelligence
HR	Hoge Raad
IMK	Instituut voor het Midden- en Kleinbedrijf

JenV	Ministerie van Justitie en Veiligheid
KI	Kunstmatige Intelligentie (ook wel: AI)
LOVS	Landelijk Overleg Vakinhoud Strafrecht
NJB	Nederlands Juristenblad
ODR	Online Dispute Resolution
RDW	Dienst Wegverkeer
RO	Wet op de Rechterlijke Organisatie
r.o.	rechtsoverweging
Rv	Wetboek van Burgerlijke Rechtsvordering
SAPAI	Strategisch Actieplan voor Artificiële Intelligentie
Sr	Wetboek van Strafrecht
Sv	Wetboek van Strafvordering
TAV	Tijdschrift Aansprakelijkheids- en Verzekeringsrecht
UAVG	Uitvoeringswet AVG
VNG	Vereniging Nederlandse Gemeenten
VR	Verkeersrecht (*tijdschrift*)
Wahv	Wet administratiefrechtelijke handhaving verkeersvoorschriften
Wjsg	Wet justitiële en strafvorderlijke gegevens
WODC	Wetenschappelijk Onderzoek- en Documentatiecentrum van het ministerie van Justitie & Veiligheid
Wpg	Wet politiegegevens
WRR	Wetenschappelijke Raad voor het Regeringsbeleid
WVW	Wegenverkeerswet

Hoofdstuk 1. Inleiding

Iedereen heeft in het dagelijks leven te maken met beslissingen die worden genomen door, of met behulp van algoritmen. Door het gebruik van algoritmen kunnen grote hoeveelheden gegevens in korte tijd worden verwerkt. Door de brede inzetbaarheid van algoritmegedreven technologieën in tal van maatschappelijke en economische processen kan de manier waarop onze samenleving werkt ingrijpend veranderen.

De inzet van algoritmen maakt het dagelijkse leven makkelijker en kan kansen opleveren voor het verwezenlijken van publieke waarden en belangen. Algoritmen maken processen efficiënter en dragen bij aan het vinden van oplossingen voor verschillende soorten maatschappelijke uitdagingen met behulp van data en rekenkracht.[1] Algoritmen kunnen bijvoorbeeld energieverbruik sturen, verbruik van water optimaliseren en ecosystemen monitoren, en zo bijdragen aan het oplossen van het klimaatprobleem en de bescherming van het leven op aarde. In de gezondheidszorg kunnen algoritmen ziektes diagnosticeren en bijdragen aan beter werkende medicijnen en sneller herstel van mensen.[2] Ook kunnen vluchtelingenstromen in kaart worden gebracht en kunnen algoritmen helpen om statushouders onder te brengen op plekken waar ze het meeste kans hebben om te integreren.[3] De mogelijkheden lijken eindeloos.

Tegelijkertijd kan de inzet van dergelijke technologieën ook risico's met zich brengen. Zo kunnen digitale (algoritmische) assistenten onze keuzevrijheid beperken.[4] En kunnen werknemers in de '*gig economy*' door algoritmische controle die over hen wordt uitgeoefend in de knel raken.[5] Meer algemeen doen algoritmen niet altijd recht aan de complexiteit en veelzijdigheid van de echte wereld, waardoor een risico op discriminatie kan ontstaan. Privacyrisico's ontstaan mogelijk als met algoritmen grote hoeveelheden persoonsgegevens worden verwerkt en algoritmen worden gebruikt om gegevens aan elkaar te koppelen. En rechtsbescherming en rechtsherstel kunnen in het gedrang komen door de gebrekkige inzichtelijkheid van de werking van algoritmen.

Op verschillende niveaus is er aandacht voor het toegenomen belang van algoritmen in onze samenlevingen. Zo werd in 2018 in de mededeling *Kunstmatige Intelligentie voor Europa* de noodzaak beschreven van een gemeenschappelijke aanpak voor de ontwikkeling van kunstmatige intelligentie, waarbij ruimte is voor innovatie maar tegelijkertijd de Europese waarden gewaarborgd

[1] World Economic Forum 2018.
[2] Kourou e.a*., Computational and Structural Biotechnology Journal* 2015, p. 8-17; Huang e.a*., Scientific Reports* 2018, p. 1-8.
[3] Bansak e.a*., Science* 2018, p. 325-329. Zie voor de mogelijke toepassingen daarvan in Nederland Gerritsen, Kattenberg & Vermeulen 2018, p. 14.
[4] Gal, *Michigan Technology Law Review* 2018, p. 59-104.
[5] Wood e.a., *Work, Employment and Society* 2019, p. 56-75.

worden.[6] In februari 2020 werd in vervolg daarop door de Europese Commissie een *Witboek over Kunstmatige Intelligentie* gepresenteerd, waarmee onder meer werd beoogd een ethische en juridische basis te leggen voor een Europees ecosysteem van excellentie en vertrouwen waarbinnen kunstmatige intelligentie zich op een verantwoorde manier kan ontwikkelen.[7]

Ook in Nederland staan de mogelijke gevolgen van de ontwikkeling en inzet van algoritmen hoog op de agenda. Dat komt bijvoorbeeld tot uitdrukking in de beleidsbrieven over *AI, publieke waarden en mensenrechten*[8] en *Waarborgen tegen risico's van data-analyses door de overheid*,[9] die tegelijkertijd met het *Strategisch Actieplan voor Artificiële Intelligentie* (SAPAI)[10] in oktober 2019 zijn verschenen. De mogelijke consequenties van de inzet van algoritmen voor de bescherming van publieke waarden en belangen zijn niet enkel onderwerp van politiek en beleidsmatig debat, maar ook van verschillende juridische procedures. Voorbeelden zijn de rechtszaken over het Systeem Risico Indicatie (SyRI)[11] en het Programma Aanpak Stikstof (PAS).[12]

De gevolgen van de inzet van algoritmen en de ontwikkeling van kunstmatige intelligentie in bredere zin zijn reeds in aantal onderzoeken in kaart gebracht. Zo heeft het Rathenau Instituut in 2017 twee rapporten gepubliceerd die in dit kader relevant zijn, te weten *Opwaarderen. Het borgen van publieke waarden in de digitale samenleving*[13] en *Mensenrechten in het robottijdperk*.[14] Het onderzoek van de WRR naar de invloed van *big data* is beschreven in het rapport *Big Data in een vrije en veilige samenleving*.[15] Deze rapporten vormden voor de overheid mede aanleiding om meer verdiepend onderzoek te laten doen naar de impact van kunstmatige intelligentie op publieke waarden en belangen. Het ministerie van Binnenlandse Zaken en Koninkrijksrelaties heeft onderzoek laten verrichten naar de impact van algoritmen op de grondrechten[16] en het toezicht op gebruik van algoritmen door de overheid.[17] In aanvulling op deze onderzoeken voert de WRR op dit moment een overkoepelend, multidisciplinair onderzoek uit.[18]

Het onderzoeksrapport dat voor u ligt is primair geschreven vanuit een toegepast-juridische invalshoek. De opdracht tot het uitvoeren van dit onderzoek is gegeven door het Wetenschappelijk Onderzoeks- en Documentatiecentrum (WODC) op verzoek van Directie Wetgeving en Juridische

[6] COM(2018) 237 final.
[7] COM(2020) 65 final.
[8] Brief van de Minister van Binnenlandse Zaken en Koninkrijksrelaties van 8 oktober 2019, *Kamerstukken II* 2019/20, 26643, nr. 642.
[9] Brief van de Minister voor Rechtsbescherming van 8 oktober 2019, *Kamerstukken II* 2019/20, 26643 en 32761, nr. 641.
[10] SAPAI 2019.
[11] Rechtbank Den Haag 5 februari 2020, ECLI:NL:RBDHA:2020:865 (*SyRI)*.
[12] ABRvS 17 mei 2017, ECLI:NL:RVS:2017:1259 (*Stichting Werkgroep Behoud de Peel/GS Noord Brabant)*.
[13] Kool e.a. 2017.
[14] Van Est & Gerritsen 2017.
[15] WRR 2016.
[16] Zie Vetzo, Gerards & Nehmelman 2018.
[17] Frissen, Van Eck en Drouen 2019.
[18] Zie daarover wrr.nl/onderwerpen/artificiele-intelligentie.

Zaken (DWJZ) van het Ministerie van Justitie en Veiligheid. Met deze onderzoeksopdracht wordt tevens invulling gegeven aan één van de aangekondigde actiepunten in het kader van het *Strategisch Actieplan voor Artificiële Intelligentie*.[19]

In dit onderzoeksrapport verkennen we de juridische aspecten van algoritmische besluitvorming. Dat doen we in het licht van de huidige stand van de techniek en de ontwikkelingen die ten aanzien daarvan in de komende vijf tot tien jaar te verwachten zijn. De focus ligt daarbij op de toepassingen van algoritmen in een viertal door het WODC en DWJZ geselecteerde domeinen, waarnaar aan de hand van casestudy's onderzoek gedaan wordt. We besteden aandacht aan specifieke toepassingen van algoritmen en brengen in kaart in hoeverre de bestaande juridische kaders voldoende bestendig zijn om de kansen van algoritmische besluitvorming te realiseren en de risico's daarvan te vermijden of de gevolgen daarvan te mitigeren.

1.1 Algemene vraagstelling

De tweeledige vraag die centraal staat in dit onderzoek luidt als volgt:

Welke kansen en risico's doen zich voor bij algoritmische besluitvorming met betrekking tot de bescherming en realisering van publieke waarden en belangen, en zijn de bestaande juridische kaders voldoende bestendig om kansen te verwezenlijken en het intreden van geïdentificeerde risico's te voorkomen of de gevolgen daarvan te mitigeren?

Het eerste deel van deze hoofdvraag richt zich op een identificatie van de kansen en risico's die zich kunnen voordoen als gevolg van algoritmische besluitvorming in het licht van publieke waarden en belangen. Het tweede deel is gericht op een beoordeling van de bestendigheid van de huidige toepasselijke juridische kaders in het licht van die kansen en risico's voor publieke waarden en belangen.

1.2 Afbakening

Het onderzoek dat nu voor u ligt is langs een aantal lijnen begrensd. Ten eerste strekt ons onderzoek ertoe de bestaande juridische kaders te beoordelen. Daarbij leunen wij op onze bevindingen ten aanzien van de inzet van algoritmen in vier door het WODC en DWJZ geselecteerde domeinen, die wij aan de hand van casestudy's in kaart brengen. Ons onderzoek heeft daarmee een (toegepast) juridisch karakter. Beleidsmatige en ethische vraagstukken vallen buiten de reikwijdte van dit onderzoek. Wel geldt vanzelfsprekend dat ethische uitgangspunten ook door het geldende recht gewaarborgd kunnen zijn.

[19] Zie SAPAI 2019, p. 58.

Ten tweede geldt ten aanzien van de bestudeerde technologieën dat wij ons hebben beperkt tot het inventariseren van kansen en risico's die zich de komende vijf tot tien jaar kunnen voordoen. De voorspelbaarheid van ontwikkelingen op het gebied van algoritmische besluitvorming is vanwege de grote innovatiesnelheid beperkt. Het is dan ook weinig zinvol om de ontwikkeling op (nog) langere termijn te voorspellen. Wij nemen daarom ook in de casestudy's steeds het huidige gebruik van een algoritme tot uitgangspunt, waarbij zoveel mogelijk de op dit moment voorzienbare ontwikkelingen in de technologie en het gebruik van algoritmen worden meegenomen. Vraagstukken die samenhangen met de inzet van toekomstige technologieën die hun intrede nog niet hebben gedaan, zoals *general artificial intelligence*, die in staat zou zijn om alle menselijke taken over te nemen, blijven daarom buiten beschouwing.[20]

Tot slot is het onderzoek begrensd door de invulling die wordt gegeven aan een aantal begrippen die centraal staan in dit onderzoek. Dit onderzoek is gericht op *algoritmische besluitvorming*; op *kansen en risico's* voor de bescherming en realisering van *publieke waarden en belangen*; en op (de bestendigheid van) de *juridische kaders*. Hieronder gaan we nader in op deze begrippen.

1.2.1 Algoritmische besluitvorming

Dit onderzoek richt zich op algoritmische besluitvorming. Daarmee doelen wij op besluitvormingsprocessen waarin een algoritme een rol speelt. Het gaat daarbij om gevallen waarin een algoritme zelf een besluit neemt, of waarin de uitvoer van een algoritme wordt meegenomen in een menselijk besluitvormingsproces.

Een *algoritme* is in de kern een geschreven computerinstructie die met een bepaalde waarde of een reeks waarden als invoer, een bepaalde waarde of reeks waarden als uitvoer produceert.[21] De technologische aspecten van verschillende soorten algoritmen worden uitgebreider besproken in par. 2.1 van dit rapport.

Onder *besluitvorming* verstaan wij in dit rapport processen die erop gericht zijn om beslissingen te nemen die raken aan de rechtspositie van rechtssubjecten of hen anderszins in hun belangen treffen. Ons begrip van algoritmische besluitvorming beperkt zich daarom niet tot puur juridische besluiten, waarbij het rechtsgevolg centraal staat, maar omvat ook besluiten met *feitelijke* gevolgen die raken aan de belangen van een rechtssubject.

[20] Met *general artificial intelligence* wordt bedoeld een vorm van kunstmatige intelligentie die autonoom complexe problemen in meerdere gebieden kan oplossen en die daarmee menselijke intelligentie benadert of overtreft. Deze vorm van kunstmatige intelligentie wordt veelal onderscheiden van *narrow artificial intelligence*, waarvan sprake is als een systeem in één of meerdere specifieke gebieden een probleem kan oplossen, zoals tijdens het spelen van een schaakspel of het besturen van een auto. Zie daarover uitvoerig Pennachin & Goertzel 2007.
[21] Cormen e.a. 2009, p. 5. Zie ook Gillespie 2016.

Dit onderzoek omvat, op verzoek van het WODC, ook de beslissingen die worden genomen door algoritmen in zelfrijdende auto's. Veel van deze beslissingen, zoals het wisselen van rijbaan, zijn triviaal en zijn er niet op gericht de rechtspositie van rechtssubjecten te veranderen, noch raken ze direct aan belangen van rechtssubjecten. De beslissingen van zelfrijdende auto's zijn dan ook niet direct aan te merken als algoritmische besluitvorming in de zin van dit onderzoek. Deze beslissingen kunnen echter wel (grote) gevolgen hebben als daarbij iets misgaat. In het licht van de groeiende rol die algoritmen spelen in het dagelijks leven, en de gevolgen die dat kan hebben, is de zelfrijdende auto daardoor desalniettemin een relevante casestudy.[22]

Ook van belang is dat het begrip 'besluitvormingsproces' in dit onderzoek meer omvat dan alleen het proces van invoer naar uitvoer van een bepaald algoritme, en ook meer dan de uitvoer van het algoritme alleen. Ook besluiten die door mensen worden genomen op basis van bijvoorbeeld informatie of adviezen van algoritmen, zoals het geval is bij adviserende en ondersteunende algoritmegebaseerde systemen, zijn in dit onderzoek betrokken.

1.2.2 Kansen en risico's voor publieke waarden en belangen

Dit onderzoek spitst zich toe op de kansen en risico's voor de verwezenlijking van publieke waarden en belangen die zich kunnen voordoen als gevolg van algoritmische besluitvorming. Bij de beoordeling of sprake is van kansen en risico's hanteren wij publieke waarden en belangen als normatief kader.[23] Een kans wordt in dat licht gedefinieerd als de mogelijkheid om een publieke waarde of publiek belang te realiseren. Een risico wordt gedefinieerd als de mogelijkheid dat een publieke waarde of publiek belang niet (volledig) verwezenlijkt wordt of zelfs geschaad wordt.

Bij de inzet van algoritmen in concrete besluitvormingsprocessen zullen zich zowel kansen als risico's voordoen. Kansen van algoritmische besluitvorming bestaan dan niet zonder de daarmee samenhangende risico's. Overheden en private actoren die beslissen over de inzet van algoritmen zullen daarbij, met inachtneming van toepasselijke juridische kaders, uiteraard een gedegen afweging moeten maken. In dit onderzoek worden de kansen en risico's van de inzet van algoritmen desalniettemin afzonderlijk besproken om op inzichtelijke wijze aan te geven hoe publieke waarden en belangen in het geding kunnen komen en wat dat betekent voor de bestendigheid van bestaande juridische kaders.

Voor het definiëren van publieke waarden zoeken we aansluiting bij de grondrechten. Grondrechten vormen een belangrijke juridische uitdrukking en concretisering van een groot aantal van de in een samenleving breed gedragen publieke waarden en belangen, in het bijzonder van belangrijke waarden als autonomie en menselijke waardigheid van burgers. De onderhavige studie

[22] Zie voor een verdere verantwoording met betrekking tot de casestudy's par. 1.3.1.
[23] De belangrijke rol van publieke waarden komt ook uitdrukkelijk naar voren in Kool e.a. 2017. Zie Van Est e.a. 2019.

richt zich daarom hoofdzakelijk op publieke waarden en belangen die tot uitdrukking zijn gebracht in de grondrechten. Deze aanpak vindt nadrukkelijk steun in de visie van de *High Level Group on AI* (AI HLEG) van de Europese Commissie, namelijk dat '[e]erbiediging van grondrechten, binnen een kader van democratie en de rechtsstaat, (...) de meest veelbelovende grondslagen [biedt] voor het vaststellen van abstracte ethische beginselen en waarden die in het kader van [kunstmatige intelligentie] kunnen worden geoperationaliseerd'.[24] De keuze voor een dergelijke grondrechtelijke invulling van publieke waarden en belangen stelt ons bovendien in staat om bestaande juridische kaders op zo objectief mogelijke wijze te toetsen. De grondrechten maken immers al deel uit van het breed geaccepteerde stelsel van wet- en regelgeving en verdragen.

Uit vooronderzoek naar de toepassing van algoritmen in de verschillende domeinen waarop de casestudy's betrekking hebben, is gebleken dat een drietal als grondrechten beschermde waarden en belangen steeds van belang is.[25] Het gaat daarbij om het recht op gegevensbescherming, het recht op non-discriminatie en het recht op rechtsbescherming. Deze drie grondrechten en de wijze waarop zij tot uitdrukking komen in wet- en regelgeving, nemen daarom een belangrijke plek in dit onderzoek in. Wij zullen hiernaar verwijzen als 'casusoverstijgende publieke waarden en belangen'. De drie genoemde grondrechten worden nader besproken in hoofdstuk 3 van dit rapport.

In aanvulling op die casusoverstijgende publieke waarden en belangen spelen in de verschillende casestudy's ook andere, veelal meer specifieke, publieke waarden en belangen een rol. Zij worden hierna 'casusspecifieke publieke waarden en belangen' genoemd. Vanwege hun casusspecifieke aard worden ze in de casestudy's zelf besproken en blijven ze in deze algemene inleiding buiten beschouwing. Ook deze publieke waarden en belangen kunnen beschermd zijn door grondrechten. Echter, in de casestudy's wordt er bij het definiëren van de casusspecifieke belangen ook aansluiting gezocht bij waarden en belangen die niet (uitdrukkelijk) geconcretiseerd zijn in grondrechten.

1.2.3 Bestendigheid juridisch kader

In dit onderzoek toetsen we de bestendigheid van de juridisch kaders die het specifieke domein van een casestudy reguleren. Door de ontwikkeling en toepassing van technologieën zoals algoritmen kunnen besluitvormingsprocessen op een zodanig andere manier gaan verlopen dat het juridisch kader niet meer geschikt is om publieke waarden en belangen afdoende te waarborgen of om keuzes en afwegingen te sturen.[26] Enerzijds stelt de inzet van algoritmen de samenleving potentieel in staat om publieke waarden en belangen te garanderen, bijvoorbeeld

[24] AI HLEG 2019. Zie voor deze koppeling van publieke waarden aan grondrechten ook *Kamerstukken II* 2019/20, 26643, nr. 642, p. 4 en 8.
[25] Zie in dat verband ook de grondrechten die worden genoemd in *Kamerstukken II* 2019/20, 26643, nr. 642, p. 17.
[26] Agenda Digitale Overheid (NL Digibeter) 2019, p. 29.

doordat de technologie helpt om discriminatie te detecteren en te bestrijden. Anderzijds is denkbaar dat het bestaande juridisch kader de toepassing van algoritmen niet toelaat of belemmerend werkt ten aanzien van toekomstige toepassingen van die technologieën. Bovendien is het denkbaar dat het juridisch kader onvoldoende in staat stelt om specifieke risico's van de inzet van algoritmen ten aanzien van publieke waarden en belangen te voorkomen of te mitigeren.

Omdat het onderzoek zich toespitst op de in de casestudy's onderzochte domeinen en de ontwikkelingen die in de komende vijf tot tien jaar in die domeinen te verwachten zijn, richten wij ons primair op de knelpunten die volgen uit het juridisch kader ten aanzien van die ontwikkelingen. We kunnen dan ook geen uitspraken doen over andere domeinen waarin de inzet van algoritmen tot kansen zou kunnen leiden en mogelijk op juridische grenzen stuit. Wel kunnen wij in algemene zin identificeren waar het juridisch kader belemmerend kan werken ten aanzien van de meer technologische aspecten die samenhangen met de inzet van algoritmen in besluitvormings-processen.

In de casestudy's onderzoeken wij hoe de juridische kaders zich verhouden tot de bescherming van publieke waarden en belangen in het licht van de verschillende algoritmische technologieën en toekomstige ontwikkelingen daarvan. Daarbij is van belang dat de toepassing van algoritmen in deze domeinen niet enkel gereguleerd wordt door de specifieke kaders die in dat domein gelden, maar dat daarin ook algemene kaders met een breder toepassingsgebied gelden. Beide typen kaders komen in de casestudy's aan bod.

Van belang is ook dat algoritmen functioneren in een bredere organisatorische, maatschappelijke en technologische context. Daardoor is voor het op bestendige wijze borgen van publieke waarden en belangen niet alleen een rol weggelegd voor klassieke juridische reguleringsinstrumenten.[27] Ook organisaties die algoritmen inzetten kunnen daarbij een rol spelen, net als het maatschappelijk veld in het algemeen. Het is bovendien denkbaar dat het realiseren van publieke waarden en belangen plaatsheeft in of vanwege de technologieën zelf. De organisatorische, maatschappelijke en technologische aspecten van het borgen van publieke waarden worden daarom eveneens in onze analyse betrokken.

1.3 Methodologie

De onderzoeksvraag wordt beantwoord aan de hand van een onderzoek naar algoritmische besluitvorming in algemene zin en casestudy-onderzoek. Het onderzoek is op 16 april 2020 afgesloten. Ontwikkelingen van na die datum zijn in dit onderzoek niet meegenomen. Zoals gebruikelijk bij onderzoeken in opdracht van het WODC is ook voor dit onderzoek een

[27] Zie over die context ook par. 2.2.

Begeleidingscommissie ingesteld.[28] De opdrachtgever heeft daarnaast een klankbordgroep ingesteld bestaande uit beleidsverantwoordelijke ambtenaren bij de verschillende betrokken ministeries, die elk op hun eigen casus hebben meegelezen en konden controleren op feitelijke en juridische onjuistheden.[29]

1.3.1 Casestudy's

De casestudy's maken het mogelijk om ten aanzien van concrete toepassingen van algoritmische besluitvorming 'diepteboringen' te doen. Voor specifieke onderwerpen kunnen daarbij de kansen en risico's in kaart worden gebracht en aan de hand daarvan kan de bestendigheid van de toepasselijke juridische kaders worden getoetst.[30] De casestudy's zijn zo een goede illustratie van de variëteit van mogelijke toepassingen van algoritmische systemen en de manier waarop de bestaande juridische kaders daarop inhaken.

Tegelijkertijd is het belangrijk om te onderkennen dat deze methode beperkingen met zich brengt. De belangrijkste daarvan voor dit onderzoek is dat de bevindingen uit een casestudy ten aanzien van geïdentificeerde kansen en risico's en de bestendigheid van de toepasselijke juridische kaders beperkt generaliseerbaar zijn.[31] De casestudy's verschillen voor wat betreft het domein waarin algoritmen worden ingezet. Ook het geheel van toepasselijke juridische kaders die van toepassing zijn op de inzet van algoritmen is anders per casestudy. Bovendien is van belang dat de insteek van de bestudering van de inzet van algoritmen verschilt per casestudy. Zo wordt in de casestudy naar de rechtspraak niet één toepassing van algoritmische besluitvorming in de rechtspraak bestudeerd maar wordt de (mogelijke) inzet van algoritmen in dit domein in brede zin onderzocht. In de casestudy naar overheidsincasso bij verkeersboetes wordt daarentegen juist een concrete toepassing van algoritmische besluitvorming door het Centraal Justitieel Incassobureau (CJIB) onder de loep genomen. Wij zullen daarom met de nodige voorzichtigheid op een hoger abstractieniveau reflecteren op de implicaties van de inzet van algoritmen voor verschillende domeinen in de samenleving en op de vraag in hoeverre publieke waarden en belangen daarbij op bestendige wijze geborgd zijn.

[28] Zie Bijlage 1 voor de leden van de begeleidingscommissie.
[29] Zie Bijlage 2 voor de leden van de klankbordgroep.
[30] Vgl. Gagnon 2010, p. 2.
[31] In algemene zin over die generaliseerbaarheid Gagnon 2010, p. 3. Zie over de relevantie van de selectie van casestudy's Yin 2009, p. 255.

In het kader van dit onderzoek zijn casestudy's verricht naar algoritmische besluitvorming in de volgende toepassingsgebieden:

1) Contentmoderatie door online platformen
2) Zelfrijdende auto's
3) De rechtspraak
4) Overheidsincasso bij verkeersboetes

Deze casestudy's zijn door het WODC in overleg met betrokkenen binnen de ministeries geselecteerd, waarbij deels gebruik is gemaakt van een door de onderzoekers aangeleverde longlist van mogelijke casestudy's. De keuze voor de casestudy Overheidsincasso bij verkeersboetes is in onderling overleg met het ministerie gemaakt en is ingegeven door de wens om tegemoet te komen aan een toezegging aan de Tweede Kamer. Deze strekt ertoe om onderzoek te laten doen naar de inzet van algoritmen ten behoeve van de vraag "welke kansen kunstmatige intelligentie kan bieden voor een tijdige signalering, vooral in massale besluitvormingsprocessen, van mensen die door omstandigheden (tijdelijk) mogelijk niet zelfredzaam zijn", teneinde hen maatwerk te kunnen bieden.[32] Voor een nadere toelichting op deze casestudy wordt verwezen naar par. 7.1. Aanvankelijk was naast bovengenoemde vier casestudy's ook een casestudy naar peer-to-peer handel in energie voorzien. Deze kon door overmacht echter niet tijdig worden afgerond en is daarom, in overleg met de opdrachtgever, niet in dit rapport opgenomen. Het streven is deze casestudy op een later moment af te ronden, waarna deze als afzonderlijk document naast onderhavig hoofdrapport zal worden gepubliceerd op de website van het WODC.

Om een goed beeld te kunnen vormen van de kansen en risico's voor publieke waarden en belangen, en de bestendigheid van de juridische kaders, komen per casestudy steeds de volgende aspecten aan de orde:

- de techno-sociale ontwikkelingen in het besproken domein;
- de voor de casestudy relevante publieke waarden en belangen en hun vertaling in het toepasselijk juridisch kader;
- de bestendigheid van de juridische kaders ten behoeve van de borging van publieke waarden en belangen;
- de noodzaak of wenselijkheid van aanpassingen in dit juridisch kader.

1.3.2 Literatuuronderzoek

Het fundament van het onderzoek wordt zowel bij de algemene hoofdstukken (hoofdstukken 1-3 en 8-10) als bij de casestudy's gevormd door een studie naar wet- en regelgeving, rechtspraak en wetenschappelijke literatuur. Om het technologische domein en de werking van de verschillende

[32] *Kamerstukken II* 2017/18, 34775-VI, nr. 88, par. 4.

systemen goed te kunnen doorgronden hebben wij vanzelfsprekend ook literatuur over die onderwerpen bestudeerd. Een volledig overzicht van de door ons bestudeerde literatuur is te vinden in de bibliografie.

1.3.3 Interviews

Om de informatie die is vergaard tijdens het literatuurzoek aan te vullen en de inzichten die op basis daarvan zijn verkregen te verifiëren, zijn voor dit onderzoek diepte-interviews afgenomen. Voor iedere casestudy en ook voor de algemene hoofdstukken van dit rapport zijn ten minste vijf interviews uitgevoerd. In totaal is met 42 respondenten in 32 interviews gesproken.[33] De overgrote meerderheid van de interviews is *face-to-face* afgenomen. Zeven interviews zijn telefonisch of via een videoverbinding verricht, bijvoorbeeld omdat de respondent zich in het buitenland bevond. Een lijst met respondenten alsmede de organisatie waar zij werkzaam zijn, is te vinden in Bijlage 4 bij dit rapport.

De onderzoekers hebben er bij het selecteren van de te interviewen personen naar gestreefd om steeds te spreken met (1) de producent of ontwikkelaar van het betreffende algoritme; (2) een gebruiker van het algoritme en/of een persoon die ermee werkt; (3) een beleidsmaker die actief is in het domein waarin het algoritme wordt gebruikt; (4) een persoon die is onderworpen aan de betreffende algoritmische besluitvorming of een vertegenwoordiger van een organisatie die de belangen van betrokkenen behartigt; en (5) een (juridisch) expert of onderzoeker die kan reflecteren op het gebruik van het algoritme en de toepasselijke juridische kaders. Waar relevant wordt de selectie van geïnterviewden nader toegelicht in de verschillende casestudy's, nu deze selectie tevens samenhangt met de specifieke vragen en informatiebehoefte in het betreffende casestudy-onderzoek.

Bij de interviews hebben wij gewerkt met een semigestructureerde vragenlijst. De structuur van het interview, alsmede de vragen die in ieder geval aan de orde moesten komen, zijn daarbij van tevoren bepaald en door alle betrokken (casestudy-)onderzoekers gehanteerd. Deze vragen zijn te vinden in Bijlage 5 bij dit rapport. Binnen deze kaders bestond voor de interviewers ruimte om nieuwe of aanvullende vragen toe te voegen als daaraan behoefte was. In het onderzoek is daarnaast gewerkt met open vragen, waardoor geïnterviewden de mogelijkheid hadden om verder uit te weiden en verbanden te leggen met andere relevante ontwikkelingen in het betreffende domein.

Van ieder interview is door de onderzoekers een verslag gemaakt. Aan geïnterviewden is steeds medegedeeld dat niet rechtstreeks uit de interviews of de verslagen daarvan zou worden geciteerd

[33] Zeven interviews betroffen groepsinterviews, met twee of meer respondenten gezamenlijk. Zie Bijlage 4 voor nadere toelichting.

en dat uitspraken niet te herleiden zouden zijn tot personen of organisaties, tenzij anders aangegeven en in dat geval onder voorwaarde van expliciete toestemming.

Voor het formuleren van interviewvragen met betrekking tot de ontwikkeling en werking van de technologieën hebben we gebruikgemaakt van algemene (technische) literatuur, richtlijnen en aanbevelingen met betrekking tot het verantwoordelijk gebruik van algoritmen. Voor het opstellen van de vragenlijst hebben we verder aansluiting gezocht bij de 'levenscyclus' van een algoritme. Globaal is die cyclus onder te verdelen in drie fases.[34] De eerste fase is de *probleemanalyse*, waarin het probleem dat met het algoritme moet worden opgelost wordt onderzocht en geanalyseerd en het te bereiken doel wordt bepaald. In de tweede fase wordt het algoritme *ontwikkeld*. In deze fase wordt het algoritme ontworpen (geprogrammeerd) en eventueel getraind aan de hand van verzamelde en bewerkte data. In deze fase wordt het algoritme ook getest en eventueel aangepast. In de derde fase wordt het algoritme in gebruik genomen door de organisatie en ingezet in een *besluitvormingsproces*. Dit is ook de fase waarin mensen gevolgen kunnen ondervinden van het gebruik van het algoritme. Hoewel deze indeling noodzakelijkerwijze een versimpeling van de werkelijkheid vormt, bood zij structuur aan de interviews. Bovendien draagt deze indeling bij aan het doel om op consistente wijze een zo volledig mogelijk beeld te krijgen van de ontwikkeling en het gebruik van de algoritmen.

De functie van de interviews is afhankelijk van de specifieke informatiebehoefte in een casestudy. Zo geldt dat de ontwikkelingen in bepaalde door de casestudy's bestreken gebieden vrij goed in de literatuur gedocumenteerd zijn. De interviews ondersteunen dan vooral de selectie van literatuur en de verificatie van relevante bevindingen. In andere gebieden – dit geldt met name voor de casestudy Overheidsincasso bij verkeersboetes – is de beschikbaarheid van (openbare) informatie minder groot. De interviews vormen in die gevallen een belangrijke primaire informatiebron. In de introductie van de verschillende casestudy's wordt, waar nodig, nader ingegaan op de rol van de interviews.

1.3.4 Expertmeeting

Op 27 november 2019 is ten behoeve van onderhavig onderzoek een expertmeeting georganiseerd, waaraan 21 experts deelnamen. De casestudy-onderzoekers presenteerden daarbij de uitkomsten van hun casestudy en hebben in kleinere breakoutsessies met experts gesproken over hun bevindingen. Ook is er in een plenair gedeelte in meer algemene zin gesproken over de kansen en risico's van algoritmische besluitvorming, de bestendigheid van de betrokken juridische kaders, en mogelijke oplossingsrichtingen. De lijst met deelnemers aan deze bijeenkomst is te vinden in Bijlage 3 bij dit rapport.

[34] AI HLEG 2019, p. 47.

1.4 Leeswijzer

Dit onderzoeksrapport is onderverdeeld in algemene hoofdstukken die hoofdzakelijk betrekking hebben op generieke aspecten van algoritmische besluitvorming (hoofdstukken 1-3 en 8-10) en vier casestudy's, waarin de gevolgen van algoritmische besluitvorming voor het realiseren van publieke waarden in vier domeinen centraal staan (hoofdstukken 4-7).

In hoofdstuk 2 gaan wij in op de verschillende typen algoritmen die in dit onderzoek worden onderscheiden. In hoofdstuk 3 staan vervolgens de drie algemene publieke waarden en belangen centraal. In de hoofdstukken 4-7 bespreken wij de inzet van algoritmen in vier specifieke domeinen in het licht van de drie algemene publieke waarden en belangen. Ook worden daar – waar relevant – casestudyspecifieke waarden en belangen geïdentificeerd. Aan de hand van de in die specifieke domeinen geïdentificeerde kansen en risico's, gaan de verschillende betrokken onderzoekers in de casestudy's in op de bestendigheid van de betrokken juridische kaders. In hoofdstuk 8 bespreken wij vervolgens in algemene lijnen de kansen en risico's van de inzet van algoritmen voor de algemene publieke waarden en belangen. Daarbij wordt geput uit de bevindingen van de casestudy's. In hoofdstuk 9 bespreken wij de bestendigheid van de algemene juridisch kaders in het licht van de geïdentificeerde kansen en risico's. Eén en ander wordt tot slot samengebracht in het concluderende hoofdstuk 10.

Hoofdstuk 2. Algoritmen: een introductie

Centraal in dit hoofdstuk staan de verschillende typen algoritmen die gebruikt kunnen worden bij algoritmische besluitvorming. Het gaat daarbij om algoritmen die informatie aandragen op basis waarvan een besluit wordt genomen, maar ook om algoritmen die zelfstandig een besluit nemen. In dit hoofdstuk bespreken we de mogelijkheden en beperkingen die samenhangen met de aard van de verschillende typen algoritmen, die onafhankelijk van de inzet van dergelijke systemen in een specifiek domein bestaan. Om te kunnen doorgronden hoe algoritmen werken en welke gevolgen hun inzet kan hebben, kan echter niet worden volstaan met een beschouwing van de eigenschappen van de verschillende typen algoritmen. In het tweede deel van dit hoofdstuk benadrukken we daarom het belang om algoritmen te beschouwen in de organisatorische context waarin een algoritme wordt ingezet en bespreken we de bredere maatschappelijke context.

2.1 Algoritmen

Een algoritme is in de kern een omschreven computerinstructie die met een bepaalde waarde of een reeks waarden als *invoer*, een bepaalde waarde of reeks waarden als *uitvoer* produceert. Het algoritme is dus de opeenvolging van stappen die de invoer verwerkt tot een uitvoer om daarmee een specifiek probleem op te lossen.

De invoer van een algoritme kan bestaan uit een grote verscheidenheid aan gegevens; van demografische en sociaaleconomische gegevens tot geografische of financiële informatie. De uitvoer van het algoritme kan een zelfstandig besluit inhouden (zoals de beslissing om een snelheidsovertreding te beboeten), maar kan ook worden meegenomen in een complexer besluitvormingsproces waarin een mens uiteindelijk beslist. Dit is bijvoorbeeld het geval bij de ondersteuningsalgoritmen die beschreven worden in de casestudy's naar de rechtspraak en het modereren van online content door platformen.[35]

Algoritmen kunnen verschillen afhankelijk van de doelen waarvoor ze worden ingezet. Daarbij kan een onderscheid gemaakt worden tussen beschrijvende, diagnostische, voorspellende en voorschrijvende algoritmen.[36] Beschrijvende algoritmen geven een analyse van wat er in bepaald geval *gebeurt*. Diagnostische algoritmen strekken ertoe een bepaald geval *te verklaren*. Voorspellende algoritmen doen een poging te voorspellen wat er in een bepaald geval *zal gebeuren*. Voorschrijvende algoritmen geven aan wat in een bepaald geval *moet gebeuren*.

[35] Zie respectievelijk hoofdstukken 4 en 6.
[36] Bijlage bij *Kamerstukken II* 2018/19, 26643, nr. 641.

Waar met algoritmen beslissingen worden genomen ten aanzien van personen, dan kan daarnaast een onderscheid gemaakt worden tussen algoritmen die (helpen) beslissen op grond van individuele kenmerken of omstandigheden, en algoritmen waarbij sprake is van profilering. In het laatste geval worden bepaalde persoonskenmerken van een persoon gekoppeld aan een bepaald profiel op grond waarvan aannames worden gedaan over die persoon.[37] Dergelijke algoritmen kunnen bijvoorbeeld aangeven hoe groot de kans is dat een persoon tot een bepaalde groep behoort of bepaald gedrag vertoont of zal vertonen.

In dit onderzoek speelt het onderscheid dat bovendien gemaakt kan worden tussen *regelgebaseerde* algoritmen en *zelflerende* algoritmen een belangrijke rol. Dat onderscheid wordt hieronder toegelicht. Het is daarbij goed te beseffen dat in de praktijk ook gebruikgemaakt wordt van combinaties van (verschillende typen) algoritmen. Zo kan een regelgebaseerd algoritme gebaseerd zijn op verbanden die door een zelflerend algoritme zijn vastgesteld.[38]

2.1.1 Regelgebaseerde algoritmen

Een belangrijk kenmerk van regelgebaseerde algoritmen is dat zij van tevoren geprogrammeerd zijn volgens een 'als dit, dan dat'-structuur.[39] Daartoe worden processen vertaald in variabelen en regels zodat computers daarmee kunnen werken.[40] De werking van deze algoritmen ligt dus bij voorbaat vast. Op basis van dezelfde invoergegevens zal een regelgebaseerd algoritme dan ook altijd tot dezelfde uitkomst komen, waarmee de werking in beginsel voorspelbaar is. Regelgebaseerde algoritmen zijn, door hun capaciteit om verschillende feitencomplexen te analyseren en daarop regels toe te passen, in staat om snel taken uit te voeren die voor individuen te complex zijn of die te veel tijd kosten. Zij kunnen gebruikers bijvoorbeeld wijzen op afwijkende gegevens of grote hoeveelheden gegevens in korte tijd classificeren.[41] Regelgebaseerde algoritmen zijn bij uitstek geschikt voor het automatiseren van beslisbomen voor processen waarin alle denkbare scenario's vooraf kunnen worden voorzien.

Regelgebaseerde algoritmen worden vaak toegepast in zogenaamde expertsystemen.[42] Bij de ontwikkeling van een expertsysteem werken experts in een bepaald gebied samen met programmeurs met het doel om dat specifieke gebied te modelleren. De kennis van de experts wordt dan omgezet in variabelen en regels. Het expertsysteem kan vervolgens aan de hand van invoergegevens door toepassing van de regels tot een uitvoer komen.[43] Om een expertsysteem te programmeren, dient dus sprake te zijn van kennis die zich ertoe leent om in regels omgezet te

[37] Hildebrandt & Gutwirth 2008, p. 19. Zie ook art. 4 onder 4 AVG.
[38] Te denken valt aan een zelflerend algoritme dat in een grote gegevensverzameling relevante factoren ontdekt die vervolgens worden gebruikt in een regelgebaseerd algoritme dat de factoren afweegt.
[39] Grosan & Abraham 2011, p. 149.
[40] Surden, *Georgia State University Law Review* 2019, p. 1316; Shoham, *Communications of the ACM* 2016, p. 47-49.
[41] Surden, *Georgia State University Law Review* 2019, p. 1318.
[42] Grosan & Abraham 2011, p. 154.
[43] Surden, *Georgia State University Law Review* 2019, p. 1316-1317.

worden. Daarbij kan onder meer gedacht worden aan medische kennis, maar ook aan wet- en regelgeving.

Al met al lenen regelgebaseerde algoritmen zich goed voor het automatiseren van relatief simpele en overzichtelijke besluitvormingsprocessen waarin overeenstemming bestaat over het doel van de besluitvorming en de daarbij relevante criteria en waarin de benodigde informatie aanwezig is om het algoritme naar behoren te laten functioneren.[44] Regelgebaseerde algoritmen stellen organisaties in staat om consistente beslissingen te nemen in gelijke gevallen en menselijke fouten bij het uitvoeren van repetitieve taken te elimineren of in ieder geval het risico daarop te beperken. Dergelijke algoritmen zijn over het algemeen ook goed uitlegbaar, nu de uitvoer van het algoritme is te achterhalen door aan de hand van de gegeven invoer de verschillende stappen in het algoritme te doorlopen.[45]

Regelgebaseerde algoritmen kennen echter ook beperkingen. Van belang is dat het geprogrammeerde domein en de daarin door mensen gehanteerde kennis zich ertoe moet lenen om in 'als dit, dan dat'-regels gevat te worden.[46] Regelgebaseerde algoritmen zijn daarom niet geschikt voor toepassing in minder imperatieve besluitvormingsprocessen, waarin interpretatie is vereist of discretionaire ruimte bestaat. Daarbij kan het bijvoorbeeld gaan om de interpretatie van culturele begrippen of open normen, waarvoor menselijke inzicht van belang is. In het juridische domein kan daarbij gedacht worden aan besluiten die niet rechtstreeks volgen uit de strikte toepassing van regels, maar waarbij ook een interpretatie in het licht van de omstandigheden van een geval is vereist.[47]

Als algoritmen worden gebruikt om juridische besluitvormingsprocessen te automatiseren, dan zal de programmeur de juridische regel altijd moeten omzetten in een door computers te verwerken regel. Daarvoor moet de juridische regel worden geïnterpreteerd, ook als die op het oog duidelijk en imperatief lijkt. Een onjuiste of controversiële interpretatie kan daarbij leiden tot onwenseljike belissingen. Als algoritmen worden ingezet op grote schaal dan kunnen controversiële interpretaties, fouten of verkeerde aannames in het systeem direct grote en wijdverbreide gevolgen hebben, anders dan wanneer een individuele menselijke beslisser in afzonderlijke gevallen afzonderlijke besluiten neemt.[48]

Een ander nadeel van systemen die werken op basis van regelgebaseerde algoritmen is dat zij een statisch karakter hebben. Zij kunnen zich niet zonder tussenkomst van een programmeur

[44] Stolk, Boot & Spanninga, montesquieu-instituut.nl 26 november 2018.
[45] Voor een uitgebreider overzicht van voordelen, zie Grosan & Abraham 2011, p. 175.
[46] Zie daarover ook Grosan & Abraham 2011, p. 219.
[47] Stolk, Boot & Spanninga, montesquieu-instituut.nl 26 november 2018.
[48] Van Eck 2018, p. 40; Van Eck 2013.

aanpassen aan ontwikkelingen of zichzelf corrigeren. Als tekortkomingen niet worden opgemerkt of gecorrigeerd, dan kan de inzet van regelgebaseerde algoritmen voor langere tijd tot onjuiste of onterechte beslissingen leiden.

Ten slotte geldt dat de hoeveelheid aan variabelen en toe te passen regels een regelgebaseerd algoritme zo complex kan maken dat de uitvoer van het algoritme niet goed te voorspellen en uit te leggen is. Bovendien is het van belang dat regelgebaseerde algoritmen deel uit kunnen maken van een groter en complexer systeem waarin algoritmen met elkaar interacteren.[49] Ook dat kan het voorspellen en uitleggen van de uitvoer van regelgebaseerde algoritmen bemoeilijken.[50]

2.1.2 Zelflerende algoritmen

Zelflerende algoritmen (ook wel *machine learning*-algoritmen) kunnen uit grote gegevensverzamelingen zelf modellen afleiden die worden toegepast op een invoer.[51] Op basis van grote hoeveelheden gegevens die als trainingsdata aan het algoritme worden aangereikt, gaat het algoritme op zoek naar verbanden en patronen in die gegevens.[52] Ook is het mogelijk dat zelflerende algoritmen verder leren op basis van later ingevoerde gegevens en feedback.[53] Voor zelflerende systemen is het niet nodig dat vooraf alle relevante variabelen en regels worden geprogrammeerd. Naarmate het algoritme meer gegevens wordt aangereikt, kan het zijn taak idealiter beter en nauwkeuriger uitvoeren.[54] Hieronder bespreken we eerst de eigenschappen van zelflerende algoritmen in algemene zin. Daarna gaan we in op verschillende typen zelflerende algoritmen, te weten: *supervised learning*-, *unsupervised learning*-, *reinforcement learning*- en *deep learning*-algoritmen.

De kwaliteit en nauwkeurigheid van zelflerende algoritmen is voor een belangrijk deel afhankelijk van de beschikbaarheid van gestructureerde, hoogwaardige en representatieve gegevens die door een computer verwerkt kunnen worden. Wordt een algoritme bijvoorbeeld gevoed met gegevens die vooroordelen of ongewenste stereotypen bevatten, en wordt daarop onvoldoende getest en gecorrigeerd, dan kunnen die vooroordelen en stereotypen deel gaan uitmaken van het door het zelflerende algoritme ontwikkelde model en zich daarmee doorvertalen in de uitvoer van het algoritme.

[49] Zie par. 2.2.
[50] Voor een uitgebreider overzicht van nadelen, zie Grosan & Abraham 2011, p. 176.
[51] Larus e.a. 2018, p. 7.
[52] Surden, *Georgia State University Law Review* 2019, p. 1311; Grosan & Abraham 2011, p. 261-268.
[53] Een voorbeeld daarvan zijn spamfilters in e-mailprogramma's die 'doorleren' aan de feedback van gebruikers. Zie over het onderscheid tussen 'statische' of offline modellen en dynamische of online modellen: Datatilsynet 2018, p. 10.
[54] Surden, *Georgia State University Law Review* 2019, p. 1312.

Van belang is verder dat zelflerende algoritmen in beginsel geen causale of (juridisch) redengevende verbanden leggen, maar dat zij vooral correlaties identificeren.[55] Zelflerende algoritmen kunnen in trainingsdata verbanden vinden zonder dat daarbij sprake hoeft te zijn van een *oorzakelijk* verband. Dat is te illustreren aan de hand van een simpel voorbeeld. Een systeem zou een verband kunnen vinden tussen lichaamslengte en kaalheid. Er is in werkelijkheid echter (voor zover bekend) geen causaal verband tussen lichaamslengte en kaalheid.[56] Wel is het zo dat mannen over het algemeen langer zijn dan vrouwen en ook sneller kaal worden; dát het algoritme een verband vindt tussen lichaamslengte en kaalheid is dus niet zo gek. De toepassing van een verband dat is gevonden door een zelflerend algoritme kan op die manier leiden tot juiste uitkomsten, zonder dat daaraan causale of redengevende verbanden ten grondslag liggen. Een gevolg daarvan is dat de wijze waarop een zelflerend algoritme bepaalde ingevoerde gegevens classificeert voor mensen doorgaans niet is te doorgronden.[57] De uitvoer vanzelflerende algoritmenis dan ook inherent moeilijk uitlegbaar.

Een andere beperking van zelflerende algoritmen vloeit voort uit het feit dat zij getraind zijn met gegevens over het verleden. Als zich in de loop van de tijd nieuwe nieuwe gevalstypes voordoen, die niet of onvoldoende gerepresenteerd werden in de trainingsdata, dan zal het algoritme niet langer (alle) juiste of wenselijke verbanden leggen. Als met behulp van dergelijke algoritmen over mensen wordt beslist, dan kan dat bijvoorbeeld tot gevolg hebben dat niet iedereen op passende wijze wordt behandeld. Om zelflerende algoritmen naar behoren te laten werken kan het daarom ook nodig zijn dat zij blijven leren of worden herijkt.

Een ander probleem kan ontstaan als de uitvoer van het ene algoritme wordt ingezet als invoer in een volgend algoritmisch besluitvormingsproces; als de uitvoer die invoer beïnvloedt; of als de uitvoer wordt gebruikt om nieuwe algoritmen te trainen. Dan kan een vicieuze cirkel ontstaan waarbij het algoritme komt tot *self-fulfilling prophecies*. Een voorbeeld daarvan wordt gegeven in recent onderzoek van het Britse *Royal United Services Institute* naar de inzet van algoritmen door de politie in het kader van preventief toezicht. Uit dergelijke algoritmen kan bijvoorbeeld volgen dat in een bepaalde regio sprake is van een verhoogd risico op crimineel gedrag. Als aan de hand daarvan de aanwezigheid van politie in dat gebied verhoogd wordt, is de kans ook groter dat daar meer strafbare feiten worden geconstateerd. Als die gegevens vervolgens worden gebruikt om het algoritme verder te trainen, is de kans groter dat die regio als nóg risicovoller wordt aangemerkt en ontstaat er dus een vicieuze cirkel.[58]

[55] Pearl, *Communications of the ACM* 2019, p. 54-60. Zie over de ontwikkelingen rondom het identificeren van causale relaties aan de hand van zelflerende systemen echter ook Schölkopf 2019.
[56] Zie voor een verdere uitwerking van dit voorbeeld en voor andere voorbeelden Schoonen, Trouw.nl 28 april 2019.
[57] Ribeiro, Singh & Guestrin 2016; Hardman & Beauxis-Aussalet 2018, slide 26.
[58] Babuta & Oswald 2019, p. 12.

Zelflerende algoritmen kunnen op verschillende manieren worden getraind. Bij *supervised learning* wordt het algoritme getraind aan de hand van gelabelde voorbeelden. Dat wil zeggen dat gebruik wordt gemaakt van gegevens die al gekwalificeerd of gecategoriseerd zijn met het oog op de gewenste uitvoer. Het algoritme wordt dan geleerd hoe het met de aangereikte voorbeelden dient om te gaan. Vanwege de benodigde gelabelde gegevenssets is de kwaliteit van de uitvoer bij deze algoritmen niet alleen afhankelijk van de kwaliteit van de gegevens, maar ook van de kwaliteit van de aangebrachte labels. Een voordeel van dergelijke algoritmen is niettemin dat relatief gemakkelijk te bepalen is hoe goed het algoritme werkt, namelijk door ongelabelde gegevens in te voeren en te controleren of de uitkomst overeenstemt met de verwachting.[59]

Bij *unsupervised learning* wordt de trainingsdata niet gelabeld. Het algoritme wordt dan zo geprogrammeerd dat het na verloop van tijd autonoom structuren of patronen in de ingevoerde gegevens kan herkennen. *Unsupervised learning*-algoritmen hebben als voordeel dat zij overweg kunnen met gegevens die niet eerder gelabeld zijn. Op die manier kan een dergelijk systeem bijvoorbeeld gebruikt worden om vergelijkbare (juridische) documenten te doorzoeken en relevante patronen in de tekst te identificeren. Een lastigheid bij dit soort algoritmen is echter dat de nauwkeurigheid of juiste werking moeilijker te bepalen is.[60] Het algoritme moet namelijk zelfstandig verbanden ontdekken in de aangereikte data. De kwaliteit van de gebruikte trainingsdata en de manier waarop het algoritme met die data omgaat zijn bij *unsupervised learning*-algoritmen dus extra belangrijk. Gebruikers moeten daarom vertrouwen hebben in het algoritme en de wijze waarop het totstandkomt, of de uitvoer van die algoritmen steeds laten controleren door mensen.[61]

Reinforcement learning door algoritmen is gebaseerd op drie essentiële componenten: de agent, de omgeving en de acties. De agent is het algoritme dat getraind moet worden. De omgeving bestaat uit alle data die de agent nodig heeft om met zijn omgeving te kunnen interacteren. De acties omvatten alles wat het algoritme in zijn omgeving kan doen. Bij *reinforcement learning* navigeert de agent in de omgeving en moet het daarbij een bepaald doel bereiken, waarop een beloning (meestal in de vorm van punten) volgt. *Reinforcement learning* lost het probleem op dat het niet mogelijk of te lastig is om alle mogelijke scenario's te berekenen en te waarderen. In geval van *reinforcement learning* leert het algoritme 'spelenderwijs' welke acties lucratiever zijn in bepaalde scenario's en dus welke keuzes het beste kunnen worden gemaakt.[62]

[59] Schapire 2003.
[60] Burkov 2019.
[61] Dickson, *TechTalks* 28 augustus 2017; Grosan & Abraham 2011, p. 282.
[62] Qu, towardsdatascience.com 22 oktober 2018.

Deep learning-algoritmen zijn gebaseerd op modellen die trainingsdata, simpelgezegd, op verschillende abstractieniveaus kunnen verwerken.[63] Daarvoor wordt veelal gebruik gemaakt van neurale netwerken. Waar de eerder beschreven vormen van *machine learning* in hoge mate afhankelijk zijn van de manier waarop gegevens aan het systeem gepresenteerd worden, zijn *deep learning*-algoritmen in staat om ongestructureerde gegevens te analyseren en daarin verbanden te ontdekken.[64] Een belangrijk voordeel van *deep learning*-algoritmen is dat ze structuren kunnen ontdekken in ruwe data, terwijl andere leermethoden gestructureerde invoer nodig hebben waarbij mensen bijvoorbeeld relevante factoren dienen aan te geven. Op dit moment werken *deep-learning* technieken goed voor toepassingen zoals objectherkenning en automatisch vertalen. *Deep learning*-algoritmen zijn op dit moment echter nog niet in staat om natuurlijke taal te begrijpen. Daardoor kan dit soort algoritmen nog niet worden ingezet om bijvoorbeeld zelfstandig in een grote hoeveelheid jurisprudentie de voor het betreffende juridische probleemgebied relevante factoren te herkennen.[65]

2.2 Organisatorische en maatschappelijke context

Algoritmen staan niet op zichzelf, maar zijn ingebed in een bredere organisatorische en maatschappelijke context die hun inzet kleurt. Binnen een organisatie kunnen algoritmiten functioneren met verschillende gradaties van zelfstandigheid. Bij algoritmen met een beschrijvende of diagnostische rol wordt de uitvoer van het systeem veelal door een mens meegenomen in een verder besluitvormingsproces. In andere gevallen is menselijke tussenkomst minder vanzelfsprekend. Dat is bijvoorbeeld het geval bij voorspellende algoritmen of bij algoritmen die een bepaalde handeling voorschrijven. Naarmate de menselijke tussenkomst afneemt, neemt de rechtstreekse impact van het algoritme op rechtssubjecten vanzelfsprekend toe.[66] Binnen organisaties kan de inzet van algoritmen een belangrijke rol spelen in het beter, nauwkeuriger of sneller laten verlopen van besluitvormingsprocessen, maar daarbij is het van belang om in het oog te houden dat algoritmen ook de dynamiek van processen binnen een organisatie kunnen veranderen. De gebruikte algoritmen interacteren immers met mensen en kunnen daarmee mogelijk ook invloed hebben op het gevoel van verantwoordelijkheid, controle en invloed van medewerkers van een organisatie.[67] Het is daarom voor het inschatten van kansen en risico's van de inzet van algoritmen in een concreet geval van belang dat er nadrukkelijk aandacht is voor de aard en kenmerken van de organisatorische context waarin algoritmen worden ingezet.

[63] LeCun, Bengio & Hinton, *Nature* 2015, p. 436-444.
[64] Goodfellow, Bengio & Courville 2016, hoofdstuk 1.
[65] Wilks 2019. In de casestudy naar de rechtspraak (hoofdstuk 6) wordt hierop verder ingegaan.
[66] Bijlage bij *Kamerstukken II* 2019/20, 26643, nr. 641, p. 4.
[67] Zie over *agency* en technologie verder onder meer Rammert 2008.

Daarnaast is het belangrijk om vast te stellen dat algoritmen worden ingezet in een samenleving waarin processen steeds verder digitaliseren en systemen aan elkaar worden gekoppeld. In deze maatschappelijke context vormen zogenaamde *enabling technologies* de basis voor veel innovatieve toepassingen en nieuwe vormen van gebruik.[68] Een goed voorbeeld daarvan is het ontstaan van een *Internet of Things*, waarin niet alleen traditionele computers aan elkaar worden gekoppeld worden, maar ook tal van andere apparaten met het internet worden verbonden.[69] Daarnaast is het door toegenomen rekenkracht van computers en toegenomen mogelijkheden om grote hoeveelheden informatie op te slaan, mogelijk geworden grote (*big*) dataverzamelingen te verwerken en daarop analyses los te laten.[70] Deze ontwikkelingen maken het bovendien mogelijk om algoritmen op een zodanige wijze in te zetten dat er systemen ontstaan die taken kunnen uitvoeren die normaal gesproken een vorm van menselijke intelligentie vereisen. Er is dan sprake van (een vorm van) *Kunstmatige (of Artificiële) Intelligentie* (KI, of ook wel AI).[71]

De grotere verbondenheid van systemen, de bijbehorende interacties en de hoeveelheden data die daarmee gegenereerd en verwerkt worden, brengen nieuwe uitdagingen met zich. De context waarin systemen functioneren, wordt steeds complexer en daarmee minder overzichtelijk. Als systemen aan elkaar worden gekoppeld in een netwerk, kunnen zij van elkaar afhankelijk worden en elkaar beïnvloeden. Zo is in een netwerk van apparaten de beveiliging voor een groot deel afhankelijk van de zwakste schakel in het netwerk en bestaat er een risico op datalekken en hacks.[72] Als algoritmen met elkaar interacteren kan dit tot onvoorspelbare resultaten leiden. Bovendien kan de koppeling van systemen leiden tot zogenaamde *cascading failures*, waarbij het falen van één systeem leidt tot het falen van de gekoppelde systemen. De risico's die daarbij kunnen ontstaan, laten zich illustreren aan de hand van de *flash crashes* die zich in de financiële wereld hebben voorgedaan. Zo vond in 2010 een dergelijke crash plaats mede als gevolg van algoritmische beurshandel.[73] Om verliezen te beperken, waren de algoritmen die in de beurshandel werden ingezet geprogrammeerd om aandelen te verkopen als beurskoersen in een bepaalde mate daalden. Toen op 6 mei 2010 de beurskoersen in New York daalden, leidde dat tot grootschalige automatische verkopen. Deze verkopen leidden tot verdere koersdalingen, die weer leidden tot nieuwe automatische verkopen. De beurskoersen herstelden zich, maar wie gedurende de *flash crash* verkocht, leed grote verliezen.[74]

[68] Mitrou 2019, p. 16.
[69] Ashton, *RFID Journal* 22 juni 2009.
[70] Laney 2001.
[71] Russell & Norvig 2010, hoofdstuk 1; Surden, *Georgia State University Law Review* 2019, p. 1307.
[72] Alaba e.a., *Journal of Network and Computer Applications* 2017, p. 10-28; Weber, *Computer Law & Security Review* 2015, p. 618-627.
[73] Zie over deze gebeurtenissen ook CFTC & SEC 2010.
[74] Kirilenko e.a., *The Journal of Finance* 2017, p. 967-998.

2.3 Afsluitende opmerkingen

Algoritmen kunnen als gezegd worden onderscheiden op grond van de doelen waarvoor ze worden ingezet. Er kan in dat verband een onderscheid gemaakt worden tussen beschrijvende, diagnostische, voorspellende en voorschrijvende algoritmen. Ook verschilt de rol die algoritmen spelen bij het beslissen over personen en de wijze waarop zij dat doen. Algoritmen kunnen beslissingen nemen of ondersteunen op grond van individuele kenmerken of omstandigheden, maar zij kunnen ook door middel van profilering persoonskenmerken koppelen aan profielen op grond waarvan aannames worden gedaan over die persoon.

In dit hoofdstuk stond het verschil tussen regelgebaseerde algoritmen en zelflerende algoritmen centraal. In *regelgebaseerde* algoritmen worden de relevante variabelen en regels die de basis vormen van het algoritme van tevoren geprogrammeerd. De kwaliteit van de werking van dergelijke algoritmen hangt af van de mate waarin een op te lossen probleem kan worden omgezet in variabelen en regels. Regelgebaseerde systemen werken met name goed in relatief overzichtelijke domeinen waar het aantal mogelijke uitkomsten beperkt is. Regelgebaseerde algoritmen zijn, vanwege de voorgeprogrammeerde regels, echter niet in staat om om te gaan met onvoorziene scenario's.

Waar in regelgebaseerde algoritmen dus de vooraf bepaalde variabelen en regels centraal staan, destilleren *zelflerende algoritmen* uit aangereikte data zelf variabelen, regels en modellen. De kwaliteit van de werking van zelflerende algoritmen is dan ook voor een groot deel afhankelijk van de kwaliteit van de data die het systeem aangereikt krijgt, en de mate waarin die data representatief zijn. Een belangrijk voordeel van zelflerende algoritmen is dat zij in staat zijn om zich te blijven ontwikkelen als zij op basis van meer of andere data getraind worden. Wel geldt daarbij dat zelflerende algoritmen doorgaans getraind worden met gegevens die betrekking hebben op het verleden. Een zelflerend algoritme zal dus ontdekte verbanden in gegevens over het verleden, ook toepassen op nieuwe gevallen, waarmee het algoritme dus in zekere zin het verleden op de toekomst projecteert. Dat kan onjuiste of onwenselijke besluitvorming in de hand werken als situaties veranderen.

Regelgebaseerde algoritmen hebben, vanwege hun voorgeprogrammeerde karakter, in het algemeen een grotere voorspelbaarheid dan zelflerende algoritmen. De invoer van dezelfde invoergegevens zal immers altijd tot dezelfde uitkomsten leiden. In een individueel geval is de uitkomst van een algoritme in beginsel uitlegbaar in voor mensen te begrijpen termen. Als voldoende technische transparantie wordt geboden kan de 'redenering' van een algoritme namelijk aan de hand van ingevoerde gegevens en toepassing van gebruikte regels gevolgd worden. Zelflerende algoritmen zijn daarentegen niet gebaseerd op geprogrammeerde regels, maar leggen (doorgaans statistische) verbanden in de data, zonder dat daarbij sprake hoeft te zijn van causale

of redengevende verbanden. Dat maakt de zelflerende systemen, zeker waar gewerkt wordt met dynamische modellen (dat wil zeggen modellen die zich gedurende de inzet van het systeem kunnen ontwikkelen), minder voorspelbaar en moeilijker uitlegbaar op een voor mensen te begrijpen manier. Zelflerende systemen kunnen in technologische zin wel transparant zijn door bijvoorbeeld de broncode en invoervariabelen, parameters en drempelwaarden die gebruikt zijn inzichtelijk te maken. Het bieden van dergelijke technische transparantie kan bijdragen aan de uitlegbaarheid van de werking van het algoritme, maar zelflerende algoritmen zijn tegelijkertijd veelal dusdanig complex dat het lastig en vaak ook onmogelijk is om daaruit de gronden voor een uitkomst in een individueel geval af te leiden.[75] Er wordt op dit moment veel onderzoek verricht naar *Explainable AI* ('XAI'). Het doel van dat onderzoek is om manieren te vinden om algoritmische beslissingen, en ook de gegevens die tot die beslissingen leiden, aan eindgebruikers en andere betrokkenen in niet-technische termen uit te leggen.[76] Veelal wordt voor deze hogere uitlegbaarheid echter een prijs betaald in de vorm van een afname van de nauwkeurigheid van het systeem.[77] Op dit moment is het maken van een afweging tussen uitlegbaarheid en nauwkeurigheid van zelflerende algoritmen dus veelal nog noodzakelijk.

Voor een goed begrip van de gevolgen van de inzet van algoritmen in besluitvormingsprocessen, kan niet worden volstaan met een bespreking van de technologische onderscheidingen en technologische beperkingen van de verschillende soorten algoritmen. Belangrijk is ook dat algoritmen met verschillende gradaties van zelfstandigheid functioneren in een bredere organisatorische context. Daarnaast worden algoritmen ingezet in een steeds verder digitaliserende samenleving, waarin systemen aan elkaar gekoppeld worden en er complexe netwerken van systemen ontstaan. Dit kan ertoe leiden dat de toepassing van combinaties van op zichzelf simpele algoritmen tot zeer moeilijk uitlegbare uitkomsten leidt. Naast dergelijke potentiële complexiteit van interacties, zijn algoritmen voor hun goede werking in de gedigitaliseerde samenleving steeds vaker afhankelijk van andere systemen, waarmee eventuele fouten steeds kunnen worden doorgegeven en een trapsgewijs falen kan ontstaan. Ook dat beïnvloedt de kansen en risico's die samenhangen met de inzet van algoritmen in beslissingsprocessen.

[75] Zie daarover Bijlage bij *Kamerstukken II* 2019/20, 26643, nr. 641, p. 5.
[76] Adadi & Berrada, *IEEE Access* 2018, p. 52140.
[77] Adadi & Berrada, *IEEE Access* 2018, p. 52142 en 52145.

Hoofdstuk 3. Publieke waarden en belangen

De publieke waarden en belangen die centraal staan in dit onderzoek zijn nauw verwant aan grondrechten.[1] Vanwege het juridisch karakter van dit onderzoek worden deze publieke waarden en belangen dan ook nader geconcretiseerd aan de hand van die grondrechten. De publieke waarden en belangen die in dit hoofdstuk geïntroduceerd worden – te weten bescherming van persoonsgevens, non-discriminatie en rechtsbescherming – vormen de achtergrond waartegen in de kansen en risico's van algoritmische besluitvorming worden geïdentificeerd en de bestendigheid van het juridisch kader wordt beoordeeld.

De hier besproken publieke waarden en belangen, en hun concretisering in juridische kaders, zijn voor vrijwel alle vormen van algoritmische besluitvorming relevant. De rol en invulling van die publieke waarden verschillen afhankelijk van het domein en de dynamiek waarin algoritmen worden ingezet. De casestudy's die op de geselecteerde domeinen betrekking hebben, besteden daar verder aandacht aan.

3.1 Bescherming van persoonsgegevens

Persoonsgegevens worden, als onderdeel van tot de persoonlijke levenssfeer behorende informationele privacy,[2] beschermd door zowel grondrechtelijke normen als meer specifieke regelgeving.[3] Het belang van informationele privacy is nauw verbonden met de autonomie van het individu en hangt samen met identiteitsvorming, menselijke waardigheid en vrijheid.[4] Informationele privacy kan individuen handvatten bieden om te voorkomen dat informatie over hen wordt verzameld en om controle uit te oefenen over de informatie waartoe anderen toegang hebben.[5] Naast deze op het individu gerichte belangen, heeft de bescherming van persoonsgegevens en privacy ook bredere maatschappelijke betekenis. Zo kan het verzekeren van deze rechten bijdragen aan de vrijheid van gedachte en van meningsuiting.[6]

Een van de grondgedachten van het huidige Europese gegevensbeschermingsrecht is om burgers controle te geven over de informatie die over hen beschikbaar is. Dit heeft zich vertaald in verschillende juridische instrumenten. De belangrijkste daarvan is in de praktijk momenteel de

[1] Zie par. 1.2.2 voor onze definitie van publieke waarden en belangen.
[2] Van der Jagt 2013, p. 163.
[3] Zie ten aanzien van de grondrechten o.a.: art. 8 EVRM en EHRM (GK) 4 december 2008, ECLI:NL:XX:2008:BH1813, par. 67 (*S. en Marper/Verenigd Koninkrijk)* en artt. 7 en 8 Handvest.
[4] Mitrou 2019, p. 23.
[5] Zie voor een definitie van *informational privacy*: Koops e.a., *University of Pennsylvania Journal of International Law* 2017, p. 568.
[6] Regan, *Information, Communication & Society* 2002, p. 382-405.

Algemene Verordening Gegevensbescherming (AVG), waarmee onder meer uiting wordt gegeven aan het door art. 8 Handvest beschermde grondrecht op bescherming van persoonsgegevens. De AVG laat op bepaalde punten ruimte voor een verdere invulling door de lidstaten.[7] In Nederland is aan deze ruimte invulling gegeven in de Uitvoeringswet AVG (UAVG), waarin onder meer handhavingsmechanismen en specifieke uitzonderingsgronden zijn uitgewerkt.

De AVG is over het algemeen niet van toepassing op de verwerking van persoonsgegevens in het kader van het strafrecht. Dergelijke vormen van persoonsgegevensverwerking worden gereguleerd door de Richtlijn gegevensbescherming opsporing en vervolging.[8] In Nederland is deze Richtlijn geïmplementeerd in de Wet politiegegevens (Wpg) en de Wet justitiële en strafvorderlijke gegevens (Wjsg).[9]

Voor de bescherming van persoonsgegevens kunnen verder internationale verdragsrechtelijke regels relevant zijn, zoals art. 8 EVRM en het Databeschermingsverdrag van de Raad van Europa, evenals de invulling die daaraan is gegeven door onder meer het EHRM. Nu in de praktijk de AVG de meest concrete handvatten biedt, laten we deze verdragsrechtelijke regels en de daaraan gegeven uitleg en toepassing hierna echter buiten beschouwing en richten we ons primair op de AVG. Deze aanpak vindt ook steun in de uitspraak van de rechtbank Den Haag in de *SyRI*-zaak, waarin de inhoud van art. 8 EVRM mede uitgelegd wordt aan hand van de verwerkingsbeginselen die in de AVG zijn neergelegd.[10]

De AVG geldt (in beginsel) in alle gevallen waarin persoonsgegevens worden verwerkt, ongeacht of dat door de overheid of door private organisaties gebeurt.[11] In de AVG zijn een aantal verwerkingsbeginselen neergelegd, die kortgezegd het volgende inhouden. De verwerking van persoonsgegevens dient gebaseerd te zijn op één van de in de AVG genoemde verwerkingsgrondslagen, zoals de toestemming van de persoon op wie de gegevens slaan, of een wettelijke grondslag.[12] Verwerking moet bovendien 'behoorlijk' en 'transparant zijn'.[13] Daarnaast mogen

[7] Zie bijvoorbeeld art. 6 lid 2, art. 9 lid 4, art. 35 lid 10, art. 52 lid 5, art. 80 lid 2 en art. 87 AVG.
[8] Richtlijn (EU) 2016/680 van het Europees Parlement en de Raad van 27 april 2016 betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens door bevoegde autoriteiten met het oog op de voorkoming, het onderzoek, de opsporing en de vervolging van strafbare feiten of de tenuitvoerlegging van straffen, en betreffende het vrije verkeer van die gegevens en tot intrekking van Kaderbesluit 2008/977/JBZ van de Raad (*PbEU* 2016, L 119). Zie ook art. 3 lid 7 van de Richtlijn en de *Wegwijzer Europese richtlijn gegevensbescherming opsporing en vervolging (EU) 2016/680* van het ministerie van Justitie en Veiligheid, p. 4-5
[9] Deze wetgeving is onder meer relevant voor de bescherming van persoonsgegevens in de rechtspraak. In par. 6.3 wordt daarop verder ingegaan.
[10] Rb. Den Haag 5 februari 2020 ECLI:NL:RBDHA:2020:865 (*SyRI*).
[11] Zo moeten overheden bijvoorbeeld per definitie een functionaris voor gegevensbescherming aanstellen (art. 37 lid 1 onder a AVG). Op het moment van schrijven van dit rapport wordt door de Universiteit Tilburg een onderzoek uitgevoerd naar de horizontale aspecten van het recht op privacy, met name in het licht van technologische ontwikkelingen (WODC projectnummer 3062). Over de horizontale aspecten van het recht op privacy, zie ook Vetzo, Gerards & Nehmelman 2018, p. 77.
[12] Art. 5 lid 1 onder a en art. 6 AVG. Zie art. 7 AVG over de voorwaarden waaraan de toestemming moet voldoen. Ten aanzien van de toestemming door kinderen bevat de AVG in art. 8 een aparte bepaling.
[13] Art. 5 lid 1 onder a AVG. Zie afdeling 1 van de AVG over transparantie en afdeling 2 van de AVG over de informatie die verstrekt moet worden over de verwerking van persoonsgegevens. Art. 15 AVG beschrijft het recht op inzage.

persoonsgegevens alleen voor ‘voor welbepaalde, uitdrukkelijk omschreven en gerechtvaardigde doeleinden’ worden verzameld en mogen zij niet verder op een met die doeleinden onverenigbare wijze worden verwerkt.[14] De gegevens moeten toereikend zijn en ter zake dienend. Bovendien moeten de persoonsgegevens beperkt zijn tot wat noodzakelijk is voor de doeleinden waarvoor de gegevens worden verwerkt.[15] De gegevens moeten juist zijn en waar nodig moeten zij worden geactualiseerd.[16] Persoonsgegevens mogen ook niet langer opgeslagen worden dan nodig en dienen te worden geanonimiseerd als identificatie van de gegevens niet langer noodzakelijk is.[17] Tot slot dienen de gegevens ook op een passende wijze beveiligd te worden.[18]

Waar ‘gewone’ persoonsgegevens, dat wil zeggen tot personen herleidbare gegevens,[19] verwerkt mogen worden mits daarbij bovengenoemde verwerkingsbeginselen in acht worden genomen, geldt voor bijzondere categorieën persoonsgegevens een strenger regime. Het gaat daarbij om gegevens die samenhangen met de intieme aspecten van iemands leven, zoals ras of etnische afkomst, seksueel gedrag of seksuele gerichtheid, politieke opvattingen, religieuze of levensbeschouwelijke overtuigingen, etc.[20] Dergelijke gegevens mogen niet verwerkt worden, tenzij een uitdrukkelijk in de AVG genoemde uitzondering van toepassing is en verwerking plaatsvindt in overeenstemming met de hierboven beschreven verwerkingsbeginselen.

Tot slot is art. 22 AVG in het kader van dit onderzoek van belang. Deze bepaling houdt een ‘in-beginselverbod’ in op volledig geautomatiseerde besluitvorming. Ook profilering, waarbij persoonlijke kenmerken van een natuurlijke persoon door algoritmen worden geanalyseerd of voorspeld, valt binnen de reikwijdte van art. 22 AVG.[21] Voor toepasselijkheid van de bepaling is dan wel vereist dat daaraan voor een persoon rechtsgevolgen zijn verbonden of dat het een persoon op aanmerkelijke wijze treft. De AVG bevat daarnaast een recht op informatie met betrekking tot de wijze waarop dergelijke besluitvorming tot stand komt. Deze bepalingen geven uiting aan het zelfstandige belang dat personen hebben om geïnformeerd te worden over de wijze waarop hen betreffende gegevens worden verwerkt. Omdat ook de rechtsbescherming is gediend met dergelijke informatieverstrekking, bespreken wij deze bepalingen in de paragraaf over rechtsbescherming (par. 3.3).

[14] Art. 5 lid 1 onder b AVG.
[15] Art. 5 lid 1 onder c AVG.
[16] Art. 5 lid 1 onder d AVG. Zie ook afdeling 3 AVG over het recht op rectificatie en wissing van persoonsgegevens.
[17] Art. 5 lid 1 onder e AVG.
[18] Art. 5 lid 1 onder f AVG.
[19] Art. 4 onder 1 AVG.
[20] Zie art. 9 AVG voor een volledig overzicht.
[21] Profilering wordt in art. 4 lid AVG gedefinieerd als: ‘elke vorm van geautomatiseerde verwerking van persoonsgegevens waarbij aan de hand van persoonsgegevens bepaalde persoonlijke aspecten van een natuurlijke persoon worden geëvalueerd, met name met de bedoeling zijn beroepsprestaties, economische situatie, gezondheid, persoonlijke voorkeuren, interesses, betrouwbaarheid, gedrag, locatie of verplaatsingen te analyseren of te voorspellen.’

3.2 Non-discriminatie

Het recht op non-discriminatie houdt in de kern in dat voorkomen moet worden dat gelijke gevallen ongelijk behandeld worden, of ongelijke gevallen gelijk worden behandeld, als daarvoor geen goede rechtvaardiging bestaat.[22] Van een verdenking van het ontbreken van een objectieve rechtvaardiging is al snel sprake als een ongelijke behandeling is gebaseerd op gronden zoals ras, geslacht, seksuele geaardheid, leeftijd, handicap of geloofsovertuiging. Die gronden worden dan ook vaak als 'verdachte' gronden van onderscheid aangemerkt. Die verdachtheid komt eruit voort dat benadeling op een van deze gronden veelal nauw samenhangt met ongewenst geachte of al te brede stereotypen of vooroordelen, of met ideeën over inferioriteit van bepaalde groepen of kenmerken. Ook kan de verdachtheid samenhangen met een geschiedenis van stigmatisering en buitensluiting van groepen met bepaalde persoonskenmerken.

Zulke verdachte gronden zijn veelal expliciet als 'verboden gronden' van onderscheid opgenomen in codificaties van het discriminatieverbod of als verdacht aangemerkt in de rechtspraak van bijvoorbeeld het EHRM.[23] Als een onderscheid rechtstreeks is ingegeven door een van deze gronden, en hiervoor geen rechtvaardiging is te geven, dan is er sprake van *directe* discriminatie. Van *indirecte* discriminatie is sprake als een ongelijke behandeling niet direct is gebaseerd op een verdachte of verboden grond, maar die wel het effect heeft dat leden van een groep die worden gekenmerkt door een 'verdachte' grond in het bijzonder benadeeld worden. Waar bij directe vormen van onderscheid het aantal rechtvaardigingsgronden veelal heel specifiek en limitatief is opgesomd, zijn de mogelijkheden voor rechtvaardiging bij indirect onderscheid meestal ruimer geformuleerd.[24]

Verschillende codificaties van het discriminatieverbod hebben alleen gelding in verhoudingen tussen de overheid en burgers, zoals de codificaties in internationale verdragen en de Nederlandse Grondwet. Zij kunnen niettemin via het leerstuk van de positieve verplichtingen of door invulling van open normen van privaatrecht doorwerken in onderlinge relaties van burgers.[25] Door middel van codificatie in specifieke wet- en regelgeving is het recht op non-discriminatie daarnaast uitdrukkelijk van toepassing verklaard in horizontale verhoudingen. Zo verbiedt de AWGB directe en indirecte discriminatie op de werkvloer en bij dienstverlening op een aantal limitatief opgesomde gronden en bestaat er specifieke wetgeving ten aanzien van de gelijke behandeling van mannen

[22] Vetzo & Gerards, *Computerrecht* 2019, p. 14; Vetzo, Gerards & Nehmelman 2018, p. 81.
[23] Zie bijvoorbeeld EHRM 15 september 2016, ECLI:CE:ECHR:2016:0915JUD004481811, *EHRC* 2016/101, m.nt. J.H. Gerards (*British Gurkha Welfare Society e.a./Verenigd Koninkrijk*). Zie voor de oorsprong van de term 'verdacht' in relatie tot gronden voor onderscheid, en voor meer informatie ten aanzien van welke gronden als verdacht worden aangemerkt Vetzo, Gerards & Nehmelman 2018, p. 83-84.
[24] Zie hierover verder Vetzo, Gerards & Nehmelman 2018, p. 85 e.v.; Waaldijk 2005; Gerards 2018; Holtmaat & Rodrigues 2015, p. 144.
[25] Uit art. 14 EVRM kan bijvoorbeeld een positieve verplichting voortvloeien om wetgeving in het leven te roepen die beschermt tegen discriminatie door werkgevers (EHRM 6 november 2012, ECLI:CE:ECHR:2012:1106JUD004733506 (*Redfearn/Verenigd Koninkrijk*)). Zie voor meer daarover: Besson, *Human Rights Law Review* 2008, p. 660; Vetzo, Gerards & Nehmelman 2018, p. 89.

en vrouwen, gehandicapten en chronisch zieken, en op grond van leeftijd en arbeidsduur.[26] Deze Nederlandse wetgeving dient in de meeste gevallen ook als implementatie van relevante EU-richtlijnen hierover, vooral waar het het terrein van de arbeid betreft, maar ook waar het gaat om bepaalde aspecten van sociale zekerheid en sociale voorzieningen.

3.3 Rechtsbescherming

In dit rapport hanteren wij een ruime definitie van de publieke waarde van rechtsbescherming. Daarbij gaat het om alle juridische instrumenten die aan iemand ter beschikking staan die in zijn belangen wordt geraakt als gevolg van een besluitvormingsproces waarin algoritmen zijn gebruikt. Rechtsbescherming speelt daarbij ook een sleutelrol in het realiseren en beschermen van andere publieke waarden en belangen. Door toegang tot geschilbeslechting kunnen individuen immers hun (grond)rechten effectueren als die (door middel van) algoritmische besluitvorming kunnen worden aangetast.[27]

De ruime uitleg die wij geven aan de publieke waarde van rechtsbescherming omvat zowel het recht op een effectief rechtsmiddel als het recht op een eerlijk proces, alsmede de eisen die in dat kader aan een rechter en een rechterlijke procedure kunnen worden gesteld. Het recht op een eerlijk proces komt het meest prangend naar voren als algoritmen worden ingezet in de rechtspraak.[28]

Art. 13 EVRM bepaalt dat eenieder wiens rechten uit het EVRM geschonden zijn, toegang moet hebben tot een effectief rechtsmiddel. Dit recht is ook erkend in het Unierecht, eerst als algemeen beginsel van Unierecht, en later in art. 47 Handvest.[29]

De ‘effectiviteit’ van het rechtsmiddel hangt sterk samen met de mogelijkheden tot rechtsherstel. In de context van de inzet van algoritmen moet daarbij vooral worden gedacht aan de mogelijkheden van herstel van de situatie die rechtens is in het geval van een onjuist of onterecht genomen besluit. Te denken valt aan ongedaanmaking van een besluit, maar ook aan de mogelijkheid van reparatie, compensatie van nadeel of vergoeding van schade die het gevolg is van een algoritmisch besluitvormingsproces.

Het recht op een effectief rechtsmiddel omvat verder een recht op een gemotiveerde en bindende beslissing in de individuele zaak.[30] De woordkeuze ‘rechtsmiddel’, in plaats van ‘toegang tot een

[26] Wet gelijke behandeling mannen en vrouwen; Wet gelijke behandeling op grond van handicap of chronische ziekte; Wet gelijke behandeling op grond van leeftijd bij arbeid; Wet onderscheid arbeidsduur.
[27] Vetzo, Gerards & Nehmelman 2018, p. 166.
[28] Zie de casestudy naar de rechtspraak, in het bijzonder par. 6.5.
[29] Widdershoven 2011, p. 103.
[30] Council of Europe 2018, p. 24.

rechter', houdt de mogelijkheid open dat kan worden volstaan met alternatieve vormen van conflictbeslechting zoals mediation.[31] Het recht op een effectief rechtsmiddel vereist niettemin steeds dat er procedures beschikbaar zijn en dat deze voldoende toegankelijk zijn. Ook moet de procedure zelf voldoen aan eisen, zoals een vereiste van voldoende snelheid in de beoordeling en vereisten ten aanzien van de eerlijkheid van de procedure.

3.4 Conclusie

In dit hoofdstuk hebben wij een overzicht gegeven van drie publieke waarden en belangen die in veel gevallen van algoritmische besluitvorming relevant zullen zijn. Het recht op gegevensbescherming vergt dat de gegevens die voor de ontwikkeling en toepassing van algoritmen nodig zijn, zorgvuldig en in overeenstemming met de gestelde wettelijke vereisten worden verwerkt. Het recht op gegevensbescherming is bij de inzet van algoritmen relevant omdat voor de ontwikkeling en de toepassing daarvan grote hoeveelheden persoonsgegevens nodig kunnen zijn.

Het recht op non-discriminatie vereist dat de personen wier belangen door algoritmische besluitvorming worden geraakt op een gelijkwaardige manier worden behandeld; een benadeling van personen en groepen is alleen aanvaardbaar als daarvoor een objectieve en redelijke rechtvaardiging bestaat. De inzet van algoritmen mag dan ook geen ongerechtvaardigd (indirect) onderscheid opleveren.

Het recht op rechtsbescherming garandeert dat de rechten van betrokkenen en de daarbij behorende waarborgen in de praktijk verwezenlijkt kunnen worden. Om het recht op rechtsbescherming in de praktijk te garanderen is onder meer nodig dat de personen wier belangen door algoritmische besluitvorming aangetast worden, weten tot wie zij zich kunnen wenden. Daarnaast vergt het recht op rechtsbescherming dat betrokkenen voldoende informatie over de werking van het algoritme hebben om aan te kunnen tonen dat zij op onrechtmatige wijze in hun belangen zijn geraakt.

[31] Vetzo, Gerards & Nehmelman 2018, p. 112.

Hoofdstuk 4. Casestudy Contentmoderatie door online platformen

Stefan Kulk & Thom Snijders

4.1 Introductie

Het internet biedt talloze nieuwe mogelijkheden waardoor individuen online met elkaar kunnen interacteren en communiceren. Online dienstverleners zoals Facebook, Youtube en Twitter bieden mensen een platform om zichzelf te uiten en informatie met elkaar te delen. De platformen die worden geboden kunnen echter ook gebruikt worden voor het verspreiden van content die om tal van redenen als ongewenst kan worden beschouwd door mensen, de dienstverlener en de maatschappij in brede zin. Te denken valt aan het verspreiden van desinformatie, het doen van discriminerende uitingen en het delen van terroristische propaganda. Maar bijvoorbeeld ook de verspreiding van naaktfoto's kan door bepaalde platformen als ongewenst worden bestempeld.

Online platformen kunnen een belangrijke rol spelen in het voorkomen en stoppen van de verspreiding van dergelijke content. Zij stellen regels op en nemen beslissingen over de wijze waarop bepaalde content wordt weergegeven en bepalen of content minder prominent zichtbaar moet zijn, verwijderd dient te worden of überhaupt online komt. Dit proces wordt ook wel contentmoderatie genoemd.[1] Voor het modereren van content wordt door platformen in toenemende mate gebruik gemaakt van algoritmen.

Het modereren van online content door online platformen is het onderwerp van deze casestudy. Daarin schenken we in het bijzonder aandacht aan het modereren van *hate speech*, waaronder ook discriminerende uitingen. Omdat het vaststellen van de onrechtmatigheid van dergelijke content afhankelijk is van feiten en omstandigheden, en algoritmen maar in beperkte mate rekenschap kunnen geven van de context waarin een uiting wordt gedaan, kan met een nadere bestudering van het modereren van *hate speech* worden blootgelegd wat de mogelijkheden en onmogelijkheden zijn van algoritmen in dit domein.

4.1.1 Methodologie

Het is allereerst van belang om op te merken dat het slechts beperkt mogelijk is om een volledig beeld te schetsen van de wijze waarop online platformen content modereren. De technologieën die platformen inzetten om content te modereren zijn in private handen. Platformen zijn in beginsel

[1] Contentmoderatie is geen nieuw fenomeen. De noodzaak van het maken van keuzes met betrekking tot welke content wordt getoond aan gebruikers of consumenten is ouder dan de hedendaagse online platformen, en in zekere zin zelfs ouder dan het internet; kranten, televisiezenders en andere traditionele media hebben allemaal te maken met het cureren van hun aanbod. Zie Gillespie 2018, p. 74. Contentmoderatie wordt in deze casestudy in meer detail uitgelegd in par. 4.1.3; voor een wetenschappelijke definitie van *content moderation*, zie bijvoorbeeld Roberts 2019.

niet verplicht om openheid van zaken te geven met betrekking tot de werking van hun technologieën en de inzet van algoritmen om content te modereren. Als het gaat om de werking en toepassing van algoritmen voor contentmoderatie baseren we ons daarom enerzijds op de publieke (journalistieke) informatie over het modereren van content door platformen en anderzijds op wetenschappelijk onderzoek naar de wijze waarop algoritmen ingezet *kunnen* worden.

Om vergaarde informatie aan te vullen en verkregen inzichten te verifiëren, zijn semigestructureerde interviews afgenomen. Meer in het bijzonder was het doel van de interviews om een beter beeld te krijgen van de werking en het gebruik van contentmodereeralgoritmen, evenals de context waarin ze worden gebruikt. Daarnaast is verkregen input gebruikt ten behoeve van de inventarisatie van de kansen en risico's die zich kunnen voordoen met betrekking tot de geïdentificeerde publieke waarden en de houdbaarheid van het juridisch kader.[2]

4.1.2 Opzet van de casestudy

In deze casestudy bespreken we eerst, in algemene zin, hoe algoritmen kunnen worden gebruikt voor het modereren van onrechtmatige en onwenselijke online content (par. 4.1.3). Vervolgens spitsen we de casestudy toe op het fenomeen *hate speech* en bespreken we de algoritmen die platformen inzetten om de verspreiding van *hate speech* tegen te gaan (par. 4.2). Daarna worden per publieke waarde de kansen en risico's van het gebruik van dergelijke algoritmen geïnventariseerd en wordt het relevante juridisch kader in kaart gebracht en geëvalueerd (par. 4.3 t/m 4.6). In tegenstelling tot andere casestudy's bespreken wij daar niet eerst de casestudy-overstijgende waarden, maar vangt onze analyse aan met een bespreking van de casestudyspecifieke publieke waarde van vrijheid van meningsuiting. Wij hebben daarvoor gekozen omdat het modereren van online content in de kern vooral deze publieke waarde raakt.

4.1.3. Contentmoderatie door algoritmen

Een scherp afgebakende definitie van 'online content' is moeilijk te geven. Wij verstaan daaronder iedere tekstuele, visuele of hoorbare inhoud die online beschikbaar wordt gesteld. Te denken valt aan berichten, foto's, muziek en alle andere inhoud die door gebruikers op online platformen worden geplaatst. Ook reacties op nieuwsberichten, Facebook- of Instagramposts of tweets zijn 'content'. Advertenties die op platformen worden geplaatst kunnen ook als content worden aangemerkt. Hieronder bespreken we wanneer dergelijke content wordt aangemerkt als onrechtmatig of onwenselijk, leggen we uit wat contentmoderatie inhoudt en gaan we in op de wijze waarop algoritmen daarbij ingezet kunnen worden.

[2] In deze casestudy wordt daarom niet verwezen naar de interviews en de daarin verkregen informatie en door de geïnterviewden naar voren gebrachte zienswijzen.

4.1.3.1 Onrechtmatige en onwenselijke online content

Er is een breed scala aan typen onrechtmatige en onwenselijke online content die met behulp van algoritmen kunnen worden gemodereerd. Zij kunnen langs tenminste drie lijnen worden onderscheiden.

In de eerste plaats dient er een onderscheid gemaakt te worden tussen onrechtmatige content en ongewenste content. In het geval van ongewenste content is er geen sprake van overtreding van een juridische norm door het maken of verspreiden van de content. Te denken valt aan online desinformatie, waarvan de creatie en verspreiding een verstorend effect kan hebben op bijvoorbeeld democratische processen, maar die niet per definitie onrechtmatig is.[3] Ook kan worden gedacht aan content die door online platformen als ongewenst wordt aangemerkt en die verwijderd kan worden op grond van de beleidsregels van het platform, maar die niet onrechtmatig is, zoals naaktbeelden.[4] Het zijn dan de opvattingen en denkbeelden van de platformbeheerder die de doorslag geven.

In de tweede plaats kan er ten aanzien van onrechtmatige content een onderscheid worden gemaakt naar het karakter van de onrechtmatigheid. Zo is er content die naar zijn aard onrechtmatig is. Voorbeelden zijn kinderpornografisch materiaal en content waarin aangezet wordt tot haat of geweld, of waarin een groep wordt beledigd.[5] Daarnaast is er een categorie van content die op zichzelf niet onrechtmatig is, maar waarbij de onrechtmatigheid is gelegen in de verspreiding ervan en de voorwaarden waaronder dat gebeurt. Voorbeelden zijn de ongeautoriseerde openbaarmaking van auteursrechtelijk beschermd werk of privacyinbreuken. De content in kwestie (bijvoorbeeld een film of een muziekstuk, of een nietsverhullende foto) zelf is dan niet onrechtmatig, maar de handeling om die zonder toestemming openbaar te maken wel.[6]

In de derde plaats kan er onderscheid gemaakt worden naargelang het effect dat de content heeft. Een inbreuk op een intellectueel eigendomsrecht zal veelal economische schade veroorzaken.[7]

---

[3] De High-Level Expert Group on fake news and online disinformation zegt hierover het volgende: '*Disinformation as defined here includes forms of speech that fall outside already illegal forms of speech, notably defamation, hate speech, incitement to violence, etc. but can nonetheless be harmful*.' HLEG on Fake News and Disinformation 2018, p. 10. Zie in dat verband ook de uitspraak van het EHRM in de zaak *Salov*, waarin is bepaald dat ook het verspreiden van informatie, waarvan een sterk vermoeden bestaat dat die niet waar is, onder de bescherming van art. 10 EVRM valt: EHRM 6 december 2005, ECLI:CE:ECHR:2005:0906JUD006551801 (*Salov/Oekraïne*).

[4] Het delen van dergelijke content zou echter wel in strijd kunnen zijn met contractuele voorwaarden waaraan gebruikers van platformen zich hebben verbonden.

[5] Art. 240b Sr verbiedt bijvoorbeeld de vervaardiging en verspreiding van kinderpornografisch materiaal. Art. 137d Sr verbiedt het aanzetten tot haat of geweld en art. 137c Sr stelt groepsbelediging strafbaar. Daarnaast kan in het kader van contentmoderatie ook het strafrechtelijke verbod van art. 137e Sr relevant zijn, dat ziet op de openbaarmaking van uitlatingen in de zin van art. 137c juncto art. 137d Sr, en art. 137f Sr, dat onder andere deelname aan een activiteit gericht op discriminatie strafbaar stelt.

[6] Hoewel er ten aanzien van dit soort content strikt genomen geen sprake is van onrechtmatige content, spreken we uit overwegingen van de leesbaarheid van deze casestudy wel van onrechtmatige content.

[7] Persoonlijkheidsrechten kunnen echter wel een rol spelen.

Voor content zoals 'wraakporno'[8] of discriminerende uitingen is vooral de menselijke waardigheid in het geding en zullen economische overwegingen een minder grote rol spelen. Een tussenvorm is wellicht de schending van portretrechten waarbij zowel economische belangen als privacybelangen een rol kunnen spelen.[9] Daarnaast zijn er typen content die de samenleving als geheel kunnen raken en waarvan de verspreiding dus om die reden onwenselijk wordt geacht. Daartoe behoren bijvoorbeeld terroristische inhoud of online desinformatie, maar ook discriminerende uitingen.

4.1.3.2 Contentmoderatie

Ten aanzien van content die is aangemerkt als onrechtmatig, ligt het verwijderen daarvan voor de hand. Als een rechter heeft bepaald dat, bijvoorbeeld, een post op Facebook onrechtmatig is, dan dient die post verwijderd te worden. Contentmoderatie is echter een proces dat doorgaans geheel plaatsvindt bij het platform zelf. Het platform ontvangt meldingen over, of gaat zelf op zoek naar, onrechtmatige of onwenselijke content. Het platform beoordeelt de aangebrachte of gevonden content zelf, en neemt vervolgens actie waar nodig. Van rechterlijke tussenkomst is zeer zelden sprake.

De laatste jaren is er vanuit verschillende hoeken aandacht voor de verantwoordelijkheid van online platformen om de verspreiding van onrechtmatige en ongewenste content op hun platformen tegen te gaan.[10] Het kan daarbij gaan om het (snel) verwijderen van content, maar ook het deprioriteren (lager *ranken*) daarvan.[11] Beheerders van platformen kunnen er ook zelf belang bij hebben om bepaalde vormen van onrechtmatige of onwenselijke content te weren, bijvoorbeeld om gebruikers en adverteerders aan zich gebonden te houden.[12] Om onrechtmatige en onwenselijke content te vinden, wordt tegenwoordig lang niet meer alleen gebruik gemaakt van

[8] Het begrip 'wraakporno' is een misleidende term omdat het eigenlijk gaat om de ongeautoriseerde verspreiding van beelden van seksuele aard. Wraak hoeft daarin geen rol te spelen. Bovendien wordt door het begrip 'wraak' gesuggereerd dan het verspreiden van de beelden in zekere zin te billijken zou zijn. Meer daarover Sebastian, *Feminist Media Studies* 2017, p. 1107.

[9] Zie voor een recent voorbeeld daarvan in de Nederlandse rechtspraak: Rb. 11 november 2019, ECLI:NL:RBAMS:2019:8415 (*John de Mol/Facebook*), over de 'nepadvertenties' voor cryptovaluta waarin beelden van John de Mol werden gebruikt.

[10] Zie bijvoorbeeld Chesney & Citron, *Foreign Affairs* 2019, p. 147.

[11] Het ranken van informatie als zodanig valt buiten het bereik van deze casestudy. Zie daarover onder meer de Platform-to-business Verordening waarin in art. 5 een verplichting tot verstrekking van informatie over het ranken van informatie door zoekmachine en e-commerce platformen is opgenomen (Verordening (EU) 2019/1150 van het Europees Parlement en de Raad van 20 juni 2019 ter bevordering van billijkheid en transparantie voor zakelijke gebruikers van onlinetussenhandelsdiensten (Voor de EER relevante tekst), *PbEU* 2019, L 186). Zie ook het werk van het Observatory on the Online Platform Economy (platformobservatory.eu). Zie ook het burgerinitiatief 'internetpesters aangepakt' en de Kamerbrief daarover van 17 juli 2019 (*Kamerstukken II* 2018/19, 34602, nr. 2).

[12] Bijvoorbeeld Van Noort, *NRC* 13 februari 2018; Dang, *Reuters* 21 februari 2019; Harding, *CBS NEWS* 11 mei 2018; 'Mars pulls ads from YouTube drill videos', BBC.com 4 augustus 2018; zie met betrekking tot Dumpert, het videoplatform van GeenStijl: 'Adverteerders stoppen met adverteren op Dumpert en GeenStijl', *NOS*.nl 3 mei 2017.

meldingen van mensen.[13] Om proactief te kunnen optreden zetten platformen ook algoritmen in om deze typen content te detecteren zodra gebruikers die online beschikbaar willen stellen.[14]

In zowel het privaatrecht als het publiekrecht zijn er normen die ertoe strekken dat platformen bepaalde onrechtmatige content verwijderen en, afhankelijk van de content in kwestie, ook zelf opsporen. Het aansprakelijkheidsrecht, met name de aansprakelijkheidsbeperkingen in art. 6:196c BW, stimuleren platformen om onrechtmatige content te verwijderen zodra zij daarvan kennis hebben.[15] Maar ook in het publiekrecht zien we terug dat van platformen wordt verlangd dat zij content modereren. Een voorbeeld is het conceptvoorstel voor een EU-verordening ter voorkoming van de verspreiding van terroristische online-inhoud, dat platformen verplicht terroristische content binnen een uur na melding van een autoriteit te verwijderen en verwijderd te houden.[16]

Platformen kunnen ook zelf op grond van een gebruikersovereenkomst (*terms of service*) paal en perk stellen aan de verspreiding van onwenselijke of onrechtmatige content op het platform. Deze overeenkomsten stellen online platformen in staat om content van gebruikers te verwijderen of de toegang daartoe te beperken als die de huisregels van het platform schenden.[17] De typen content die door platformen niet zijn toegestaan komen in grote lijnen overeen met de typen content die ook juridisch gezien onrechtmatig zijn. Als private ondernemingen staat het platformen echter vrij om ook strengere regels te hanteren en content die op zichzelf niet onrechtmatig is, zoals naaktbeelden, van het platform te weren.

4.1.3.3 Inzet van algoritmen

Verschillende typen onrechtmatige en ongewenste content vragen om een eigen aanpak. De rol die platformen daarin (kunnen) spelen en de wijze waarop algoritmen worden ingezet, verschilt daarom ook per type content.

Algoritmen kunnen een rol spelen bij het herkennen van onrechtmatige of ongewenste content. Zo kan een algoritme door tekstanalyse herkennen of er in een bepaald geval mogelijk sprake is van het aanzetten tot haat, geweld of discriminatie. Die content kan dan worden doorgeleid naar een medewerker van het platform, die er een definitief oordeel over velt. Het is ook denkbaar dat een algoritme detecteert dat een gebruiker onrechtmatige content probeert te delen en zelf besluit de

[13] Ook partijen die klagen over content bij platformen kunnen zelf algoritmen gebruiken. Zo is bekend dat rechthebbenden algoritmen inzetten voor het vinden van auteursrechtinbreuken op internet (Urban, Karaganis & Schofield 2016). Een ander voorbeeld is de deels geautomatiseerde detectie van *hate speech* door 'Hatebusters'. Dit programma legt gebruikers reacties op YouTube voor die mogelijk *hate speech* bevatten zodat zij die reacties kunnen aanbrengen bij Youtube (hatebusters.org).
[14] Gillespie 2018, p. 77.
[15] Deze bepaling implementeert artt. 12-14 van de E-Commercerichtlijn (Richtlijn 2000/31/EG van het Europees Parlement en de Raad van 8 juni 2000 betreffende bepaalde juridische aspecten van de diensten van de informatiemaatschappij, met name de elektronische handel, in de interne markt ("Richtlijn inzake elektronische handel"), *PbEG* 2000, L 178).
[16] Zie voor het wetgevingsdossier met betrekking tot deze verordening 2018/0331(COD).
[17] Zie voor een voorbeeld van een dergelijke contractuele bepaling punt 2 van de Facebook servicevoorwaarden (perma.cc/3H7P-BQXS). Voor een voorbeeld van dergelijke huisregels zie facebook.com/communitystandards.

content te weren, zoals reeds gebeurt met betrekking tot auteursrechtelijk beschermde werken.[18] Op basis van een *hash* - een digitale vingerafdruk die wordt gemaakt van een beschermd werk - kan dat werk door een algoritme worden herkend en kan worden voorkomen dat bepaalde content online verschijnt. Een gevaar is dat dergelijke algoritmen in geüpload materiaal ten onrechte een auteursrechtelijk beschermd werk herkennen.[19] Maar ook als materiaal correct wordt herkend, is nog niet gezegd dat dat materiaal ook daadwerkelijk inbreukmakend is. Zo staat het auteursrecht gebruik van werken toe als er bijvoorbeeld sprake is van een parodie of een citaat. Of daarvan sprake is, vergt soms een complexe juridische afweging, die (vooralsnog) niet door het algoritme gemaakt kan worden.

Algoritmen die *hashes* vergelijken kunnen ook worden ingezet om te voorkomen dat reeds als onrechtmatig aangemerkte content zich verder verspreidt. Te denken valt aan kinderpornografie of terroristische content. Ook van dergelijke content kan een *hash* gemaakt worden die door een algoritme kan worden vergeleken met nieuw geüploade content.[20]

Andere algoritmen worden ingezet om de authenticiteit van content vast te stellen en worden zodoende gebruikt om bepaalde vormen van online desinformatie te detecteren. Een voorbeeld van dergelijke content betreft zogeheten '*deep fakes*': gemanipuleerde video's die (bijna) niet van echt zijn te onderscheiden en het publiek in verwarring kunnen brengen.[21] Voorbeelden zijn video's waarin wereldleiders woorden in de mond gelegd worden die zij niet hebben gezegd.[22] Zelflerende algoritmen kunnen worden ingezet om op basis van gelabelde datasets van video's, gemanipuleerde video's te leren herkennen.[23] Een andere methode kan zijn om video's op bepaalde veelvoorkomende eigenschappen van *deep fakes*, zoals de lage resolutie van gemanipuleerde beelden, met behulp van neurale netwerken te ontdekken.[24] Met betrekking tot desinformatie kunnen algoritmen ook worden ingezet om links naar gefactcheckte artikelen te vinden, zodat automatisch een factchecklabel bij berichten kan worden geplaatst.[25]

[18] Google heeft daarvoor 'ContentID' ontwikkeld. Het systeem van Facebook heet 'Rights Manager'. Beide systemen checken geüpload materiaal tegen een database van beschermde werken.
[19] Lester & Pachamanova, *UCLA Entertainment Law Review* 2017, p. 51-73. Zie daarover ook Kulk 2019, p. 280.
[20] Zie bijvoorbeeld gifct.org/joint-tech-innovation/, waarin Facebook, Microsoft, Twitter, and YouTube samenwerken om de verspreiding van terroristische content te voorkomen en een '*database of hashes*' onderhouden waarmee wordt voorkomen dat bepaalde content opnieuw op een van de platformen verschijnt. Zie ook 'Open-Sourcing Photo- and Video-Matching Technology to Make the Internet Safer', fb.com 1 augustus 2019, waarin Facebook open source software beschikbaar stelt om identieke of bijna identieke beelden te herkennen.
[21] Voorbeelden zijn video's waarin het gezicht of alleen de mond en lippen van iemand zijn vervangen. De term is een porte-manteau van de termen '*deep learning*' en '*fake*'.
[22] Zie over de opkomst en gevaren van *deep fakes* in geopolitieke context: Citron & Norton, *Boston University Law Review* 2011, p. 1435.
[23] Zie voor een dergelijke dataset: Deepfake Detection Challenge, deepfakedetectionchallenge.ai. Zie over deze dataset: Dolhansky e.a. 2019. Zie ook over de FaceForensics Dataset: Rössler e.a. 2018.
[24] Zie bijvoorbeeld Li & Lyu 2019. Voor een overzicht van verschillende methoden om *deep fakes* te herkennen, zie Nguyen e.a. 2019.
[25] Zie over zulke labels bijvoorbeeld Gibbs, *The Guardian* 7 april 2017.

In alle gevallen is het proportionaliteitsvereiste van belang voor de invulling van contentmoderatie en de rol die algoritmen daarin kunnen spelen. De aanpak die wordt gekozen, moet in verhouding staan tot onder andere de effecten van de content in kwestie en de gevolgen van de inzet van algoritmen. Daarnaast is van belang dat het modereren van content slechts één middel is om de verspreiding van onrechtmatige en ongewenste content tegen te gaan. Het bestrijden van online desinformatie is daarvan een goed voorbeeld. De aanpak van online desinformatie vergt een combinatie van maatregelen, zoals het bevorderen van mediawijsheid, de inzet van *factcheckers*, en het verzekeren van de pluriformiteit van de media.[26] Ook discriminerende uitingen kunnen dichter bij de bron worden voorkomen, bijvoorbeeld door middel van educatieve campagnes.[27]

Het modereren van content door middel van algoritmen hoeft bovendien niet altijd te leiden tot het verwijderen van content. Om de negatieve effecten van bijvoorbeeld online desinformatie te mitigeren, zetten platformen algoritmen in om de betrouwbaarheid van bepaalde content te wegen en wordt onbetrouwbare informatie een lagere positie gegeven in de ordening van de content (*ranking*).[28] Het modereren van online content is dus meer dan alleen het verwijderen van onrechtmatige of onwenselijke content.

4.2 De aanpak van *hate speech* door online platformen

In deze casestudy gaan we nader in op de inzet van algoritmen voor de aanpak van online *hate speech*. Daarvoor zal allereerst het fenomeen van online *hate speech* kort worden besproken. Daarna bespreken we de werking en toepassing van de verschillende soorten algoritmen die worden ingezet om *hate speech* te detecteren. Ook gaan we in op hoe het gebruik van algoritmen zich in dit domein in de toekomst zou kunnen ontwikkelen.

4.2.1 Hate speech

De term *hate speech*, die oorspronkelijk uit de Verenigde Staten afkomstig is, kan het best worden beschouwd als een verzamelbegrip dat wordt gebruikt om (onderdelen van) een spectrum van schadelijke of anderszins onwenselijke uitingen aan te duiden. Daaronder vallen het oproepen tot geweld, haatdragende en haatzaaiende uitingen, maar mogelijk ook andere zeer beledigende uitingen en uitingen die getuigen van extreme vooroordelen en/of vooringenomenheid.[29] Het gaat daarbij onder andere om openbare op schrift gestelde of door middel van afbeelding gedane uitingen. Over zowel de precieze reikwijdte van de term *hate speech* als het onderscheidende

[26] HLEG on Fake News and Disinformation 2018, p. 35. Zie ook McGonagle e.a. 2018.
[27] Zie Titley, Keen & Földi 2014, p. 41.
[28] Bijvoorbeeld Google, over de inzet van algoritmen voor het herkennen van desinformatie ten aanzien van zijn producten 'How Google Fights Disinformation', Google februari 2019, blog.google/documents/37/How_Google_Fights_Disinformation.pdf, p. 4.
[29] McGonagle 2013, p. 4. Over de oorsprong van de term *hate speech*, zie Brown, *Law and Philosophy* 2017, p. 424.

karakter ervan bestaan uiteenlopende opvattingen.[30] Volgens sommigen kenmerkt *hate speech* zich in het bijzonder door een intentie om de (gelijk)waardigheid van bepaalde groepen personen - en daarmee de maatschappelijke acceptatie - te ondermijnen.[31] Voor anderen is vooral het extreme karakter van de uiting van centraal belang.[32]

Hoewel *hate speech* niet is voorbehouden aan het online domein, is dit een ruimte waarin *hate speech* zich makkelijker, sneller en verder kan verspreiden dan in de offline wereld, en die gebruikers daarbij ogenschijnlijk een hoge mate van anonimiteit biedt.[33] Online *hate speech* kan bovendien een katalysator zijn voor offline geweld, en de impact ervan reikt vaak verder dan de persoon waartegen een specifieke uiting is gericht.[34] Nagenoeg alle grote online platformen hebben regels met betrekking tot *hate speech* en verbieden daarbij in ieder geval het oproepen tot geweld en haatdragende en haatzaaiende uitingen; hierbij wordt veelal geput uit juridische terminologie uit bijvoorbeeld Amerikaanse antidiscriminatiewetgeving, met name ten aanzien van beschermde of kwetsbare groepen.[35] De in deze regels gehanteerde definities van *hate speech*, en de soorten uitlatingen die als voorbeeld worden gegeven van wat als ontoelaatbaar wordt beschouwd, verschillen desalniettemin per platform. Sommige platformen geven er de voorkeur aan zich niet te snel te mengen in het online debat en slechts de grootste uitwassen, zoals het oproepen tot geweld, te bestrijden, terwijl andere platformen strengere regels hanteren en proactiever optreden.[36]

Er is op zowel Europees als nationaal niveau aandacht voor de bestrijding van (online) *hate speech*. Zo heeft de Raad van Europa diverse aanbevelingen gedaan om *hate speech* te bestrijden.[37] In de EU is met name het Kaderbesluit racisme en vreemdelingenhaat van belang.[38] Dit Kaderbesluit vormt ook de basis voor de *Code of Conduct on Countering Illegal Hate Speech Online* die later in deze casestudy besproken zal worden.[39] Daarnaast verplicht de Richtlijn

[30] Zie voor verschillende definities o.a. Rosenfeld, *Cardozo Law Review* 2003, p. 1523; Cohen-Almagor, *Policy & Internet* 2011, p. 1.
[31] Waldron 2012, p. 5.
[32] Post 2009, p. 123.
[33] Zie Wilson 2012, p. 3; López & López 2017, p. 11-12. Zie met betrekking tot de rol van anonimiteit Mondal, Silva & Benevenuto 2017.
[34] Een bekend voorbeeld van hoe online *hate speech* fysiek geweld kan aanwakkeren is de bewuste inzet van *hate speech* in Myanmar. Zie Mozur, *The New York Times* 15 oktober 2018; Wilson 2012, p. 4. Zie ook Müller & Schwarz 2018.
[35] Gillespie 2018, p. 58. Zie voor het beleid van Facebook nl-nl.facebook.com/communitystandards/hate_speech. Zie voor het beleid van Twitter: help.twitter.com/nl/rules-and-policies/hateful-conduct-policy. Zie voor het beleid van YouTube: support.google.com/youtube/answer/2801939?hl=nl.
[36] Zie in dat verband bijvoorbeeld de aanscherping van het beleid dat Twitter voert (Conger, *The New York Times* 9 juli 2019). En met betrekking tot het online lastigvallen van mensen, zie Pater e.a. 2016, p. 369.
[37] Zie bijvoorbeeld Aanbeveling (97) 20 van het Comité van Ministers van de Raad van Europa (30 oktober 1997), *On hate speech*. Voor een overzicht van de activiteiten van de Raad van Europa op dit gebied, zie 'Freedom of expression. Hate speech', coe.int. Daarnaast monitort de *European Commission against Racism and Intolerance* (ECRI) het bestaan van racisme en intolerantie in Europa, zie daarvoor 'European Commission against Racism and Intolerance (ECRI)', coe.int. En zie met betrekking tot *hate speech* ECRI General Policy Recommendation no. 15 (8 december 2015) *On Combating Hate Speech*.
[38] Kaderbesluit 2008/913/JBZ van de Raad van 28 november 2008 betreffende de bestrijding van bepaalde vormen en uitingen van racisme en vreemdelingenhaat door middel van het strafrecht (*PbEU* 2008, L 328).
[39] Zie par. 4.3.1 voor een nadere bespreking. *Code of conduct on countering illegal hate speech online*, 30 juni 2016, ec.europa.eu/newsroom/just/document.cfm?doc_id=42985.

Audiovisuele mediadiensten lidstaten om ervoor te zorgen dat videoplatformen passende maatregelen nemen om het publiek te beschermen tegen video's die aanzetten tot geweld of haat.[40]

In Nederland zijn de belangrijkste juridische instrumenten ten aanzien van de bestrijding van *hate speech* het strafrechtelijke verbod op groepsbelediging en het verbod op haatzaaien.[41] Ook het verbod op de openbaarmaking van zulke uitingen en het verbod op o.a. deelname aan activiteiten die gericht zijn op discriminatie zijn in het kader van online *hate speech* relevant.[42] Ten aanzien van de aanpak van online *hate speech* speelt in Nederland ook het Meldpunt Internetdiscriminatie (MiND) een rol.[43] Personen kunnen discriminerende uitingen melden bij MiND, dat vervolgens een inschatting maakt van de strafbaarheid van de uiting in kwestie en een platform kan verzoeken de uiting te verwijderen. Als een verwijderverzoek niet wordt opgevolgd, kan dat ook leiden tot een melding aan het Openbaar Ministerie.

4.2.2 De werking van contentmodereeralgoritmen

Als het gaat om het domein van *hate speech*, dan worden algoritmen primair door platformen gebruikt om uitingen te detecteren die mogelijk kwalificeren als *hate speech*. Deze uitingen worden dan ter beoordeling voorgelegd aan een menselijke moderator. *Hate speech* kan vele vormen aannemen. Het hoeft bij *hate speech* niet alleen maar te gaan om geschreven tekst, maar er kan ook sprake zijn van (een combinatie van) afbeeldingen, audio en video's. Te denken valt aan zogeheten internetmemes, waarin tekst bijvoorbeeld wordt gecombineerd met een sprekende afbeelding. Het geheel van tekst en afbeelding kan dan gelden als *hate speech*. Het zijn met name beelden, en teksten die zijn opgenomen *in* een afbeelding of video, die voor algoritmen moeilijk te herkennen zijn.[44]

Er bestaat een breed spectrum van technologieën die kunnen worden ingezet voor het detecteren van *hate speech*. Een betrekkelijk eenvoudige manier om *hate speech* te detecteren is het gebruik van woordfilters. Er wordt dan door software gecheckt of er sprake is van gebruik van een woord op basis van een zwarte lijst van 'verboden' woorden.[45] In feite gaat het dan om een regelgebaseerd algoritme dat aanslaat op gebruik van vooraf bepaalde woorden. Eenvoudige woordfilters doen echter geen recht aan de context waarin het woord of de combinatie van woorden

[40] Richtlijn 2018/1808 van het Europees Parlement en de Raad van 14 november 2018 tot wijziging van Richtlijn 2010/13/EU betreffende de coördinatie van bepaalde wettelijke en bestuursrechtelijke bepalingen in de lidstaten inzake het aanbieden van audiovisuele mediadiensten (richtlijn audiovisuele mediadiensten) in het licht van een veranderende marktsituatie (*PbEU* 2018, L 303). Zie over die verplichting art. 28 ter lid 1 onder b van de richtlijn.
[41] Artt. 137c en 137d Sr.
[42] Respectievelijk artt. 137e en 137f Sr.
[43] Zie mindnederland.nl.
[44] Maar zie ook Sivakumar & Gordo, Paluri, engineering.fb.com 11 september 2018; 'Advancing self-supervision, CV, NLP to keep our platforms safe', ai.facebook.com 1 mei 2019.
[45] Instagram heeft een functionaliteit aan gebruikers aangeboden om zelf een (aanvullende) lijst van verboden woorden te bepalen. 'Keeping Comments Safe on Instagram', instagram.tumblr.com/post/150312324357/160912-news/embed.

wordt gebruikt en is er een grote kans dat uitingen door het algoritme ten onrechte als mogelijke *hate speech* worden gekwalificeerd, of dat het algoritme bepaalde vormen van *hate speech* juist níet aanbrengt.[46] Dergelijke algoritmen zijn namelijk niet goed in staat om de daadwerkelijke betekenis van een uiting, die voor een kwalificatie als *hate speech* van groot belang is, goed te interpreteren. Uitingen die minder expliciet zijn, of die sarcasme of ironie bevatten, zijn moeilijk te herkennen voor computersystemen.[47] Tegelijkertijd hoeft het gebruik van scheldwoorden niet vanzelfsprekend een indicatie te zijn van *hate speech*. Daarnaast kunnen bijvoorbeeld woorden die in principe een niet-pejoratieve betekenis hebben ook als scheldwoord worden gebruikt (denk aan '*gay*' of 'homo').[48] Bovendien bestaat er een kans dat gebruikers algoritmen misleiden door bijvoorbeeld woorden opzettelijk verkeerd te spellen of woorden met een positieve connotatie toe te voegen.[49]

Online platformen gebruiken dan ook steeds vaker zelflerende algoritmen die zij zelf ontwikkelen of die worden aangeboden door derde partijen.[50] Bij vormen van *supervised machine learning* worden uitingen van *hate speech* eerst handmatig als zodanig gelabeld. Die data worden gevoed aan het algoritme opdat het patronen gaat herkennen en deze 'kennis' kan toepassen bij het beoordelen van andere toekomstige uitingen. Voor het creëren van de benodigde datasets kunnen platformen de eerdere beslissingen van menselijke contentmoderators gebruiken, maar het labelen van de data kan ook worden uitbesteed via platformen als Amazon's *Mechanical Turk* of *CrowdFlower*, waar derden de data labelen tegen een kleine vergoeding.[51]

In geval van *supervised machine learning* is de juistheid van een beslissing sterk afhankelijk van de kwaliteit van de (gelabelde) data die wordt gebruikt in trainingsproces. In dat verband is het van belang dat data op consistente wijze worden gelabeld, door mensen met voldoende inhoudelijke kennis.[52] Een probleem kan zijn dat de datasets van uitingen die worden gebruikt om algoritmen te trainen, niet representatief zijn voor het type content dat wordt gemodereerd. Als ook gebruik wordt gemaakt van informatie over gebruikers, dan ligt het gevaar van een bevooroordeeld algoritme op de loer. Daarvan kan sprake zijn als de dataset waarop is getraind een onvolledig beeld schetst van de *hate speech* postende gebruiker.[53] Veranderende interne regels van het platform ten aanzien van wat precies mag en wat niet, kunnen de consistentie van de dataset – en daarmee de voorspelbaarheid van de uitkomst – eveneens negatief beïnvloeden.

[46] Warner & Hirschberg 2012; Davidson e.a. 2017; MacAvaney e.a., *PLoS ONE* 2019.
[47] Pavlopoulos, Malakasiotis & Androutsopoulos 2017, p. 1125.
[48] Davidson e.a. 2017.
[49] Gröndahl e.a. 2018.
[50] Voorbeelden van derde partijen zijn: Utopia AI Moderator (utopiaanalytics.com/utopia-ai-moderator/) en Hatebase (hatebase.org/).
[51] Matsakis, WIRED.com 22 maart 2018.
[52] Waseem 2016.
[53] MacAvaney e.a., *PLoS ONE* 2019.

Een daaraan gerelateerd probleem is dat wat geldt als *hate speech* en de wijze waarop mensen zich uitdrukken na verloop van tijd kan veranderen. Als algoritmen niet worden doorontwikkeld, kan dat leiden tot een verminderde nauwkeurigheid en dus een grotere kans op fout-positieve en fout-negatieve resultaten. Daarnaast kunnen opvattingen over wat *hate speech* is verschillen per taal en cultuur. Dat betekent dat wanneer online platformen bij de ontwikkeling van algoritmen vertrekken vanuit een specifiek cultuurgebonden begrip van *hate speech*, dit nadelige gevolgen kan hebben voor de nauwkeurigheid waarmee *hate speech* in andere culturen en talen wordt herkend.[54]

Natural language processing-technologie speelt een grote rol in het vinden van geschreven *hate speech*, met name als het gaat om het beoordelen van de inhoud van de uiting. *Natural language processing* is een subdomein binnen het domein van kunstmatige intelligentie dat zich bezighoudt met geschreven taal. Binnen dit subdomein wordt onder andere gebruik gemaakt van zelflerende algoritmen die in datasets patronen kunnen ontdekken met betrekking tot de zinsstructuur en inhoud van uitingen. Voorbeelden zijn aspecten zoals het sentiment van een tekst of tekstdeel.[55] Zulke elementen worden betrokken in het oordeel dat een modereeralgoritme vervolgens velt over een uiting.[56]

Ook andere gegevens over zowel de uiting als de gebruiker kunnen worden betrokken in het beoordelingsproces.[57] Gegevens over individuele gebruikers, zoals een online *track record* met daarin bijvoorbeeld informatie over of de gebruiker eerder is berispt voor het schenden van huisregels, kunnen worden gecombineerd met de resultaten van de bovengenoemde inhoudelijke analyse.[58] Andere gegevens over een specifieke uiting, zoals de lengte van de uiting en de mate waarin de inhoud is gerelateerd aan de inhoud van de originele post waaronder deze is geplaatst, kunnen eveneens als indicatoren worden gebruikt. Ook de kans dat bepaalde content *hate speech* ontlokt kan betrokken worden in de afweging om reacties daarop als *hate speech* aan te brengen.[59]

In het algemeen kan worden gesteld dat het bij de inzet van zelflerende algoritmen moeilijker is om achteraf te bepalen hoe een specifieke beslissing of inschatting tot stand is gekomen dan bij regelgebaseerde algoritmen. Bovendien vermindert deze inzichtelijkheid bij meer geavanceerde *machine learning* technieken zoals *deep learning*. Voor modereeralgoritmen kan de inzichtelijkheid in de totstandkoming van de uitkomst afnemen naarmate meer randgegevens, zoals gegevens over de gebruiker, worden aangewend om uiteindelijk tot een inschatting te komen. In de grote

[54] Kaye 2018, p. 18.
[55] Gillespie 2018, p. 103.
[56] Schmidt & Wiegand 2017, p. 3.
[57] Schmidt & Wiegand 2017, p. 5.
[58] Gillespie 2018, p. 104. Zie ook Cheng, Danescu-Niculescu-Mizil & Leskovec 2015. Zie ook Mishra e.a. 2018, p. 1088: '*[p]revious research suggests that [...] abusive content tends to come from users who share a set of common stereotypes and form communities around them*.'
[59] Schmidt & Wiegand 2017, p. 6.

hoeveelheid data die dan wordt aangewend kunnen steeds moeilijker de doorslaggevende factoren worden aangewezen, aan de hand waarvan de beslissing zou kunnen worden verklaard.

Van zelfstandige besluitvorming door algoritmen ten aanzien van mogelijke *hate speech* op platformen is voor zover wij weten geen sprake. Bij platformen zoals Facebook en YouTube beslist uiteindelijk een medewerker van een platform over de door het algoritme aangebrachte content. De bescheiden rol van het algoritme in het besluitvormingsproces is te verklaren door de complexiteit van de afwegingen die moeten worden genomen en het belang van context voor de beoordeling van een uiting.[60] Wel is het mogelijk dat algoritmen worden ingezet om reeds als *hate speech* aangemerkte content te identificeren als die opnieuw gedeeld worden en de content dan automatisch te verwijderen.[61]

Hoewel er doorgaans uiteindelijk menselijke moderators beslissen over het verwijderen van content, staat of valt de juistheid van beslissingen bij de zorgvuldigheid die zij betrachten in het besluitvormingsproces. In dat verband is van belang dat de zuiverheid van de beoordeling kan lijden onder zowel de werkdruk waaronder contentmoderators opereren als de emotionele en psychische stress die dit werk met zich meebrengt.[62] Ook de subjectiviteit van de beoordelaar en diens opvattingen kunnen ertoe leiden dat bepaalde content niet altijd goed wordt beoordeeld.

4.2.3 Blik op de toekomst

In ons onderzoek zijn we geen platformen tegengekomen die het detecteren, beoordelen en verwijderen van *hate speech* in zijn geheel overlaten aan algoritmen.[63] Een toekomstbeeld dat zich in de nabije toekomst zou kunnen voltrekken is dat content, waarover bij het algoritme nauwelijks twijfel bestaat dat er sprake is van *hate speech*, automatisch wordt verwijderd of in quarantaine wordt gezet.[64] Er is dan sprake van zogenaamde ‘semi-automatische’ contentmoderatie. Platformen zouden het probleem van *hate speech* dan afkunnen met minder menselijke moderators, die alleen zouden hoeven te beslissen over de in quarantaine geplaatste twijfelgevallen.

Op de langere termijn is het denkbaar dat algoritmen een grotere rol gaan spelen in het *voorkomen* van *hate speech*. Zo kunnen algoritmen gaan reageren op uitingen van *hate speech* in een poging

[60] Zie interview van TechCrunch met Timothy Quinn, CEO van Hatebase (Coldewey, techcrunch.nl 10 september 2019).
[61] Zie in dat verband ook HvJ EU 3 oktober 2019, ECLI:EU:C:2019:821 (*Eva Glawischnig-Piesczek/Facebook*).
[62] Zie in dat verband bijvoorbeeld Roberts 2016; Roberts 2014.
[63] Instagram is een voorbeeld van een platform dat met betrekking tot reacties die worden geplaatst op posts wel automatisch filtert met behulp van algoritmen. Instagram heeft een algoritme getraind dat ‘*offensive comments*’ op Instagramposts kan herkennen en verbergen. Strikt genomen gaat het hier niet om contentmoderatie door platformen, maar om een functionaliteit die door gebruikers van Instagram vrijwillig aangezet kan worden met betrekking tot de reacties die zij ontvangen op hun posts. Zie daarover Instagram, ‘Keeping Instagram a Safe Place for Self-Expression’, 29 juni 2017, instagram-press.com/blog/2017/06/29/keeping-instagram-a-safe-place-for-self-expression/. Zie ook Roberts, *VICE* 2 mei 2018.
[64] Pavlopoulos, Malakasiotis & Androutsopoulos 2017.

de ‘uiter’ te bewegen zijn gedrag in de toekomst aan te passen.[65] Mensen kunnen er, met behulp van algoritmen, in dit verband mogelijk ook toe worden bewogen om een concrete uiting niet te posten of die aan te passen.[66] Een gevaar is dan dat mensen onvoldoende vrij zijn zichzelf te uiten en dat de inhoud en de grenzen van het publieke debat mede worden bepaald door algoritmen.

4.3 Vrijheid van meningsuiting en informatie

Het verbieden van een bepaalde categorie uitingen en het handhaven daarvan staan in direct verband met de publieke waarde van vrijheid van meningsuiting, die op zichzelf weer een belangrijke pijler vormt onder onze democratische rechtsstaat. Hieronder wordt eerst de vrijheid van meningsuiting als publieke waarde geïntroduceerd en de wijze waarop die beschermd wordt. Daarna gaan we met betrekking tot het modereren van *hate speech* nader in op de kansen en risico’s voor de vrijheid van meningsuiting en evalueren we het relevante juridisch kader.

4.3.1 Vrijheid van meningsuiting en contentmoderatie

De vrijheid van meningsuiting omvat niet alleen het hebben en delen van een bepaalde mening, maar ook het ontvangen van meningen en informatie van anderen. Het modereren van online *hate speech* raakt daarmee aan zowel de vrijheid van gebruikers om hun mening te uiten en informatie te delen, als de vrijheid van ontvangers om meningen en informatie tot zich te nemen. In Nederland is de bescherming van de vrijheid van meningsuiting verankerd in art. 7 Gw, art. 10 EVRM en art. 11 Handvest. Ook choquerende en beledigende uitingen worden beschermd door de vrijheid van meningsuiting, omdat de belangrijke waarden in de democratische samenleving van pluralisme, tolerantie en ruimdenkendheid dat eisen.[67]

De vrijheid van meningsuiting is echter niet absoluut. Hoewel het belangrijk is dat schokkende of politiek gevoelige meningen bescherming genieten, betekent dit geen *carte blanche* voor haatdragende of haatzaaiende uitingen.[68] De wetgever kan de vrijheid van meningsuiting dan ook inperken ten behoeve van de bescherming van bepaalde publieke belangen en de bescherming van andere grondrechten en rechten van anderen. Een dergelijke inperking van de vrijheid van meningsuiting moet bij wet zijn voorzien, noodzakelijk zijn in een democratische samenleving, en

[65] Zie voor een voorbeeld waarin een Twitterbot racistische uitingen ‘sanctioneert’: Munger, *Political Behavior* 2017, p. 629-649.
[66] Zie voor een voorbeeld Santos e.a. 2018. Zie Jurgens, Chandrasekharan & Hemphill 2019 voor een oproep voor een proactieve inzet van *natural language processing* technologieën.
[67] EHRM 7 december 1976, ECLI:CE:ECHR:1976:1207JUD000549372 (*Handyside/Verenigd Koninkrijk*), r.o. 49.
[68] In dat verband moet ook worden opgemerkt dat het EHRM in bepaalde gevallen van *hate speech* een beroep op art. 10 EVRM niet-ontvankelijk of kennelijk ongegrond heeft verklaard, omdat de uitingen in kwestie de rechten van het EVRM aanvallen of ondermijnen. Zie in dat verband Keane, *Netherlands Quarterly of Human Rights* 2007, p. 641.

moet proportioneel zijn aan het nagestreefde legitieme doel. Zo kunnen ook bepaalde vormen van *hate speech* aan banden worden gelegd.[69]

Platformen zijn vrijgesteld van civielrechtelijke aansprakelijkheid voor opgeslagen content zolang zij niet weten dat er onrechtmatige content op hun servers staat.[70] In het geval zij kennis verkrijgen van onmiskenbaar onrechtmatige content, dienen platformen die onrechtmatige content 'prompt' te verwijderen.[71] Platformen hebben er daarom belang bij om onrechtmatige content te verwijderen als zij daarvan op de hoogte worden gesteld, omdat zij zich anders niet op de aansprakelijkheidsbeperking kunnen beroepen. Met deze regels wordt een balans gezocht tussen enerzijds de rechtszekerheid van internetdienstverleners met betrekking tot hun aansprakelijkheidspositie en anderzijds het belang om de verspreiding van illegale informatie en onrechtmatige activiteiten op het internet tegen te gaan.[72]

Uit art. 6:196c lid 5 BW volgt dat platformen, ondanks de aansprakelijkheidsvrijwaring, een verbod of bevel opgelegd kunnen krijgen om de verspreiding van onrechtmatige content te stoppen of te voorkomen. Wanneer een dergelijk verbod of bevel uitgevaardigd moet worden, wordt niet door het Europees recht bepaald. Art. 15 van de E-Commercerichtlijn stelt echter een grens aan wat er verlangd mag worden van platformen.[73] Verplichtingen mogen niet strekken tot een 'algemene toezichtverplichting' waarbij platformen alle geüploade materialen controleren op mogelijke illegale informatie of onrechtmatige activiteiten. Uit jurisprudentie van het HvJ EU volgt echter dat art. 15 van de E-Commercerichtlijn niet in de weg staat aan een rechterlijk bevel om reeds als onrechtmatige aangemerkte specifieke content ook in de toekomst verwijderd te houden.[74] Het is naar EU-recht ook geoorloofd om een bevel uit te vaardigen dat ertoe strekt ook 'overeenstemmende' content verwijderd te houden, voor zover het platform 'geautomatiseerde technieken en onderzoeksmethoden' kan toepassen en de content geen autonome beoordeling behoeft.[75] Het is daardoor denkbaar dat een Nederlandse rechter in de toekomst beslist dat een platform dient te voorkomen dat een specifieke of overeenstemmende onrechtmatige uiting op het

[69] Zie ook EHRM 6 juli 2006, ECLI:CE:ECHR:2006:0706JUD005940500 (*Erbakan/Turkije*), r.o. 56. Ook de aansprakelijkheid van bepaalde online dienstverleners voor dergelijke uitingen kan volgens het EHRM de toets van art. 10 EVRM doorstaan, zie EHRM 16 juni 2015, ECLI:CE:ECHR:2015:0616JUD006456909 (*Delfi AS/Estland*); EHRM 2 februari 2016, ECLI:CE:ECHR:2016:0202JUD002294713 (*MTE & Index.hu ZRT/Hongarije*). De uitspraken laten zich echter niet goed verhouden tot het Unierechtelijke kader waar dergelijke aansprakelijkheid is uitgesloten.
[70] Deze aansprakelijkheidsbeperkingen gelden ook voor eventuele strafrechtelijke aansprakelijkheid.
[71] *Kamerstukken II* 2003/04, 28197, nr. 15, p. 2.
[72] Nederland kent sinds 2008 een Gedragscode Notice-and-Takedown die wordt onderhouden door het Platform voor de InformatieSamenleving. Bij deze gedragscode zijn traditionele hostingbedrijven, maar ook platformen zoals Google en Facebook aangesloten. De gedragscode biedt een procedure voor notice-and-takedown en probeert onzekerheden weg te nemen ten aanzien van de te volgen procedure, de voorwaarden waaronder verwijderd dient te worden, en de timing van verwijdering. De inzet van algoritmen bij het detecteren van content door middel van algoritmen wordt echter niet gereguleerd door deze code. Zie voor de code: noticeandtakedowncode.nl.
[73] Richtlijn 2000/31/EG van het Europees Parlement en de Raad van 8 juni 2000 betreffende bepaalde juridische aspecten van de diensten van de informatiemaatschappij, met name de elektronische handel, in de interne markt (*PbEG* 2000, L 178/1. Art. 15 van deze richtlijn is niet expliciet geïmplementeerd in het Nederlandse recht.
[74] HvJ EU 3 oktober 2019, ECLI:EU:C:2019:821 (*Eva Glawischnig-Piesczek/Facebook*).
[75] HvJ EU 3 oktober 2019, ECLI:EU:C:2019:821 (*Eva Glawischnig-Piesczek/Facebook*).

platform wordt geplaatst. Een dergelijke beslissing zou ertoe kunnen leiden dat een platform algoritmen moet inzetten om die content te (helpen) detecteren en te verwijderen.

Specifiek ten aanzien van de aanpak van illegale *hate speech* door online platformen is er op Europees niveau de in 2016 opgestelde *Code of Conduct on Countering Illegal Hate Speech Online*. In die code heeft de Europese Commissie afspraken gemaakt met een aantal grote online platformen over de bestrijding van illegale online *hate speech*.[76] Daarin wordt onder andere gestreefd naar snelle en effectieve verwijdering van illegale *hate speech* naar aanleiding van meldingen.[77] Met de code verbinden de betrokken platformen zich ertoe de vrijheid van meningsuiting te bevorderen en faciliteren.[78] De code onderstreept ook het belang van de bestrijding van *hate speech*: '*The spread of illegal hate speech online not only negatively affects the groups or individuals that it targets, it also negatively impacts those who speak out for freedom, tolerance and non-discrimination in our open societies and has a chilling effect on the democratic discourse on online platforms.*'[79]

Naast deze specifieke code ten aanzien van *hate speech*, is ook de aanbeveling van de Europese Commissie met betrekking tot '*Measures to Effectively Tackle Illegal Content Online*' relevant.[80] Deze aanbeveling roept hosting providers, waaronder online platformen, op om effectieve, geschikte en proportionele maatregelen te nemen om de verspreiding van *illegal content* tegen te gaan. De aanbeveling roept providers op om ook proactief op te treden tegen illegale content, daar waar dat gepast en proportioneel is, onder meer door algoritmen in te zetten voor het detecteren van illegale content.[81] In de aanbeveling wordt bovendien benadrukt dat in alle gevallen waarin content verwijderd wordt of de toegang ertoe geblokkeerd, online platformen zich bewust moeten tonen van de centrale rol die zij hebben voor het faciliteren van het publieke debat, en voor het verspreiden en ontvangen van feiten, meningen en ideeën.[82]

[76] Microsoft, Twitter, Youtube, Instagram, (het inmiddels opgeheven) Google+, Snapchat, Dailymotion en Jeuxvideo.com zijn aangesloten bij de code. Zie voor de code en rapporten met betrekking tot de monitoring van de code 'The EU Code of conduct on countering illegal hate speech online. The robust response provided by the European Union', ec.europa.eu.
[77] Code of Conduct on Countering Illegal Hate Speech Online, 30 juni 2016, p. 1: '*The IT Companies to review the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, if necessary*.'
[78] Code of Conduct on Countering Illegal Hate Speech Online, 30 juni 2016, p. 1: 'Facebook*, Microsoft*, Twitter and YouTube (hereinafter "the IT Companies") – also involved in the EU Internet Forum – share, together with other platforms and social media companies, a collective responsibility and pride in promoting and facilitating freedom of expression throughout the online world.*'
[79] Code of Conduct on Countering Illegal Hate Speech Online, 30 juni 2016, p. 1.
[80] C(2018) 1177 final. In de aanbeveling wordt *illegal content* gedefinieerd als '*any information which is not in compliance with Union law or the law of a Member State concerned*'. Hoewel de aanbevelingen zich lijken toe te spitsen op informatie die naar haar aard onrechtmatig is, wordt er in de overwegingen bijvoorbeeld ook gewezen op de bestrijding van ongeautoriseerde verspreiding van auteursrechtelijk beschermd werk.
[81] C(2018) 1177 final, punt 18. De aanbeveling spreekt hier van '*use of automated means for the detection of illegal content*'.
[82] C(2018) 1177 final, overweging 31.

4.3.2 Kansen, risico's en bestendigheid juridisch kader

De kansen die de inzet van contentmodereeralgoritmen biedt voor de vrijheid van meningsuiting hebben met name betrekking op de bijdrage die zij kunnen leveren aan een pluralistische, tolerante en open samenleving. Contentmodereeralgoritmen vergroten de mogelijkheden om van online platformen een veiligere plek te maken voor met name minderheden om deel te nemen aan het publieke debat. Zij kunnen daarmee ook helpen voorkomen dat over bepaalde onderwerpen het debat niet wordt aangegaan omdat men bang is slachtoffer te worden van *hate speech*.[83]

Als het gaat om de bestrijding van *hate speech* dan is een grote uitdaging hoe en volgens welke maatstaven in concrete gevallen wordt vastgesteld of sprake is van *hate speech* en hoe daarbij kan worden voorkomen dat rechtmatige uitingen worden gecensureerd.[84] De schaal waarop in het online domein content moet worden gemodereerd, vergroot die uitdaging alleen maar. Algoritmen kunnen in dat opzicht een belangrijke bijdrage leveren, omdat ze het mogelijk maken om ook op grote schaal content te modereren.

Risico's die voortvloeien uit de inzet van algoritmen ten aanzien van de vrijheid van meningsuiting bestaan hoofdzakelijk in het onterecht aanbrengen van uitingen als *hate speech*. De oorzaken daarvan zijn reeds aangestipt in par. 4.2.2. Hoewel de uiteindelijke beslissing ten aanzien van de toelaatbaarheid van een uiting aan een menselijke content-moderator is, kunnen vooroordelen en andere omstandigheden die leiden tot onterechte kwalificatie als mogelijke *hate speech* door het algoritme toch een rol spelen als de menselijke moderator onder hoge druk snelle beslissingen moet nemen. Omgekeerd kunnen deze omstandigheden een rol spelen bij het onterecht niet kwalificeren van content als *hate speech*, wat de bestrijding van *hate speech* weer afhankelijk maakt van gebruikers die de content moeten rapporteren.

De *Code of Conduct on Countering Illegal Hate Speech Online*[85] is een door de EU geïnitieerd zelfreguleringsinstrument dat voorziet in concrete stappen die platformen moeten nemen om *hate speech* terug te dringen. Daarin wordt gestreefd naar snelle en effectieve verwijdering naar aanleiding van meldingen. Ook spreken platformen uit dat zij ernaar streven om onder andere inzicht te geven in de procedure die wordt gevolgd voor behandeling van meldingen, training te geven aan hun medewerkers, en voorlichting te geven aan gebruikers met betrekking tot wat er is toegestaan op het platform.

[83] Zie voor een Deense studie naar het verband tussen online speech en de deelname aan het debat, ook in relatie tot bepaalde onderwerpen Zuleta & Burkal 2017. Ook de Duitse *Netzwerkdurchsetzungsgesetz* (NetzDG) tracht zo een positieve bijdrage te leveren aan de vrijheid van meningsuiting, zie Theil, *Verfassungsblog* 8 februari 2018. Zie ook in algemene zin met betrekking tot de bestrijding van *hate speech* McGonagle 2013, p. 6.
[84] Massaro, *William and Mary Law Review* 1991, p. 214-215.
[85] Zie voetnoot 39 hierboven.

De implementatie van dit instrument wordt gemonitord door de Europese Commissie. De nadruk in de monitoring ligt daarin steeds op de '*removal rates*' van *hate speech*. Naleving en de effectiviteit van de code wordt met name uitgedrukt in en gemeten aan de hoeveelheden en percentages van *hate speech* die worden verwijderd naar aanleiding van gedane meldingen, en de snelheid waarmee dit gebeurt. De correctheid en de zorgvuldigheid van de besluitvorming komt echter nauwelijks ter sprake in de rapportages over de monitoring. Daarnaast valt op dat, hoewel de code aandacht besteedt aan de rol die onderwijs kan spelen bij het terugdringen van *hate speech*, er in de monitoring daarvoor ook geen aandacht is. Omdat de code en de monitoring ervan slechts betrekking hebben op het verwerken van meldingen van *hate speech* en niet op de inzet van algoritmen om zelf *hate speech* te detecteren, kan de code de risico's die samenhangen met de inzet van algoritmen voor de vrijheid van meningsuiting niet ondervangen.

In de algemenere Aanbeveling van de Europese Commissie met betrekking tot '*Measures to Effectively Tackle Illegal Content Online*' is er wel aandacht voor het proactief modereren van content met behulp van algoritmen. Een vraag die rijst is in hoeverre een dergelijk niet-bindend instrument platformen ertoe beweegt om het modereren van content vergezeld te doen gaan van waarborgen waarmee de vrijheid van meningsuiting voldoende kan worden beschermd.[86]

In de Aanbeveling is desalniettemin aandacht voor het gevaar dat rechtmatige content wordt verwijderd. Ook de rol die algoritmen daarin kunnen spelen komt aan de orde. Als hosting providers gebruikmaken van *automated means* om content te analyseren, dan dienen gepaste waarborgen te worden geboden om ervoor te zorgen dat genomen beslissingen precies (*'accurate'*) en op terechte ('*well-founded*') gronden worden genomen, in het bijzonder als het gaat om besluiten om content te verwijderen of de toegang ertoe te blokkeren.[87] Deze waarborgen zouden, waar gepast, moeten bestaan in het uitoefenen van toezicht ('*human oversight*') en het verrichten van verificaties door mensen. Dat zou in ieder geval moeten gebeuren waar een gedetailleerde beoordeling van de relevante context nodig is om te bepalen of bepaalde content illegaal is. Ook het menselijke toezicht dat gehouden moet worden, is niet nader aan voorwaarden verbonden.

De Aanbeveling roept ook op tot het geven van inzage in het beleid en de praktijken met betrekking tot het verwijderen van illegale content. Hosting providers worden aangemoedigd om met enige regelmaat daarover rapporten te publiceren, in het bijzonder ten aanzien van de hoeveelheid en het type content dat is verwijderd, de aantallen meldingen en bezwaren die zijn ingediend en de

[86] In dat verband is het tekenend dat de nieuwe Europese Commissie onder leiding van Von der Leyen heeft aangekondigd de aansprakelijkheid van online dienstverleners nader te willen regelen in wetgeving. Zie daarover Von der Leyen 2019, p. 13.
[87] C(2018) 1177 final, punt 19.

tijd die het heeft gekost om actie te ondernemen. De code roept niet specifiek op tot inzage in de wijze waarop algoritmen een rol spelen in het modereren van content.[88]

4.3.3 Tussenconclusie

Het reguleren en handhaven van online *hate speech* is onlosmakelijk verbonden met de publieke waarde van de vrijheid van meningsuiting. Hoewel terughoudendheid belangrijk is bij het verbieden en verwijderen van bepaalde uitingen, kan dit in bepaalde gevallen de vrijheid van meningsuiting ook ten goede komen omdat daarmee de voorwaarden worden geschept voor een inclusieve online omgeving die uitnodigt tot participatie. De inzet van algoritmen voor het detecteren van *hate speech* kan daaraan bijdragen, maar brengt ook risico's met zich mee. Als algoritmen uitingen onterecht als mogelijke *hate speech* aanmerken, en het menselijke toezicht onvoldoende is om dat recht te zetten, dan kan de inzet van algoritmen leiden tot censuur. In de bestaande juridische kaders is daarvoor echter onvoldoende aandacht.

De grondrechten, zoals de vrijheid van meningsuiting, kunnen handvatten bieden om deze problemen aan te pakken. Een complicatie in dat verband is wel dat het modereren van *hate speech* plaatsheeft op private platformen die zich door middel van *Terms of Service* en huisregels in een privaatrechtelijke relatie tot hun gebruikers verhouden. De grondrechten zijn niet direct van toepassing in deze verhoudingen. Wel kan gesteld worden dat overheden een positieve verplichting hebben om de vrijheid van meningsuiting te waarborgen als platformen content modereren.[89] Daaruit zou dan weer kunnen volgen dat overheden bestaande kaders dienen aan te passen om de risico's van algoritmische besluitvorming ten aanzien van de vrijheid van meningsuiting op effectieve wijze te mitigeren.

4.4 Bescherming van persoonsgegevens

4.4.1 Bescherming van persoonsgegevens en contentmoderatie

De bescherming van persoonsgegevens kan in het geding komen als algoritmen, naast een inhoudelijke analyse van de content, ook informatie over de gebruiker betrekken in hun beoordeling van de content. Een van de uitgangspunten van het gegevensbeschermingsrecht is dat individuen controle houden over de verwerking van hun persoonsgegevens.

4.4.2 Kansen, risico's en bestendigheid juridisch kader

Risico's ten aanzien van de bescherming van persoonsgegevens kunnen met name ontstaan als online platformen profielen bouwen van gebruikers, bijvoorbeeld met betrekking tot het type content dat zij posten en de kans dat die content onrechtmatig of onwenselijk is. Dergelijke

[88] Dat is wel het geval als het gaat om terroristische content, zie C(2018) 1177 final, punt 42.
[89] Zie in dat verband Angelopoulos e.a. 2015.

classificaties kunnen leiden tot besluitvorming waarbij stereotypes centraal komen te staan, en brengen, afhankelijk van het soort gegevens dat wordt gebruikt, mogelijk een verhoogd gevaar van discriminatie en onjuiste besluitvorming mee (zie ook par. 4.5). Ook worden er dan mogelijk gevoelige gegevens verzameld, die in geval van misbruik kunnen leiden tot schade aan de reputatie van gebruikers.[90]

De Algemene Verordening Gegevensbescherming (AVG), maar ook de gelijkheidswetgeving, biedt diverse waarborgen tegen voornoemde risico's. In de AVG zijn diverse beginselen neergelegd, zoals de beginselen van dataminimalisatie, doelbinding en transparantie, die ook zijn uitgewerkt in nadere regels in de AVG.[91] Als online content wordt gemodereerd en daarbij persoonsgegevens worden betrokken, dan dient aan die regels en beginselen te worden voldaan.[92]

4.4.3 Tussenconclusie

De bescherming van persoonsgegevens kan mogelijk in het geding komen als platformen gegevens over gebruikers betrekken in de analyse van content. Een risico dat zich kan voordoen is het overmatig gebruik van persoonsgegevens in een poging om het modereren van online content te optimaliseren. De beginselen en regels in de AVG bieden voldoende aanknopingspunten om gebruik van persoonsgegevens door online platformen te reguleren.

4.5 Non-discriminatie

4.5.1 Non-discriminatie en contentmoderatie

De aanpak van *hate speech* is nauw verbonden met de publieke waarde van non-discriminatie, in zoverre dat *hate speech* veelal de (gelijk)waardigheid van groepen personen ondermijnt of er zelfs op is gericht die te ondermijnen.[93] Racisme, seksisme en andere houdingen die discriminatie tussen groepen personen in stand houden of in de hand werken, kunnen door *hate speech* worden bestendigd.[94] De eerder genoemde *Code of Conduct on Countering Illegal Hate Speech Online*[95] beoogt daarom online *hate speech* op platformen te bestrijden.

[90] Zie met betrekking tot de risico's van profilering in algemene zin Schermer 2013, p. 137.
[91] Zie voor de beginselen art. 5 AVG.
[92] Zie in dat verband ook C(2018) 1177 final, overwegingen 13 en 39, waarin wordt benadrukt dat bij de aanpak van illegale online content onder andere het grondrecht op gegevensbescherming moet worden gerespecteerd, en dat maatregelen die worden genomen om gevolg te geven aan de Aanbeveling volledig in overeenstemming moeten zijn met de geldende regels ten aanzien van gegevensbescherming.
[93] Waldron 2012.
[94] Cowan, *Journal of Social Issues* 2002, p. 250.
[95] Zie voetnoot 39 hierboven.

4.5.2 Kansen, risico's en bestendigheid juridisch kader

De inzet van algoritmen voor het detecteren van *hate speech* biedt kansen voor de aanpak van discriminatie in de zin dat racistische, seksistische en andere haatzaaiende of haatdragende uitingen sneller kunnen worden gedetecteerd en daartegen kan worden opgetreden, wat zou moeten leiden tot inclusievere online omgevingen.

Tegenover die kans staat het aanmerkelijke risico dat de algoritmen die hiervoor worden ingezet de vooringenomenheden van hun programmeurs overnemen, of vooroordelen weerspiegelen die aanwezig zijn in de data waarmee deze algoritmen worden getraind.[96] Twee recente studies wijzen er bijvoorbeeld op dat de modereeralgoritmen van Twitter content vaker als beledigend of als haatzaaiend aanmerken wanneer deze elementen van straattaal of '*slang*' bevat.[97] Deze vooringenomenheid van het algoritme is te herleiden tot vooroordelen die leven ten aanzien van mensen die straattaal bezigen. Als ook menselijke moderators ten aanzien van content met straattaal besluiten dat er sneller sprake is van *hate speech,* en die besluiten weer de invoer vormen voor het aanscherpen van het algoritme, dan kan er een *feedback loop* ontstaan waarbij de ongelijke behandeling dieper wordt verankerd. De risico's die bestaan ten aanzien van de vrijheid van meningsuiting gelden dan ook sterker voor groepen personen waarover al vooroordelen bestaan. De impact van een dergelijk discriminerend effect kan bovendien worden versterkt als gevolg van de schaalvergroting die mogelijk wordt door de inzet van algoritmen.

Tegelijkertijd is het denkbaar dat juist met algoritmen de vooringenomenheden en vooroordelen in modereeralgoritmen kunnen worden blootgelegd. Waar de inzet van modereeralgoritmen leidt tot vormen van discriminatie van bepaalde beschermde groepen, zijn zulke algoritmen bijzonder waardevol. Het verbod op verwerking van bijzondere persoonsgegevens in de AVG staat echter mogelijk in de weg aan de succesvolle implementatie van zulke algoritmen, aangezien daarvoor juist het gebruik van bijzondere persoonsgegevens zoals ethniciteit en ras nodig kan zijn.[98]

Zowel de *Code of Conduct* als de Aanbevelingen van de Europese Commissie bevatten geen specifieke normen die strekken tot het voorkomen van vooroordelen in modereeralgoritmen en de bovengenoemde mogelijkheden van discriminatie. Daarnaast worden platformen niet opgeroepen om zelf *best practices* te ontwikkelen waarmee de risico's op discriminatie zouden kunnen worden ondervangen. Ook aan het menselijke toezicht dat gehouden moet worden, worden geen nadere voorwaarden gesteld die vooringenomenheid of discriminatie kunnen tegengaan.

[96] Binns e.a. 2017.
[97] Davidson, Bhattacharya & Weber 2019; Sap e.a. 2019. Zie ook, met betrekking tot een dataset van Wikipedia, Binns e.a. 2017.
[98] Art. 9 AVG.

4.5.3 Tussenconclusie

De inzet van algoritmen kan mogelijk bijdragen aan een inclusievere online omgeving waarin discriminatie op een efficiëntere wijze wordt bestreden. Echter, als vooringenomenheden de ingezette modereeralgoritmen binnensluipen, levert dat ook weer gevaren op voor de publieke waarde van non-discriminatie. De *Code of Conduct on Countering Illegal Hate Speech Online*[99] en de Aanbeveling van de Europese Commissie met betrekking tot '*Measures to Effectively Tackle Illegal Content Online*'[100] laten deze problematiek onbesproken, waardoor deze instrumenten geen of nauwelijks richting bieden ten aanzien van het voorkomen of anderszins mitigeren van mogelijke discriminatie door modereeralgoritmen.

4.6 Rechtsbescherming

4.6.1 Rechtsbescherming en contentmoderatie

Personen die door hen geplaatste content verwijderd zien, moeten de beslissing die daaraan ten grondslag ligt, kunnen aanvechten. Daarvoor is ook van belang dat zij begrijpen op grond waarvan hun content wordt verwijderd.

4.6.2 Kansen, risico's en bestendigheid juridisch kader

Een lastigheid is dat de relatie tussen online platformen en hun gebruikers en de beschikbaarheid van eventuele bezwaarmogelijkheden worden beheerst door overeenkomsten zoals *Terms of Service*. Omdat platformen zoals YouTube en Facebook als dominant zijn aan te merken, is van een gelijkwaardige verhouding tussen platform en gebruikers, en een vrije beslissing over het al dan niet accepteren van de voorwaarden, feitelijk echter geen sprake. Dit resulteert onder andere in een algemeen gebrek aan (serieuze) mogelijkheden voor gebruikers om bezwaar te maken tegen de beslissingen van platformen. De negatieve impact van dat gebrek op de rechtsbescherming van gebruikers geldt mogelijk nog sterker wanneer voor zulke besluitvorming ook algoritmen worden ingezet. Omdat modereeralgoritmen door allerlei factoren tot uitkomsten kunnen komen die mensen onlogisch of duidelijk verkeerd voorkomen, wordt het probleem van het gebrek aan bezwaarmogelijkheden nog prangender.[101]

Een risico met betrekking tot de rechtsbescherming van gebruikers van platformen is de gebrekkige inzichtelijkheid van het besluitvormingsproces en de rol die algoritmen daarin spelen op zowel individueel als meer algemeen niveau. Voor rechtsbescherming van platformgebruikers is het van belang dat zij begrijpen op grond waarvan hun uitingen zijn verwijderd en dat hen

[99] Zie voetnoot 39 hierboven.
[100] C(2018) 1177 final.
[101] Zie ook Meyers West, *New Media & Society* 2018, p. 4366, over het onbegrip dat bestaat bij gebruikers over besluiten die online platformen maken op het gebied van contentmoderatie.

daarover informatie wordt verstrekt. Dat vergt ook algoritmen die tot een uitlegbare uitvoer kunnen komen.[102] Als zelflerende algoritmen worden ingezet dan is dat echter geen eenvoudige opgave.[103]

Inzichtelijkheid in de werking van modereeralgoritmen in algemene zin kan eraan bijdragen dat duidelijk wordt hoe groot de risico's zijn met betrekking tot de vrijheid van meningsuiting en non-discriminatie. Als dat inzicht wordt verkregen dan wordt het mogelijk om daarop beleid te voeren en regulering daarop toe te snijden.

In dat kader wreekt zich dat de *Code of Conduct on Countering Illegal Hate Speech Online*[104] is toegespitst op het afhandelen van meldingen over content die is aangebracht door andere gebruikers en niet zozeer door modereeralgoritmen. De Aanbeveling '*Measures to Effectively Tackle Illegal Content Online*'[105] bevat wel bepalingen die raken aan de rechtsbescherming van de gebruiker. Hij dient op de hoogte te worden gebracht van de redenen om zijn content te verwijderen.[106] De gebruiker moet ook een mogelijkheid hebben om bezwaar aan te tekenen.[107] Platformen dienen volgens de Aanbeveling ook in algemene zin inzage te geven in het beleid en de praktijken met betrekking tot het verwijderen van illegale content. Platformen worden aangemoedigd om met enige regelmaat daarover rapporten te publiceren, in het bijzonder ook ten aanzien van de hoeveelheid en het type content dat is verwijderd, de aantallen meldingen en bezwaren die zijn ingediend en de tijd die het heeft gekost om actie te ondernemen.[108] De Code roept eveneens niet specifiek op tot inzage in de wijze waarop algoritmen een rol spelen in het modereren van content.[109]

Art. 22 AVG kan ook een rol spelen in de rechtsbescherming van platformgebruikers. Dit artikel bepaalt dat personen wier persoonsgegevens worden verwerkt het recht hebben om 'niet te worden onderworpen aan een uitsluitend op geautomatiseerde verwerking, waaronder profilering, gebaseerd besluit waaraan voor hem rechtsgevolgen zijn verbonden of dat hem anderszins in aanmerkelijke mate treft'.

Art. 22 AVG lijkt echter niet van toepassing op de huidige modereerpraktijk van online platformen. Het is in de eerste plaats onzeker of er in geval van het verwijderen van content sprake is van een beslissing waaraan rechtsgevolgen zijn verbonden, dan wel een beslissing die de gebruiker van een platform in aanmerkelijke mate treft. Afhankelijk van de uiting in kwestie en de gevolgen die niet-plaatsen van de content heeft, zou er mogelijk sprake kunnen zijn van een beslissing die de

[102] Zie ook Suzor e.a., *International Journal of Communication* 2019, p. 1526.
[103] Zie par. 2.1.2.
[104] Zie voetnoot 39 hierboven.
[105] C(2018) 1177 final.
[106] C(2018) 1177 final, punt 9.
[107] C(2018) 1177 final, punt 11.
[108] C(2018) 1177 final, punt 17.
[109] Dat is wel het geval als het gaat om terroristische content, zie C(2018) 1177 final, punt 42.

gebruiker in aanmerkelijke mate treft omdat daardoor zijn vrijheid van meningsuiting wordt beperkt. Echter, voor zover er sprake is van een dergelijke beslissing, dan geldt dat de beslissing waarschijnlijk geen besluit is dat is gebaseerd op een *uitsluitend* geautomatiseerde verwerking. Het modereren van *hate speech* vergt op dit moment namelijk nog steeds inmenging van een menselijke moderator die uiteindelijk beslist over en verantwoordelijkheid heeft voor het wel of niet verwijderen van content.[110]

Het kan niet worden uitgesloten dat het modereren van *hate speech* in de toekomst verder wordt geautomatiseerd en menselijke tussenkomst niet langer nodig is. Art. 22 AVG is dan, zeker als de besluitvorming discriminerende effecten heeft, waarschijnlijk wel van toepassing.[111] Platformen zouden dan de uitdrukkelijke toestemming moeten verkrijgen van platformgebruikers voor de inzet van modereeralgoritmen.[112] Voor platformen is dan van belang dat zij maatregelen nemen ter bescherming van de rechten en vrijheden van hun gebruikers. Daartoe behoren ook het recht op menselijke tussenkomst, het recht van de gebruiker om zijn standpunt kenbaar te maken en het recht om het modereerbesluit aan te vechten.[113] Ook het recht op informatie met betrekking tot de geautomatiseerde besluitvorming is dan van toepassing.[114] Dat zou de gebruiker van een online platform ook het recht geven om 'nuttige informatie over de onderliggende logica' van het besluitvormingsproces te ontvangen. Dat zou van online platforms vergen dat zij begrijpelijke uitleg geven over de waarschijnlijk zeer complexe wijze waarop een algoritme beslist.[115]

4.6.3 Tussenconclusie

Rechtsbescherming ondersteunt de andere relevante publieke waarden en kan eraan bijdragen dat eventuele risico's van de inzet van algoritmen door platforms met betrekking tot de vrijheid van meningsuiting en non-discriminatie kunnen worden gemitigeerd of in ieder geval gepaard gaan met bepaalde waarborgen voor gebruikers. Inzicht in de rol die algoritmen spelen in het besluitvormingsproces kunnen daarbij van belang zijn. Het huidige juridisch kader biedt echter geen transparantieverplichtingen. Ook de Aanbeveling '*Measures to Effectively Tackle Illegal Content Online*' roept niet op tot het inzichtelijk maken van de werking van algoritmen die onrechtmatige content aanbrengen. Hoewel art. 22 AVG verschillende van zulke waarborgen biedt ten aanzien van de rechtsbescherming bij geautomatiseerde besluitvorming, moet worden geconstateerd dat de huidige contentmoderatiepraktijk in ieder geval op dit moment nog niet zodanig geautomatiseerd is dat die bepaling van toepassing is. Tot dat moment lijkt de rechtsbescherming van gebruikers bij contentmoderatie dan ook onvoldoende gewaarborgd.

[110] Vgl. Groep gegevensbescherming artikel 29 2017b, p. 24.
[111] Vgl. Groep gegevensbescherming artikel 29 2017b, p. 26.
[112] Art. 22 lid 2 onder c AVG.
[113] Art. 22 lid 3 AVG.
[114] Art. 13 lid 2 onder f en art. 14 lid 2 onder g AVG.
[115] Vgl. Groep gegevensbescherming artikel 29 2017b, p. 25.

4.7 Conclusie

De inzet van algoritmen die onrechtmatige of onwenselijke content kunnen opsporen en aanbrengen, is onlosmakelijk verbonden met de schaal waarop content vandaag de dag wordt gegenereerd en gedeeld.[116] 'Traditionele' methoden van modereren zijn niet toegerust op deze hoeveelheden content, zodat de automatisering van een deel van het modereerproces in het huidige internetlandschap onontkoombaar is. Een kans die daarom in algemene zin geldt voor alle vormen van contentmoderatie is dat online platformen met het gebruik van algoritmen het probleem van schaal waarop illegale en onwenselijke content wordt gedeeld het hoofd kunnen bieden.

Om de rol die algoritmen (kunnen) spelen in contentmoderatie en de daarmee samenhangende kansen en risico's te kunnen waarderen, dient onderscheid te worden gemaakt naar het type content in kwestie, de onrechtmatigheid of onwenselijkheid van die content, en de gevolgen ervan. De inzet van algoritmen voor het modereren van content zal afhankelijk van deze onderscheidingen meer of minder geschikt zijn. Zo zijn uitingen die naar hun aard onrechtmatig zijn waarschijnlijk eenvoudiger te detecteren dan uitingen waarvoor een uitgebreide analyse nodig is van de context waarin een bepaalde uiting is gedaan. Daarnaast kan ook content die reeds als onrechtmatig of onwenselijk is aangemerkt eenvoudiger gedetecteerd worden dan content waarvan dat (nog) niet vaststaat.

In deze casestudy bestudeerden wij in het bijzonder de inzet van algoritmen door online platformen voor het detecteren van *hate speech*. Algoritmen maken het mogelijk om op grote schaal *hate speech* aan te pakken en dat biedt kansen voor het inclusiever maken van online omgevingen, waarin personen zich niet ontmoedigd voelen om te participeren in het publiek debat. Online platformen nemen daarin al de nodige stappen. Ook de aansprakelijkheidsnormen en zelfreguleringsinstrumenten bieden stimulansen om *hate speech* proactief op te sporen met behulp van algoritmen.

Risico's die samenhangen met de inzet van algoritmen om *hate speech* te detecteren doen zich vooral voor in relatie tot de publieke waarden van de vrijheid van meningsuiting en non-discriminatie. Er is een risico dat uitingen ten onrechte als onrechtmatig of als onwenselijk worden gekwalificeerd en aangebracht en dat werkelijk onrechtmatige dan wel onwenselijke uitingen niet als zodanig worden herkend. Daaraan draagt bij dat algoritmen niet altijd goed in staat zullen zijn om zich rekenschap te geven van de betekenis van een uiting en de context waarbinnen een uiting wordt gedaan. Deze risico's zullen tot op zekere hoogte ook gelden ten aanzien van het modereren van andere typen onrechtmatige of ongewenste content. Naarmate een verwijderbeslissing een

[116] Gillespie 2018.

genuanceerde juridische beoordeling vraagt, is de kans groter dat een modereeralgoritme onterecht bepaalde content wel of niet aanbrengt.

Ook het gevaar van vooroordelen speelt daarbij een rol. Wetenschappelijk onderzoek naar de prestaties van *hate speech* modereeralgoritmen wijst erop dat uitingen door minderheidsgroepen als gevolg van vooringenomenheden een hogere kans hebben om te worden gedetecteerd. Menselijke moderators moeten uiteindelijk beslissen over aangebrachte content en kunnen onjuiste kwalificaties rechtzetten. De hoge druk waaronder menselijke contentmoderators veelal moeten beslissen en de beperkte instructies die zij krijgen, nuanceren echter de waarde van dat menselijke toezicht.

De publieke waarde van rechtsbescherming komt met name in geding door de gebrekkige inzichtelijkheid van de werking van besluitvormingsprocessen. De informatie die door online platformen zelf naar buiten wordt gebracht is doorgaans niet inzichtelijk genoeg om een goed beeld te kunnen vormen van de contentmoderatiepraktijk als geheel en de rol van algoritmen daarin. Dat bemoeilijkt het opstellen van passend beleid en het beoordelen van de bestendigheid van bestaande regelgeving.[117]

Gebruikers komen, voor zover bekend, niet te weten of een algoritme een rol heeft gespeeld bij de totstandkoming van dat besluit en wat de eventuele bijdrage van het algoritme was. Hierdoor kan de rechtsbescherming van gebruikers jegens private online platformen – die in veel gevallen al beperkt is als gevolg van de voorwaarden die gebruikers contractueel aanvaarden – onder druk komen te staan. Als platformen in de toekomst echter (sommige) content volledig geautomatiseerd of semi-geautomatiseerd modereren, dan kan de AVG mogelijk uitkomst bieden voor het inzichtelijker maken van het modereerproces en de rol die algoritmen daarin spelen.

In regelgeving met betrekking tot het modereren van content in het algemeen en in het bijzonder ten aanzien van *hate speech* is nog maar nauwelijks aandacht voor de gevolgen van het algoritmisch aanbrengen van onrechtmatige content. Hoewel de druk op platformen wordt opgevoerd om proactief te modereren en daarvoor ook technologie in te zetten, is voor de risico's van de inzet van algoritmen, zoals de mogelijke discriminerende effecten, nog onvoldoende aandacht. Menselijk toezicht op contentmoderatie wordt wel aangemoedigd. Maar bij gebrek aan voorwaarden waaraan dat toezicht inhoudelijk moet voldoen, is een modereerpraktijk ontstaan waarin menselijke moderators onder grote druk moeten beslissen, en waarbij eventuele risico's

[117] Speciaal rapporteur voor de bevordering en bescherming van het recht op vrijheid van meningsuiting aan de VN David Kaye benadrukt in dat verband dat transparantie noodzakelijk is bij de automatisering van contentmoderatie vanwege mogelijke risico's ten aanzien van grondrechten. Daarnaast is het volgens hem van wezenlijk belang dat de samenleving en maatschappelijke belangenorganisaties worden betrokken bij de implementatie van zulke geautomatiseerde processen (Kaye 2018, p. 18).

die samenhangen met de inzet van algoritmen niet of onvoldoende door het huidige juridisch kader worden gemitigeerd.

Hoofdstuk 5. Casestudy Zelfrijdende auto's

Vicky Breemen & Anouk Wouters[1]

5.1 Introductie

Volgens de *2019 Autonomous Vehicles Readiness Index* van KPMG is Nederland voor het tweede jaar op rij het land dat het best is voorbereid op de komst van zelfrijdende auto's.[2] De index, waarin 25 landen zijn opgenomen, kijkt naar vier factoren: beleid en wetgeving; technologie en innovatie; infrastructuur en acceptatie door consumenten. Voor Nederland worden onder andere de Experimenteerwet voor zelfrijdende auto's[3] en de invoering van een rijbewijs voor voertuigen uitgelicht. Dat Nederland goed voorbereid is op de komst van zelfrijdende auto's is in lijn met de ambitie om Nederland op de kaart te zetten als land waar deze innovaties kunnen plaatsvinden.[4]

Deze casestudy richt zich op (de ontwikkeling van) de zelfrijdende personenauto,[5] oftewel de autonome, intelligente auto. In de literatuur worden de elementen van deze definitie als volgt omschreven. 'Autonomie' betreft het benodigde niveau van menselijke tussenkomst in het functioneren: minder menselijke interventie resulteert in een hogere mate van zelfstandigheid en vice versa. 'Intelligentie' ziet op de manier waarop een systeem de omgeving kan waarnemen en kan inspelen op veranderende omstandigheden. 'Auto's' zijn gemotoriseerde voertuigen die onder andere voor transport worden ingezet.[6] Deze definitie van autonome intelligente auto's impliceert verschillende niveaus van automatisering.[7] Figuur 1 toont de niveaus van automatisering van de rijtaak zoals gedefinieerd door SAE International.[8] Bij niveau 0 doet de bestuurder alles zelf, bij niveau 5 is dit de auto. Bij de niveaus 1-4 is sprake van, in toenemende mate, overname van de rijtaak door de auto. Bij niveau 1 (*driver assistance*) kan de auto één taak zelfstandig verrichten. Het gaat dan bijvoorbeeld om adaptieve cruisecontrol.[9] Bij niveau 2 (*partial driving automation*) zijn ten minste twee functies geautomatiseerd, zoals een combinatie van adaptieve cruisecontrol en

[1] De auteurs willen Roeland de Bruin en Peter Blok hartelijk danken voor hun waardevolle feedback op eerdere versies van deze casestudy, en Jonas Folkers voor zijn bijdrage aan het literatuuronderzoek.
[2] KPMG 2019, p. 14.
[3] Wet van 26 september 2018 tot wijziging van de Wegenverkeerswet 1994 in verband met mogelijk maken van experimenten met geautomatiseerde systemen in motorrijtuigen (*Stb*. 2018, 347).
[4] *Kamerstukken II* 2013/14, 31305, nr. 210.
[5] Personenauto's zijn '*the main driver of the development towards automated driving*', omdat hun grote marktaandeel investeringen in de ontwikkeling van de benodigde technologieën mogelijk maakt. Bovendien evolueren personenauto's '*level by level with more sensors, connectivity and computing power on- and off-board*'. Zie ERTRAC Working Group "Connectivity and Automated Driving" 2019, p. 13.
[6] De Bruin, *European Journal of Risk Regulation* 2016, p. 485; De Cock Buning & De Bruin, *Connection Science* 2017, p. 190.
[7] Zie over verschillende niveaus van automatisering in auto's ook Van Est & Gerritsen 2017, p. 33-36; Engelhard & De Bruin 2018, p. 4-5.
[8] SAE International is een wereldwijd verband van ingenieurs en technici in de ruimtevaart- en auto-industrie (sae.org/about/).
[9] Van den Acker, *VR* 2015, p. 367.

een systeem om in de rijstrook te blijven (*lane keeping*). De bestuurder wordt dus hulp geboden bij de besturing. Bij niveau 3 (*conditional driving automation*) kan de auto onder bepaalde (ideale) omstandigheden helemaal zelf rijden, maar is de bestuurder nog wel nodig als back-up – klaar om de controle over te nemen als de auto daar om vraagt. Bij niveau 4 (*high driving automation*) zijn auto's zodanig ontworpen dat ze in beginsel zelfstandig rijden. Mochten er toch omstandigheden zijn waaronder de auto niet veilig zijn weg kan vervolgen, dan kan de bestuurder worden gevraagd om het weer over te nemen. Als deze niet reageert, brengt de auto zichzelf op een veilige locatie tot stilstand. Bij niveau 5 (*full driving automation*) kan de auto onder alle omstandigheden zelfstandig rijden.[10]

Algoritmen zijn de drijvende kracht achter de ontwikkeling van de steeds autonomer wordende auto.[11] Deze worden, zoals later in dit stuk uitgebreider zal worden omschreven, gebruikt om de omgeving te analyseren, met andere actoren op de weg of de infrastructuur te communiceren en uiteindelijk beslissingen te nemen over welke handeling(en) de auto zal uitvoeren.

[10] Deze indeling van niveaus gaat uit van een scenario van een geleidelijke ontwikkeling van de conventionele auto naar een steeds verder geautomatiseerde auto met als eindpunt een volledig autonoom rijdend voertuig. Van Wees wijst in dit verband op het feit dat er ook toepassingsvarianten denkbaar zijn die een meer 'revolutionair implementatietraject' kennen, in die zin dat van meet af aan sprake is van een zelfrijdend voertuig zonder dat de bemoeienis van een bestuurder vereist is. Daarbij kan gedacht worden aan voertuigen die zich beperken tot vaste routes in een geografisch beperkt gebied en met relatief lage snelheden rijden. In dit verband wordt wel een onderscheid gemaakt tussen een '*something everywhere*' en een '*everything somewhere*' benadering. Zie Van Wees, *AV&S* 2015, p. 178 (met verwijzing).
[11] Europese Commissie 2018, p. 1.

Figuur 1

Level	Name	Narrative definition	DDT		DDT fallback	ODD
			Sustained lateral and longitudinal vehicle motion control	*OEDR*		
Driver* performs part or all of the *DDT						
0	**No Driving Automation**	The performance by the *driver* of the entire *DDT*, even when enhanced by *active safety systems*.	*Driver*	*Driver*	*Driver*	n/a
1	**Driver Assistance**	The *sustained* and *ODD*-specific execution by a *driving automation system* of either the *lateral* or the *longitudinal vehicle motion control* subtask of the DDT (but not both simultaneously) with the expectation that the *driver* performs the remainder of the *DDT*.	*Driver* and *System*	*Driver*	*Driver*	Limited
2	**Partial Driving Automation**	The *sustained* and *ODD*-specific execution by a *driving automation system* of both the *lateral* and *longitudinal vehicle motion control* subtasks of the *DDT* with the expectation that the *driver* completes the *OEDR* subtask and *supervises* the *driving automation system*.	***System***	*Driver*	*Driver*	Limited
***ADS* ("*System*") performs the entire *DDT* (while engaged)**						
3	**Conditional Driving Automation**	The *sustained* and *ODD*-specific performance by an *ADS* of the entire DDT with the expectation that the *DDT fallback-ready user* is *receptive* to *ADS*-issued *requests to intervene*, as well as to *DDT performance-relevant system failures* in other *vehicle* systems, and will respond appropriately.	*System*	***System***	*Fallback-ready user (becomes the driver during fallback)*	Limited
4	**High Driving Automation**	The *sustained* and *ODD*-specific performance by an *ADS* of the entire *DDT* and *DDT fallback* without any expectation that a *user* will respond to a *request to intervene*.	*System*	*System*	***System***	Limited
5	**Full Driving Automation**	The *sustained* and unconditional (i.e., not *ODD*-specific) performance by an *ADS* of the entire *DDT* and *DDT fallback* without any expectation that a *user* will respond to a *request to intervene*.	*System*	*System*	*System*	**Unlimited**

Bron: 'Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. J3016_201806', SAE International juni 2018, sae.org/standards/content/j3016_201806. Gebruikte afkortingen: *dynamic driving task* (DDT), *operational design domain* (ODD), *object and event detection and response* (OEDR), en ten slotte *automated driving system* (ADS).

Zelfrijdende auto's worden, onder andere, ontwikkeld met het idee dat zij op den duur kunnen zorgen voor een betere verkeersveiligheid.[12] Ook kan deze innovatieve vorm van vervoer door middel van rijstijl-verbetering het verbruik van brandstof optimaliseren en zo bijdragen aan duurzamer vervoer.[13] Vanwege deze (te verwachten) positieve gevolgen, wordt de ontwikkeling van zelfrijdende auto's gestimuleerd door, onder andere, de EU.[14] Potentiële risico's zijn echter ook aan de orde en bevinden zich op het terrein van publieke waarden en belangen als bescherming van persoonsgegevens, non-discriminatie en rechtsbescherming voor slachtoffers van ongevallen met zelfrijdende auto's. Al deze aspecten zullen in deze casestudy aan bod komen.

Om een juridische toetsing mogelijk te maken, geeft de casestudy allereerst een beknopt overzicht van de algoritmische besluitvorming in (deels) zelfrijdende auto's. Daartoe wordt enerzijds aandacht besteed aan de (typen) algoritmen die nu worden gebruikt en anderzijds aan de

[12] Timmer & Kool 2014, p. 11.
[13] Gawron e.a., *Environmental Science & Technology* 2018, p. 3249-3256.
[14] Vgl. Europese Commissie 2018.

mogelijke ontwikkelingen die de komende vijf tot tien jaar kunnen worden verwacht (par. 5.2). In de paragrafen daarna (par. 5.3-5.7) staan de kansen en risico's voor publieke waarden en belangen centraal, alsmede de bestendigheid van het juridisch kader in dit verband.

5.2 Beslissingsalgoritmen, sensoriek en communicatietechnologie

In lijn met het centrale onderwerp van deze studie, algoritmische besluitvorming, onderzoekt deze casestudy de besluitvorming van de zogenaamde 'beslissingsalgoritmen' van de zelfrijdende auto. Hiermee wordt bedoeld: het (samenspel van) algoritme(n) dat het uiteindelijke besluit neemt omtrent de rij-actie van de auto. Gaat de auto remmen, inhalen, uitwijken? De besluiten die de beslissingsalgoritmen nemen, zijn besluiten binnen de definitie van dit rapport omdat zij potentieel de belangen van een rechtssubject raken. Hiervan is bijvoorbeeld sprake wanneer de beslissingsalgoritmen het besluit nemen om niet uit te wijken voor een andere weggebruiker, met een ongeval tot gevolg.

5.2.1 De werking van beslissingsalgoritmen

De beslissingsalgoritmen worden door de auto van *input* (ook wel: invoer) voorzien ten behoeve van het te nemen besluit. Deze kan grofweg worden ingedeeld in twee typen: enerzijds de invoer die wordt geleverd door de sensoriek van de auto zelf en anderzijds de invoer die wordt geleverd door de zogenaamde V2X-communicatietechnologie. Het beslissingsalgoritme verwerkt en interpreteert de aangeleverde informatie vervolgens en neemt een besluit omtrent wat de auto gaat doen (bijvoorbeeld: al dan niet sturen, remmen of versnellen).[15]

Sensoriek

Een (deels) zelfrijdende auto heeft in de regel drie typen sensoren: i) laserscanners of lidars, ii) camera's en iii) radars.[16] Deze sensoren kunnen bijvoorbeeld de locatie van andere weggebruikers waarnemen, maar ook wat er op de verkeersborden langs de kant van de weg staat en of een verkeerslicht op groen of op rood staat. Datgene dat door de sensoren wordt waargenomen, wordt verwerkt door algoritmen. De *output* (ook wel: uitvoer) van deze algoritmen, bijvoorbeeld een kaart met posities van andere weggebruikers, wordt als invoer geleverd aan het beslissingsalgoritme, dat vervolgens de beslissing neemt omtrent de volgende rij-actie van de auto.[17] Het kan daarbij nuttig zijn om de invoer die door de sensoren van de auto wordt gegenereerd, aan te vullen door invoer die door de V2X-communicatietechnologie wordt gegenereerd.[18] Niet alle sensoren kunnen

[15] Surden & Williams, *Cardozo Law Review* 2016, p. 147.
[16] Van de Weijer, *De Volkskrant* 30 november 2018.
[17] Surden & Williams, *Cardozo Law Review* 2016, p. 137 (met verwijzingen).
[18] In een recent rapport van de Europese adviesraad voor wegtransport wordt tevens geconstateerd dat de *environmental perception* van de zelfrijdende auto wordt beperkt door de *range and capability* van de sensoren in de auto. De conclusie die hier in het rapport aan wordt verbonden, is dat het nuttig is om de informatie die wordt geleverd door de sensoren aan te vullen met informatie die wordt gegenereerd door middel van communicatie, bijvoorbeeld met de infrastructuur, om zo

bijvoorbeeld (goed genoeg) door fysieke objecten zoals gebouwen of bomen heen kijken. Ook kunnen objecten (nog) te ver weg zijn voor de sensoren om waar te nemen of zich juist in de dode hoek vlak naast de auto bevinden.[19]

V2X-communicatietechnologie

De term V2X (*vehicle-to-everything*) communicatietechnologie is de verzamelnaam voor technologie die voorziet in communicatie met andere gemotoriseerde weggebruikers (V2V, *vehicle-to-vehicle*), niet-gemotoriseerde weggebruikers (V2P, *vehicle-to-pedestrians*)[20] en de infrastructuur (V2I, *vehicle-to-infrastructure*). V2X-communicatietechnologie is een waardevolle aanvulling op de sensoren, omdat deze bewerkstelligt dat een beter beeld van de omgeving kan worden gevormd. Daarnaast is een grote toegevoegde waarde van het aanvullen van de sensorinformatie met V2X-informatie dat de auto niet alleen beter begrijpt hoe de omgeving er op dit moment uitziet, maar ook hoe de omgeving gaat veranderen in de komende paar seconden.[21] Dit laat zich goed illustreren aan de hand van het volgende voorbeeld. Stel, auto A rijdt over een weg en auto B komt om de hoek aangereden. Als auto A communiceert met een camera langs de kant van de weg (lees: de infrastructuur) of met auto B zelf, dan kan via beide routes worden gewaarschuwd voor de komst van auto B, terwijl de sensoren deze (nog) niet kunnen waarnemen.

5.2.2 Huidige toepassing

Nu de wisselwerking tussen de beslissingsalgoritmen, de sensoriek en de V2X-communicatietechnologie van (deels) zelfrijdende auto's in kaart is gebracht, wordt ingezoomd op de huidige toepassing hiervan in personenauto's. De toepassing van beslissingsalgoritmen manifesteert zich momenteel met name in zogenaamde *advanced driver assistance systems* in de eerste twee niveaus van automatisering, zoals hieronder zal worden beschreven. Ontwikkelaars experimenteren daarnaast momenteel volop met verdergaande automatisering, waardoor de 'huidige' toepassing niet helemaal scherp af te grenzen is van de 'toekomstige' toepassing. De experimenten illustreren wat in de toekomst mogelijk zal worden; de bespreking van lopende experimenten kan dan ook worden beschouwd als brug tussen de huidige en toekomstige mogelijkheden.

betere beslissingen te kunnen nemen. Zie ERTRAC Working Group "Connectivity and Automated Driving" 2019, p. 7 en 10. Zie tevens: *Declaration of Amsterdam on cooperation in the field of connected and automated driving* (Verklaring voor de gezamenlijke onderneming van zelfrijdende voertuigen, op 14 april 2016 door de EU-lidstaten aangenomen onder Nederlands voorzitterschap), bijlage bij *Kamerstukken II* 2015/16, 21501-33, nr. 592; Europese Commissie 2018, p. 3.

[19] Vgl. Yang e.a., *Science China Technological Sciences* 2018, p. 1454-1457.

[20] Bijvoorbeeld via communicatie met smartphones. Zie bijvoorbeeld: Jeltes, cursor.tue.nl 13 september 2018.

[21] In de literatuur wordt gesteld dat voor zelfrijdende auto's, '*[d]ecision problems in the real world are aggravated by incomplete and noisy perception and uncertain knowledge about how the world evolves over time*'. Voor zelfrijdende auto's betekent dit aldus, aan de ene kant, dat de intenties van andere verkeersdeelnemers onduidelijk kunnen zijn en hun gedrag niet met zekerheid kan worden voorspeld. Aan de andere kant zijn de waarnemingen van de sensoren gevoelig voor fouten, nu niet de gehele omgeving zichtbaar is en bovendien snel verandert. V2X-communicatietechnologie kan deze onzekerheden reduceren. Zie Brechtel, Gindele & Dillmann 2014. Vgl. Hobert e.a., *IEEE Communications Magazine* 2015, p. 64; Hecker e.a., *Journal of Communications* 2011, p. 115; Yang e.a., *Science China Technological Sciences* 2018, p. 1454-1457.

Advanced driver assistance systems

Volgens een recent radioreclamespotje zitten auto's tegenwoordig vol slimme rijhulpsystemen zoals '*lane assist*' en 'file assistent' die werk overnemen van de bestuurder.[22] Deze reclame schetst een aantal *advanced driver assistance systems* opties, waarmee auto's nu al zelfstandig rijbeslissingen nemen. Andere voorbeelden van 'rijhulpsystemen die autonoom rijden straks mogelijk moeten maken' – een stelling die de samenloop van de huidige en toekomstige toepassing illustreert – zijn adaptieve cruisecontrol en autonome noodstopsystemen.[23] Vanuit het perspectief van het beslissingsalgoritme is een rode draad in deze toepassingen dat de auto operationele rijtaken (zoals sturen, remmen en versnellen) zelfstandig uitvoert. Overige dynamische rijtaken moet de bestuurder zelf uitvoeren.[24] Deze toepassingen van automatisering die al daadwerkelijk op de weg in gebruik zijn, worden over het algemeen geschaard onder de eerste twee niveaus van automatisering.[25]

De hierbij gebruikte beslissingsalgoritmen zijn binnen de huidige toepassingen vaak regelgebaseerd.[26] Op grond van beslisbomen kan duidelijk worden uitgelegd waarom de auto een beslissing heeft genomen. Dit is van belang voor de voorspelbaarheid en uitlegbaarheid van de algoritmen. Een voorbeeld kan dit verduidelijken. Bij adaptieve cruisecontrol anticipeert de auto middels een afstandsradar en snelheidsregelaar automatisch op het tempo van de voorganger.[27] De uitvoer van het algoritme is zeer voorspelbaar en uitlegbaar. Immers, gegeven bepaalde invoer omtrent snelheid en afstand van de voorganger, kan men vooraf een bepaalde uitkomst verwachten.[28] De uitlegbaarheid hangt hier sterk mee samen. Wat betreft zelfstandigheid, rijdt het systeem automatisch een bepaalde snelheid waarbij menselijk ingrijpen niet nodig is (maar wel mogelijk). Het systeem geeft daarmee richting ten aanzien van het te nemen besluit, maar dat gegenereerde advies wordt niet altijd gevolgd.[29]

5.2.3 Blik op de toekomst

De regelgebaseerde beslissingsalgoritmen in (deels) zelfrijdende auto's zullen in de toekomst steeds meer plaats maken voor *machine learning* beslissingsalgoritmen.[30] *Machine learning* wordt vaak omschreven als een *black box*,[31] omdat de manier waarop dit type algoritme een beslissing

[22] Vgl. volkswagen.nl/modellen/passat.
[23] Surden & Williams, *Cardozo Law Review* 2016, p. 134 (met verwijzingen).
[24] Vgl. de beschrijving van de automatiseringsniveaus in par. 5.1.
[25] Zo stelt de ANWB dat de meest moderne voertuigen inmiddels voldoen aan automatiseringsniveau 2. Zie anwb.nl/auto/zelfrijdende-auto/wat-is-de-zelfrijdende-auto. In een rapport uit 2017 van de *European Road Transport Research Advisory Council* (ERTRAC) wordt hetzelfde geconcludeerd. Zie ERTRAC Working Group "Connectivity and Automated Driving" 2017, p. 6-7 en 13.
[26] Onderkend moet worden dat binnen de huidige toepassingen tot op zekere hoogte ook al gebruik wordt gemaakt van *machine learning*-algoritmen. Zie in dit verband bijvoorbeeld: Fehrenbacher, FORTUNE.com 16 oktober 2015.
[27] Vgl. de uitleg op anwb.nl/auto/zelfrijdende-auto/wat-is-de-zelfrijdende-auto.
[28] Hiervoor is wel vereist dat de variabelen en de door het algoritme toe te passen regels bekend zijn.
[29] Vgl. Calvert e.a., *IEEE Intelligent Transportation Systems Magazine* 2020.
[30] Vgl. Stilgoe, *Social Studies of Science* 2018, p. 29.
[31] Holstein, Dodig-Crnkovic & Pelliccione 2018.

neemt (zelfs voor de programmeur) relatief complex en abstract is.[32] Dit wordt ook wel het *comprehensibility principle* genoemd.[33] De voorspelbaarheid en uitlegbaarheid van de uitvoer, oftewel de rijactie waartoe het beslissingsalgoritme overgaat, is daarmee in de regel laag.[34]

Experimenten met (deels) zelfrijdende auto's

De toekomstige toepassing van de beslissingsalgoritmen in zelfrijdende auto's laat zich het best illustreren aan de hand van experimenten met verdergaande niveaus van automatisering, die nog geen gemeengoed zijn op de weg.[35] De bespreking is niet uitputtend bedoeld.[36] Beoogd wordt rode draden in kaart te brengen, vooral met betrekking tot de mogelijke gevolgen voor de voorspelbaarheid, uitlegbaarheid en zelfstandigheid van de beslissingsalgoritmen.

Concrete voorbeelden van experimenten volgen onder andere uit het mede door de EU gefinancierde project AUTOPILOT. Een van de partijen die bij dit project betrokken is, is TNO. Op de Nederlandse testlocatie Brainport experimenteert TNO met verschillende *driving modes*: i) *platooning,* waarbij een peloton van in hoge mate geautomatiseerde voertuigen een *lead vehicle* volgt op basis van V2V-communicatie, al dan niet op een specifiek aangeduide rijstrook; ii) *highway pilot*, waar een cloudservice invoer van sensoren samenbrengt om achteropkomende auto's te waarschuwen voor gevaren op de weg; iii) *automated valet parking*, waarbij de bestuurder de auto kan achterlaten op een *predefined drop-off location* om hem later weer op te halen; en iv) *urban driving*, waar het identificeer-, voorspellings- en reactievermogen van volledig geautomatiseerde voertuigen in complexe situaties wordt getest, in beginsel zonder handelen van de bestuurder (die overigens wel op elk moment kan overschakelen naar handmatig rijden).[37] Binnen de eerste drie AUTOPILOT-experimenten, waarin de gebieden en manieren van rijden van tevoren zijn

[32] Surden & Williams, *Cardozo Law Review* 2016, p. 162.
[33] Surden & Williams, *Cardozo Law Review* 2016, p. 162 (met verwijzing).
[34] Vgl. Baehrens e.a., *Journal of Machine Learning Research* 2010, p. 1803; Stilgoe, *Social Studies of Science* 2018, p. 30.
[35] Wat betreft auto's met automatiseringsniveau 3 zijn er schattingen dat deze tussen 2021-2025 hun intrede zouden kunnen doen. Zie onder andere: ERTRAC Working Group "Connectivity and Automated Driving" 2019, p. 12-13. Experimenten in gecontroleerde omgevingen met niveau 4 zijn onderweg, en voor niveau 5 geldt dat de schattingen (sterk) uiteenlopen. Zo wordt 2030 genoemd, waarmee deze ontwikkeling binnen de reikwijdte van dit onderzoek zou vallen, maar ook 2050. Zie onder andere: Gavrila (TU Delft) zoals geciteerd in Van de Weijer, *De Volkskrant* 30 november 2018; ERTRAC Working Group "Connectivity and Automated Driving" 2019, p. 4. Vooral de introductie van volledig zelfrijdende auto's in stedelijke gebieden blijft onzeker. Critici betwijfelen of dit überhaupt een realiteit zal worden in de nabije toekomst. Ondanks de meest geavanceerde technieken zou het gedrag van andere verkeersdeelnemers namelijk moeilijk voorspelbaar blijven, wat gevolgen heeft voor de betrouwbaarheid van de besluitvorming. Vgl. J. Ploeg (TU Eindhoven) zoals geciteerd op ntr.nl/site/nieuws/Zelfrijdende-auto-kansloos-in-de-stad/409; het onderzoek van SWOV en RAI Vereniging zoals genoemd op swov.nl/nieuws/volledig-zelfstandige-auto-laat-nog-jaren-op-zich-wachten; Van de Weijer, *De Volkskrant* 31 oktober 2019 en Van de Weijer, *De Volkskrant* 27 december 2019.
[36] Veel technische literatuur betreft namelijk een specifieke toepassing of een ontwikkeld model of algoritme dat getest wordt, terwijl er door de enorme diversiteit in voertuigen, sensoren en toepassingen niet een *one-size-fits-all* toepassing is. Vgl. Elfring e.a., *Sensors* 2016, p. 1668 en 1690.
[37] Zie autopilot-project.eu/pilot-sites/brainport-nl en autopilot-project.eu/pilot-sites/driving-modes. De EU financiert momenteel meerdere grote samenwerkingsprojecten en tests, waaronder in automatiseringsniveau 3, met het oog op het uitrollen van deze toepassingen in Europa. Een ander voorbeeld is het L3Pilot project, dat tests doet met '*the viability of automated driving as a safe and efficient means of transportation on public roads. It will focus on large-scale piloting of SAE Level 3 functions, with additional assessment of some Level 4 functions. […] The tests will provide valuable data for evaluating technical aspects, user acceptance, driving and travel behaviour, as well as impact on traffic and safety*', zie l3pilot.eu.

vastgesteld, is sprake van een relatief hoge voorspelbaarheid en uitlegbaarheid tegenover een relatief lage mate van zelfstandigheid van de algoritmische besluitvorming. Het experiment omtrent *urban driving* daarentegen draait om een volledig autonoom voertuig. In dit verband is eerder opgemerkt dat de voorspelbaarheid en uitlegbaarheid van *machine learning* beslissingsalgoritmen in (vergaand) autonome voertuigen laag, terwijl de zelfstandigheid hoog is.

Andere interessante voorbeelden zijn de experimenten van de TU Delft en grote fabrikanten als Bosch en Daimler. De rode draad in de onderliggende technologie bij deze experimenten is het gebruik van sensorfusie in combinatie met *machine learning*. Sensorfusie ziet op het samenvoegen van data uit verschillende sensoren om zo de beperkingen van elke sensor te ondervangen. Dit kan op twee manieren worden gedaan: ofwel door de data van alle sensoren samen te voegen en dit te laten interpreteren door de *machine learning*-algoritmen ('vroege sensorfusie'), ofwel door de data van elke sensor apart te laten verwerken en classificeren ('late sensorfusie').[38] Bij late sensorfusie zijn de beslissingen van de auto achteraf vrij goed door de mens te verklaren, hetgeen bijdraagt aan de uitlegbaarheid van de beslissingsalgoritmen. Vroege sensorfusie heeft daarentegen het karakter van een *black box*: we weten niet wat er precies in het systeem gebeurt en hoe het tot zijn beslissingen komt.[39]

5.2.4 Tussenconclusie

In deze paragraaf is allereerst de wisselwerking tussen de beslissingsalgoritmen, de sensoriek en de V2X-communicatietechnologie van (deels) zelfrijdende auto's in kaart gebracht, waarna is ingezoomd op de huidige toepassing hiervan in personenauto's. Hoewel *de* zelfrijdende auto er nog niet is, laten de besproken toepassingen van voertuigautomatisering zien dat personenauto's in steeds verdergaande mate zelfstandig worden.

In de toekomst zullen auto's niet langer alleen 'eenvoudige' handelingen uitvoeren. Als gevolg hiervan zullen de regelgebaseerde beslissingsalgoritmen overwegend plaats maken voor *machine learning* beslissingsalgoritmen. Dit betekent dat de voorspelbaarheid en uitlegbaarheid van de algoritmen in beginsel laag zullen zijn,[40] hetgeen mogelijk gevolgen heeft voor de publieke waarden en belangen die in dit onderzoek centraal staan. De volgende paragrafen richten zich op de kansen en risico's in dit verband, alsmede op (de bestendigheid van) het juridisch kader.[41]

[38] Van de Weijer, *De Volkskrant* 30 november 2018.
[39] Gavrila (TU Delft) zoals geciteerd in Van de Weijer, *De Volkskrant* 30 november 2018.
[40] Hetgeen overigens genuanceerd kan worden door bijvoorbeeld het gebruik van late sensorfusie.
[41] Vgl. 'Autonomous Road Vehicle', breakthrough.unglobalcompact.org/disruptive-technologies/autonomous-road-vehicles, 11 juli 2017.

5.3 Bescherming van persoonsgegevens

Onder 5.2.1 is de V2X-communicatietechnologie, die invoer kan leveren aan de beslissingsalgoritmen in (deels) zelfrijdende auto's, geïntroduceerd en toegelicht. Het gebruik van V2X-communicatietechnologie in (deels) zelfrijdende auto's brengt mogelijk risico's met zich mee voor de bescherming van persoonsgegevens,[42] de eerste casusoverstijgende waarde die in dit onderzoek centraal staat.

5.3.1 Bescherming van persoonsgegevens en (deels) zelfrijdende auto's

Het begrip persoonsgegevens wordt gedefinieerd als 'alle informatie[43] over een geïdentificeerde of identificeerbare natuurlijke persoon', waarbij als identificeerbaar wordt beschouwd 'een natuurlijke persoon die direct of indirect kan worden geïdentificeerd'.[44] Met andere woorden: persoonsgegevens zijn gegevens die op directe of indirecte wijze een specifiek individu kunnen identificeren. Zoals in hoofdstuk 3 van dit rapport reeds naar voren is gekomen, worden persoonsgegevens (als onderdeel van tot de persoonlijke levenssfeer behorende informationele privacy) beschermd, zowel door grondrechtelijke instrumenten alsook door uitwerkingen daarvan in meer specifieke instrumenten zoals de Algemene Verordening Gegevensbescherming (AVG).

De (deels) zelfrijdende auto kan, onder andere als gevolg van de inzet van V2X-communicatietechnologieën, enorme hoeveelheden gegevens verzamelen, gebruiken, opslaan en verzenden.[45] Deze gegevens zullen vaak op de een of andere manier verband houden met een natuurlijk persoon. Een voorbeeld hiervan zijn locatiegegevens.[46] Locatiegegevens, al dan niet *real-time*, kunnen kwalificeren als persoonsgegevens.[47] Informatie over de plek waar een zelfrijdende auto zich op een bepaald moment bevindt of bevond, is immers in principe herleidbaar tot een persoon, met name wanneer deze informatie wordt gecombineerd met andere gegevens.[48] Een tweede voorbeeld in dit verband zijn gegevens over de rijstijl van de chauffeur. Deze gegevens kwalificeren als persoonsgegevens wanneer deze kunnen worden gekoppeld aan een persoon, bijvoorbeeld

[42] Fagnant & Kockelman, *Transportation Research Part A* 2015, p. 178; De Bruin, *European Journal of Risk Regulation* 2016, p. 495; Jadaan, Zeater & Abukhalil, *Procedia Engineering* 2017, p. 644.
[43] Ongeacht inhoud (privé dan wel publiek), aard (bijvoorbeeld feiten, meningen, maar ook onwaarheden), vorm (bijvoorbeeld papier, digitaal, beeld of geluid) dan wel bron (bijvoorbeeld de betrokkene, eigen waarneming of publiek). Zie Groep gegevensbescherming artikel 29 2007, p. 6-9.
[44] Art. 4 lid 1 AVG. Locatiegegevens worden in dit artikel expliciet genoemd als voorbeeld van een identificator aan de hand waarvan natuurlijke personen kunnen worden geïdentificeerd.
[45] Peppet, *Texas Law Review* 2014, p. 104-108; Boeglin, *Yale Journal of Law and Technology* 2015, p. 181; Gaeta, *Diritto Mercato Tecnologia* 2017, p. 7 (met verwijzing); Hacker, *International Data Privacy Law* 2017, p. 268; Vgl. European Data Protection Board 2020, p. 11.
[46] Surden & Williams, *Cardozo Law Review* 2016, p. 121 en 142;
[47] Zie met betrekking tot locatiegegevens EHRM 2 september 2010, ECLI:CE:ECHR:2010:0902JUD003562305, (*Uzun t. Duitsland*) en EHRM 8 februari 2018, ECLI:CE:ECHR:2018: 0208JUD003144612 (*Ben Faiza t. Frankrijk*).
[48] Vgl. De Bruin, *European Journal of Risk Regulation* 2016, p. 498. *Gevoelige* persoonsgegevens (in de zin van art. 9 AVG) zijn locatiegegevens in beginsel overigens niet. Zie in dit verband: Griffioen 2011, p. 35.

door via het kenteken of serienummer van de auto te achterhalen wie de eigenaar van de auto is, of wie de auto huurde.[49]

5.3.2 Kansen, risico's en bestendigheid juridisch kader

Hierboven werd geconstateerd dat een deel van de gegevens die zelfrijdende auto's verzamelen, gebruiken, opslaan en verzenden kunnen kwalificeren als persoonsgegevens. Wanneer dit het geval is, is de AVG het relevante juridisch kader.[50] De AVG schrijft voor dat persoonsgegevens moeten worden verwerkt met inachtneming van een reeks beginselen, waarbij de plicht tot naleving van deze beginselen rust op de partij die als 'verwerkingsverantwoordelijke' kan worden aangemerkt.[51] Het betreft achtereenvolgens de beginselen van rechtmatigheid, behoorlijkheid en transparantie; van doelbinding; van minimale gegevensverwerking; van juistheid; en van integriteit en vertrouwelijkheid.[52]

In het licht van deze beginselen moet bij de ontwikkeling van technologieën door middel waarvan persoonsgegevens kunnen worden verwerkt, rekening worden gehouden met de noties van gegevensbescherming door ontwerp (ook wel *privacy by design*) en gegevensbescherming door standaardinstellingen (ook wel *privacy by default*).[53] Dit houdt in dat al in de ontwerpfase van de (deels) zelfrijdende auto zorg moeten worden gedragen voor een minimale inbreuk op de persoonlijke levenssfeer bij de verwerkingsactiviteiten.[54] Bovendien is de verwerkingsverantwoordelijke verplicht om, wanneer een voorgenomen verwerking waarschijnlijk een hoog risico inhoudt voor de rechten en vrijheden van natuurlijke personen, een zogenaamde gegevensbeschermingseffectbeoordeling (ook wel *Data Protection Impact Assessment* of DPIA) uit te voeren.[55] Dit is met name relevant voor verwerkingen waarbij nieuwe technologieën worden gebruikt.[56] De toezichthoudende autoriteit (in Nederland de Autoriteit Persoonsgegevens (AP)) stelt in dit verband een lijst op van het soort verwerkingen waarvoor een DPIA verplicht is.[57] Grootschalige verwerkingen en/of stelselmatige monitoring van locatiegegevens, waarbij auto's expliciet als voorbeeld worden genoemd, maken in Nederland onderdeel uit van desbetreffende lijst.[58]

[49] European Data Protection Board 2020, p. 7; Autoriteit Persoonsgegevens 2020, p. 2.
[50] Immers, de AVG is blijkens art. 2 lid 1 AVG van toepassing wanneer sprake is van (geheel of gedeeltelijk geautomatiseerde) verwerking van persoonsgegevens. Het begrip 'verwerking' is blijkens de definitie in art. 4 lid 2 AVG erg ruim, en omvat mede verzamelen, gebruiken, opslaan en verzenden.
[51] Artt. 4 lid 7 jo. 5 lid 2 jo. 24 AVG.
[52] Art. 5 lid 1 AVG.
[53] Art. 25 AVG. Vgl. European Data Protection Board 2020, p. 4.
[54] Engelhard & De Bruin 2018, p. 55. Vgl. Prins, *NJBlog* 29 september 2015; Schermer, Hagenauw & Falot 2018, p. 61.
[55] Art. 35 AVG.
[56] Art. 35 lid 1 AVG.
[57] Art. 35 lid 4 AVG.
[58] Besluit inzake lijst van verwerkingen van persoonsgegevens waarvoor een gegevensbeschermingseffectbeoordeling (DPIA) verplicht is, Autoriteit Persoonsgegevens (*Stcrt.* 2019, 64418).

Zoals gezegd rust de plicht tot naleving van het bovenstaande op de partij die als 'verwerkingsverantwoordelijke' in de zin van de AVG kan worden aangemerkt.[59] Producenten van (deels) zelfrijdende auto's kunnen worden aangemerkt als verwerkingsverantwoordelijke wanneer zij het doel en de middelen van de verwerking van persoonsgegevens bepalen.[60] Dit is bijvoorbeeld het geval wanneer zij bepalen welke persoonsgegevens worden verwerkt en wanneer en hoe deze worden opgeslagen, hetgeen in de regel het geval is.[61] Wanneer de producent in dat geval voor de verwerking van persoonsgegevens gebruikmaakt van software of hardware die door een derde partij is geproduceerd, hetgeen vaak aan de orde is, kwalificeert deze derde niet (tevens) als verwerkingsverantwoordelijke. Dit vormt een potentieel knelpunt in het licht van de bescherming van persoonsgegevens omdat deze derde partij verreweg als beste (of misschien wel: als enige) in staat is om adequate *privacy by design* maatregelen te implementeren.[62]

Dit knelpunt doet zich bijvoorbeeld voor in het licht van het volgende. Zelfrijdende auto's zijn, met name wanneer zij gebruikmaken van V2X-communicatietechnologie, kwetsbaar voor hacks.[63] Het begrip hack kan in dit verband worden gedefinieerd als iedere activiteit waarbij de technologie van de (deels) zelfrijdende auto wordt gebruikt op een manier waarvoor deze niet is ontworpen door de producent.[64] Wanneer een (deels) zelfrijdende auto wordt gehackt, kan dit uiteraard acute problemen opleveren met betrekking tot de fysieke veiligheid van zowel inzittenden als personen in de directe omgeving van de auto indien, bijvoorbeeld, de besturing kan worden overgenomen.[65] In het licht van de bescherming van persoonsgegevens kan een hack echter eveneens grote gevolgen hebben. Een hacker verkrijgt immers mogelijk toegang tot de persoonsgegevens die de (deels) zelfrijdende auto verzamelt, gebruikt en opslaat.[66] Een hack van een zelfrijdende auto kwalificeert daarmee naar alle waarschijnlijkheid als een 'datalek'.[67] Wanneer de producent kan worden aangemerkt als verwerkingsverantwoordelijke, is hij verantwoordelijk voor het waarborgen van de integriteit en vertrouwelijkheid van de persoonsgegevens en dus ook voor het voorkomen van datalekken.[68] In dit verband is echter goed denkbaar dat de producent van de (deels)

[59] Artt. 4 lid 7 jo. 5 lid 2 jo. 24 AVG.
[60] Art. 4 lid 7 AVG.
[61] Engelhard & De Bruin 2018, p. 54.
[62] Borking 2010, p. 389; Klitou 2012, p. 281.
[63] Glancy, *Santa Clara Law Review* 2012, p. 1196-1197; Boeglin, *Yale Journal of Law and Technology* 2015, p. 181; Schellekens, *Computer Law & Security Review* 2016, p. 310.
[64] Engelhard & De Bruin 2018, p. 49 (met verwijzing).
[65] Zie voor enkele tot de verbeelding sprekende voorbeelden Prins, *NJBlog* 29 september 2015. Zie bijvoorbeeld ook Parkinson e.a., *IEEE Transactions on Intelligent Transportation Systems* 2017, p. 2898-2915; Raiyn, *Transport and Telecommunication* 2018, p. 325-334; European Union Agency for Cybersecurity 2019.
[66] Engelhard & De Bruin 2018, p. 50.
[67] Een 'inbreuk in verband met persoonsgegevens', in art. 4 lid 12 AVG gedefinieerd als een inbreuk op de beveiliging die [...] op onrechtmatige wijze leidt tot de vernietiging, het verlies, de wijziging of de ongeoorloofde verstrekking van of de ongeoorloofde toegang tot doorgezonden, opgeslagen of anderszins verwerkte gegevens. De verwerkingsverantwoordelijke heeft in dat geval in beginsel de plicht het 'datalek' te melden bij de toezichthouder op grond van art. 33 AVG. Hetzelfde geldt voor, bijvoorbeeld, een hack van de servers van de producent, indien daar (ook) persoonsgegevens worden opgeslagen.
[68] De verwerkingsverantwoordelijke moet immers op grond van art. 5 lid 1 onder f AVG passende technische of organisatorische maatregelen nemen zodat de beveiliging van persoonsgegevens gewaarborgd is en zodat deze, onder meer, beschermd zijn tegen ongeoorloofde of onrechtmatige verwerking en verlies. Art. 32 lid 1 AVG biedt in dit verband

zelfrijdende auto niet de meest geschikte partij is om de benodigde maatregelen te implementeren, zoals *anonymisation by design* of *pseudonymisation by design*,[69] terwijl hij de enige is die hier op grond van de AVG verantwoordelijk voor is.[70]

5.3.3 Tussenconclusie

Wanneer zelfrijdende auto's persoonsgegevens verwerken, moet rekening worden gehouden met de beginselen uit de AVG. In dit verband moet in de ontwerpfase van technologieën door middel waarvan persoonsgegevens kunnen worden verwerkt reeds rekening worden gehouden met, ondere andere, de notie van *privacy by design*.[71] Aandachtspunt hierbij is dat de partij die verantwoordelijk is voor de naleving van de AVG veelal niet de partij is die in staat is om adequate *privacy by design* maatregelen te implementeren. Immers, de producent zal in de regel verwerkingsverantwoordelijke zijn maar deze zal vaak voor de verwerking van persoonsgegevens gebruikmaken van software of hardware die door een derde partij is geproduceerd. Deze derde kwalificeert niet (tevens) als verwerkingsverantwoordelijke, maar is wel de partij die als beste (of misschien wel: als enige) in staat is om adequate *privacy by design* maatregelen te implementeren (zoals *pseudonymisation by design*).[72] Het is vooralsnog onduidelijk of dit gesignaleerde risico door middel van een brede interpretatie van de definitie van 'verwerkingsverantwoordelijke' uit art. 4(7) AVG kan worden gemitigeerd of niet.

5.4 Non-discriminatie

Ook voor de tweede casusoverstijgende waarde van dit onderzoek, non-discriminatie, hangen de kansen en risico's sterk samen met de technische werking van (deels) zelfrijdende auto's. Vooral de werking van de sensoriek en het daarmee gegenereerde beeld van de omgeving zijn van belang, omdat deze interpretatie van de verzamelde data aan de basis ligt van de besluitvorming.

5.4.1 Non-discriminatie en (deels) zelfrijdende auto's

'Zelfrijdende auto kan gedrag voetgangers voorspellen', aldus een nieuwsbericht van 9 augustus 2019 over onderzoek van de TU Delft naar geheel zelfrijdende auto's. Dit is volgens het

een niet-limitatieve opsomming van denkbare passende beveiligingsmaatregelen, zoals pseudonimisering en versleuteling van persoonsgegevens. Daarnaast laat art. 40 AVG ruimte voor de ontwikkeling van gedragscodes waarmee nader kan worden ingekleurd welke beveiligingsmaatregelen als passend hebben te gelden. Engelhard en De Bruin merken in dit verband op dat er al initiatieven zijn genomen waarbij een voorzet wordt gedaan voor dergelijke gedragscodes, zoals bijvoorbeeld de *Data Protection Principles for Connected Vehicles* van het Duitse Verband der Automobilindustrie en de *ACEA Principles of Data Protection in Relation to Connected Vehicles and Services* van de European Automobile Manufacturers Association. Zie Engelhard & De Bruin 2018, p. 57.

[69] European Data Protection Board 2020, p. 16-17. Vgl. *Resolution on Data Protection in Automated and Connected Vehicles* 2017, p. 3; International Working Group on Data Protection in Telecommunications 2018, p. 11.

[70] Dit knelpunt is, specifiek met betrekking tot het risico van hacks van zelfrijdende auto's, reeds gesignaleerd in een rapport van het Europees Parlement uit 2018. Zie Evas 2018, p. 26. Hierbij is tevens interessant om te noemen dat de Autoriteit Persoonsgegevens slechte beveiliging van de persoongegevens die auto's verzamelen, beschouwt als iets 'waar mensen rekening mee moeten houden'. Zie autoriteitpersoonsgegevens.nl/nl/nieuws/ap-geeft-tips-voor-privacy-bij-connected-cars.

[71] Art. 25 AVG.

[72] Borking 2010, p. 389; Klitou 2012, p. 281.

nieuwsbericht een belangrijke ontwikkeling omdat de zelfrijdende auto mensen tot nu toe moeilijk kon detecteren.[73] Dit geeft een eerste mogelijke implicatie voor het non-discriminatiebeginsel. Immers, wat betekent (grotendeels) zelfstandige besluitvorming van zelfrijdende auto's voor de behandeling van voetgangers ten opzichte van andere verkeersdeelnemers? Deze redenering kan nog een stap verder worden getrokken in het licht van andere nieuwsberichten, volgens welke onderzoek had aangetoond dat de camerasystemen van zelfrijdende auto's mensen met een donkere huidskleur minder goed zouden herkennen. Daardoor zou de auto hen niet als voetgangers registreren, terwijl deze anders anticipeert op voetgangers dan op andere objecten.[74] Indien daarvan daadwerkelijk sprake is, dan is het effect dat de technologie van zelfrijdende auto's al dan niet onbewust leidt tot risico's van ongelijke behandeling van voetgangers *onderling*. Immers, mogelijk zijn niet alle voetgangers even herkenbaar, bijvoorbeeld op basis van (huids)kleur maar ook de mate waarin hun kleding kan worden onderscheiden van de verdere omgeving.[75] Andere mogelijke factoren voor onderscheid in de besluitvorming van zelfrijdende auto's die in de literatuur worden aangevoerd, zijn leeftijd en geslacht.[76]

De vraag die vervolgens opkomt is hoe we dit soort uitkomsten van algoritmische besluitvorming in zelfrijdende auto's moeten duiden in het licht van het non-discriminatiebeginsel. Het gaat hierbij immers niet zozeer om de intentie, maar om het discriminerende effect van besluiten.[77] De betekenis van de publieke waarde non-discriminatie is in hoofdstuk 3 van dit rapport geïntroduceerd op grond van zowel algemene kaders (zoals art. 14 EVRM en art. 1 Gw) als specifieke wetgeving (zoals de AWGB). Die bespreking van, onder andere, de verboden gronden van discriminatie en het onderscheid tussen directe en indirecte discriminatie dient als basis voor de analyse in deze paragraaf, aangevuld met verdere duiding in de context van zelfrijdende auto's.

Meerdere gronden die in de rechtspraak over art. 14 EVRM als 'verdacht' worden aangemerkt zijn relevant voor de context van zelfrijdende auto's. De gronden geslacht, ras en kleur raken mogelijk aan de mate waarin verkeersdeelnemers op gelijke voet worden herkend door de sensoren van zelfrijdende auto's. Ook godsdienst is een relevante grond, zoals hieronder verder zal worden toegelicht in het licht van herkenbaarheid van voetgangers met gezichtsbedekkende kleding.

[73] Van der Parre, nos.nl 9 augustus 2019.
[74] Zie onder andere in dat verband: 'Zelfrijdende auto's herkennen donkere huidskleur minder goed', nu.nl 5 maart 2019. Het onderzoek dat wordt bedoeld is Wilson, Hoffman & Morgenstern 2019.
[75] Zoals hieronder echter genuanceerd zal worden, zijn er ook andere verklaringen die maken dat sensoren gezichten mogelijk minder goed registreren.
[76] Brandão 2019.
[77] Vgl. de bespreking van direct en indirecte discriminatie in hoofdstuk 3 van dit rapport, gebaseerd op Vetzo, Gerards & Nehmelman 2018, p. 87. De ethische dimensie van algoritmische besluitvorming door zelfrijdende auto's, die ziet op dilemma's betreffende hoe te handelen in een bepaalde (verkeers)situatie, blijft in deze casestudy verder buiten beschouwing. Een voorbeeld is de keuze tussen uitwijken voor een overstekende voetganger op leeftijd of een fietser met een kind voorop. Vgl. onder andere Lin 2015; Ethik-Kommission Automatisiertes und Vernetztes Fahren 2017, p. 16-17; Roff, brookings.edu 7 december 2018.

Verder verbiedt art. 14 EVRM onderscheid ‘op welke grond ook’, inclusief ‘andere status’. Dit biedt wellicht een opening voor gronden zoals leeftijd en niet-religieuze gezichtsbedekkende kleding.

5.4.2 Kansen, risico’s en bestendigheid juridisch kader

De technische werking van zelfrijdende auto’s kan bepaalde tekortkomingen vertonen die mogelijk eerder een discriminatoir *effect* hebben dan dat sprake is van het opzettelijk achterstellen van bepaalde mensen door het systeem. Een eerste voorbeeld van mogelijke ongelijke behandeling wordt gevormd door het feit dat niet alle algoritmen die worden gebruikt om de sensoren te interpreteren elke huidskleur even goed herkennen (afhankelijk van de concrete omstandigheden). Iemand met een huidskleur die ‘minder goed reflecteert’ zou vanuit schaduw of in het donker minder goed zichtbaar kunnen zijn voor het systeem van de zelfrijdende auto, terwijl iemand met een huidskleur die ‘goed reflecteert’ mogelijk minder goed wordt geregistreerd in extreem zonnig weer. Dit levert een risico op van discriminatoire effecten van de algoritmische besluitvorming. Een tweede, concreter voorbeeld betreft de camera’s: mogelijk herkennen niet alle algoritmen die worden gebruikt om camerabeelden te interpreteren personen met gezichtsbedekkende kleding (denk aan een burka of muts tot over de wenkbrauwen) even goed als personen zonder gezichtsbedekkende kleding. Ten derde is van belang dat de algoritmen die worden gebruikt om de radar te interpreteren, oftewel ‘de wereld [te tonen] zoals de radar van een robotauto die ziet’,[78] voetgangers niet altijd als zodanig identificeren. Als een voetganger voor een lantaarnpaal wordt aangezien, is het onwenselijke effect dat deze door de auto mogelijk onterecht anders behandeld wordt dan andere verkeersdeelnemers.

Het voorgaande illustreert dat sensoren niet onfeilbaar zijn. Een radar kan meer zien dan een camera, namelijk ook objecten die zich achter andere objecten bevinden. Een camera levert daarentegen gedetailleerdere beelden (maar is wel minder accuraat in het donker of in slecht weer). Elk van de sensoren heeft dus bepaalde voordelen en beperkingen. Zo werd een voetganger die uit de schaduw tevoorschijn stapte over het hoofd gezien bij het fatale ongeluk met een zelfrijdende Uber-auto in de VS in 2018, terwijl de laserscanner of lidar haar had moeten opmerken.[79] Er bestaat dus een groter risico voor verschillende benadering van verschillende voetgangers wanneer alleen een laserscanner of lidar zou worden gebruikt (in plaats van wanneer ook camera’s worden ingezet). Een combinatie van sensoren leidt vermoedelijk tot een grotere kans op een accurate analyse van de omgeving en daarmee bovendien tot een *kans* voor verwezenlijking van het non-discriminatiebeginsel. We kunnen hier in elk geval concluderen dat de risico’s voor discriminatoire uitkomsten afhangen van het precieze samenspel van de sensoren in (deels) zelfrijdende auto’s.

[78] Van de Weijer, *De Volkskrant* 30 november 2018.
[79] Zie Marshall, WIRED.com 31 maart 2018.

De fundamentelere vraag die dan nog openstaat is die naar de oorsprong van mogelijke vooringenomenheid, en wie daarop kan worden aangesproken. Het gaat dan om *bias* in het algoritme en/of de dataset.[80] Specifiek voor zelfrijdende auto's zou dat betekenen dat onderzocht moet worden in hoeverre het gebruik van bepaalde trainingsdata of de wijze waarop het algoritme wordt getraind, resulteert in de verschillende benadering van (in bovengenoemde voorbeelden) voetgangers. Voor wat betreft trainingsdata, die contextafhankelijk zijn, gaat het dan om de wijze waarop systemen leren over de omgeving op grond van enorme datasets van afbeeldingen van verkeerssituaties en -deelnemers. Afhankelijk van de datasets die worden gebruikt om de *machine learning*-algoritmen te trainen, kan een eventuele *bias* geïncorporeerd worden in het besluitvormingsproces.[81] De zogenaamde leeftijds*bias* is hier illustratief: volwassenen en kinderen blijken in verschillende mate te worden gedetecteerd door de sensoren, waardoor leeftijd een grond voor verschillende behandeling kan zijn.[82] Dit is een 'persoonsgebonden kenmerk' waar art. 14 EVRM op ziet.[83] Brandão wijst erop dat een dergelijke *bias* kan ontstaan doordat kinderen minder aanwezig zijn in de foto's van het straatbeeld, onder andere door demografische redenen (percentueel gezien is hun aandeel in de bevolking lager) en doordat ze op school zitten tijdens werkuren.[84] Het voorgaande wijst op een '*trade-off between efficiency and fairness*'.[85]

Onderkend moet worden dat ook in de huidige situatie, waarin menselijke bestuurders beslissingen nemen, de verminderde zichtbaarheid van voetgangers in verschillende (weers)omstandigheden kan leiden tot aanrijdingen en dat de problematiek van ontoereikende perceptie van de omgeving in die zin niet nieuw is.[86] Wat wel nieuw is, is de verwachting dat zelfrijdende auto's de *human error* uit de besluitvorming zullen halen. Hoewel in theorie dus kansen bestaan voor verwezenlijking van het non-discriminatiebeginsel doordat de algoritmische besluitvorming de objectiviteit ten goede zou kunnen komen, blijkt dat er bij zelfrijdende auto's een aantal risico's bestaat dat aandacht verdient. Dit betreft bijvoorbeeld over- of ondervertegenwoordiging van bepaalde persoonskenmerken in de data waarmee de algoritmen worden getraind om objecten te classificeren en als voetganger te identificeren (en vervolgens rijbeslissingen te nemen). Als de concrete opzet van beslissingsalgoritmen het effect zal hebben van benadeling van de ene persoon of groep ten opzichte van andere, is de vraag of dat gerechtvaardigd is.[87]

[80] Vgl. Vetzo, Gerards & Nehmelman 2018, p. 142-143. Dit is het geval bij indirecte discriminatie.
[81] Vgl. Marshall, WIRED.com 31 maart 2018.
[82] Brandão 2019, p. 3.
[83] Vgl. ook Gerards, *NJCM-Bulletin* 2004, p. 179 en 183.
[84] Brandão 2019, p. 3-4.
[85] Hoewel wordt voorgesteld om de oplossing te zoeken in het genereren van meer balanceerde datasets, wordt tegelijkertijd gewezen op mogelijke privacybezwaren in het geval geslacht, leeftijd en andere (gevoelige) persoonskenmerken herkenbaar zouden zijn op *street-level datasets*. Een alternatief zou zijn om een oplossing te zoeken in de keuze voor de gebruikte algoritmen, nu wordt geconstateerd dat de keuze voor een voetgangersdetectie-algoritme '*value-laden*' is gezien de bijbehorende '*trade-off between the safety for some people and for others*'. Zie Brandão 2019, p. 3-4. Vgl. ook Liu 2016.
[86] Vgl. Calvert e.a., *Theoretical Issues in Ergonomics Science* 2019, p. 9.
[87] Vgl. Vetzo, Gerards & Nehmelman 2018, p. 80.

In het licht van de evaluatie van het juridisch kader komt een aantal vragen naar voren. De eerste vraag is of er, in de context van de werking van zelfrijdende auto's, een rechtvaardiging bestaat voor het optreden van ongelijke behandeling. Vooropgesteld moet worden dat dit meestal niet het geval is als de ongelijke behandeling kan worden teruggevoerd op een 'verdachte grond', zoals onveranderlijke persoonskenmerken (huidskleur, etnische afkomst, geslacht) of persoonskenmerken waarvan niet in redelijkheid kan worden verwacht dat mensen ze wijzigen (zoals godsdienst).[88] Een aantal van deze gronden kwam prominent naar voren in de bespreking van kansen en risico's voor het non-discriminatiebeginsel. Daarnaast kan het gaan om de 'algemene irrelevantie van een kenmerk voor het dagelijks functioneren'.[89] Dat biedt ruimte om minder voor de hand liggende kenmerken die grond kunnen zijn voor onderscheid en die gebracht kunnen worden onder de grond van 'andere status', zoals niet-religieuze hoofdbedekking, aan te merken als verdachte grond. Ook in dit geval kan ongerechtvaardigde discriminatie worden aangenomen. Het huidige juridisch kader regelt dus al relevante gronden voor discriminatie in de context van zelfrijdende auto's en mitigeert in die zin de risico's voor het non-discriminatiebeginsel.

Echter, voordat überhaupt aan de rechtvaardigingsvraag wordt toegekomen, moet worden bezien *of* sprake is van ongelijke behandeling. Het is voor benadeelden echter moeilijk aan te tonen dat sprake is van ongelijke behandeling.[90] Aan de ene kant is het door de eerder besproken lage uitlegbaarheid en voorspelbaarheid van de besluitvorming van zelfrijdende auto's mogelijk niet duidelijk of sprake is van concrete benadeling van de ene persoon ten opzichte van de andere, [91] bijvoorbeeld in geval van een ongeval met letsel tot gevolg: is dat ongeval daadwerkelijk te herleiden tot verdachte persoonskenmerken? Aan de andere kant kan er evengoed sprake zijn van discriminatie in geval van nadeel dat is gebaseerd op een beschermd persoonskenmerk, ook als het niet evident is dat er echt sprake is van een *ongelijke* behandeling. In elk geval vormt de moeilijk te ontdekken en controleren *bias* in de data een risico voor discriminatie, terwijl de kans die objectiviteit biedt voor het non-discriminatiebeginsel mogelijk niet voldoende wordt verwezenlijkt. Deze problematiek wordt niet opgelost in het huidige juridisch kader en vormt daarom een grondrechtelijk aandachtspunt. Een oplossingsrichting die in de literatuur wordt gesuggereerd voor het opsporen en corrigeren van oneerlijke en discriminatoire uitkomsten van algoritmische besluitvorming combineert technische en juridische elementen. Het voorstel van Hacker is om juridische kernconcepten zoals rechtvaardigheid en non-discriminatie te verankeren in de '*code of the digital economy itself*'.[92]

[88] Vgl. Vetzo, Gerards & Nehmelman 2018, p. 82-83.
[89] Vetzo, Gerards & Nehmelman 2018, p. 84.
[90] Vetzo, Gerards & Nehmelman 2018, p. 122.
[91] Vgl. in algemene zin Vetzo, Gerards & Nehmelman 2018, p. 99.
[92] Hacker, *International Data Privacy Law* 2017, p. 266-286.

5.4.3 Tussenconclusie

Afhankelijk van het concrete samenspel van sensoren in zelfrijdende auto's bestaat een risico voor het non-discriminatiebeginsel, ofwel doordat niet alle voetgangers of andere niet-gemotoriseerde weggebruikers even goed herkend worden door de algoritmen, ofwel doordat nadeel ontstaat dat gebaseerd is op een beschermd persoonskenmerk zoals huidskleur of leeftijd. Nu het huidige juridisch kader deze gronden expliciet erkent, worden daarmee handvatten geboden om de risico's voor het non-discriminatiebeginsel in de context van zelfrijdende auto's te adresseren. Ten opzichte van de huidige situatie met menselijke bestuurders, die in bepaalde omstandigheden voetgangers ook minder goed kunnen waarnemen, bestaan kansen voor grotere objectiviteit. De ondoorzichtigheid van de algoritmen leidt daarentegen mogelijk ook tot risico's in het licht van het non-discriminatiebeginsel, bijvoorbeeld als een *bias* in de data moeilijk aan te tonen is.

5.5 Rechtsbescherming

Onder 5.2.1 is reeds naar voren gekomen dat de beslissingsalgoritmen in (deels) zelfrijdende auto's beslissingen nemen over de rij-acties van de auto. In het licht van de laatste casusoverstijgende publieke waarde rechtsbescherming werpt dit belangrijke vragen op, met name wanneer desbetreffende beslissing schade tot gevolg heeft. Met andere woorden: '*we need to ask ourselves who is liable when a driverless vehicle is involved in an accident*'.[93]

5.5.1 Rechtsbescherming en (deels) zelfrijdende auto's

Centraal bij rechtsbescherming in het kader van de algoritmische besluitvorming in (deels) zelfrijdende auto's, staat de noodzaak tot herstel van de situatie die rechtens is wanneer de besluitvorming schade tot gevolg heeft. De vraag is dan welke rechtsmiddelen beschikbaar zijn die kunnen leiden tot herstel van de situatie. Het recht op een effectief rechtsmiddel (art. 6 EVRM jo. 13 EVRM) vereist, onder andere, dat gegarandeerd moet zijn dat een rechtsmiddel uiteindelijk kan leiden tot een bindende uitspraak en tot effectief rechtsherstel. Daarvoor is, in de context van zelfrijdende auto's, een vergoeding van de geleden schade nodig. De centrale vraag is dus: welke partij kan het slachtoffer, dat schade heeft geleden als gevolg van een besluit van een zelfrijdende auto, aanspreken? Het standpunt van de wetgever in dit verband is dat een slachtoffer van een ongeval met een (deels) zelfrijdende auto niet in een nadeliger positie mag komen dan dat het geval zou zijn geweest als er een niet-zelfrijdende auto bij het ongeval betrokken was.[94]

[93] Europese Commissie 2018, p. 2.
[94] *Kamerstukken II* 2017/18, 34838, nr. 3, p. 16. Vgl. Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 34.

5.5.2 Kansen, risico's en bestendigheid juridisch kader

Bezien zal worden in hoeverre het bestaande aansprakelijkheidsrecht zich leent voor toepassing op (deels) zelfrijdende auto's. Daartoe worden de verschillende aansprakelijkheidsgrondslagen die het slachtoffer ten dienste staan, besproken. De vraag welke (potentiële) knelpunten er aan te wijzen zijn, staat daarbij centraal. De term 'bestuurder' wordt in deze subparagraaf gebruikt om de inzittende aan te geven die de rijtaak uitvoert wanneer de auto dit niet doet.

Art. 185 WVW (Verkeersaansprakelijkheid voor ongemotoriseerde slachtoffers)

Het ongemotoriseerde slachtoffer van een ongeval veroorzaakt door een motorrijtuig wordt beschermd door art. 185 WVW. Op grond van het eerste lid van deze bepaling is de eigenaar of houder van een motorrijtuig aansprakelijk voor de met zijn motorrijtuig toegebrachte letselschade of schade aan bezittingen. Dit is (alleen) anders indien sprake is van overmacht. Van overmacht is sprake wanneer het verkeersongeval geheel te wijten is aan gedragingen van het slachtoffer of van derden.[95] Die gedragingen moesten zo onwaarschijnlijk zijn dat de gemotoriseerde verkeersdeelnemer hier naar redelijkheid geen rekening mee hoefde te houden.[96] Wanneer overmacht niet aannemelijk is, is de eigenaar of houder van het motorrijtuig aansprakelijk. Dit laat de vraag naar de omvang van de aansprakelijkheid nog open.[97] Slachtoffers ouder dan veertien jaar krijgen in beginsel minstens 50% van hun schade vergoed.[98] Voor slachtoffers jonger dan veertien jaar is dit in beginsel zelfs 100%.[99] De 'strengheid' van het systeem wordt gerechtvaardigd door de notie van *Betriebsgefahr*: het aan gemotoriseerd verkeer verbonden gevaar, waardoor de eigenaar of de houder van het motorrijtuig sneller aansprakelijk wordt gehouden.[100] Er is in feite sprake van een risicoaansprakelijkheid, met als achterliggende gedachte de bescherming van zwakkere verkeersdeelnemers.[101]

Hoe vertaalt zich dit naar de context van (deels) zelfrijdende auto's? Er is weinig twijfel over de vraag of een (deels) zelfrijdende auto onder de omschrijving van een 'motorrijtuig' valt.[102] Daarnaast is voor aansprakelijkheid op grond van art. 185 WVW niet vereist dat de eigenaar of houder zelf in de auto reed en/of controle had over het voertuig.[103] De bepaling kan dus worden toegepast op auto's binnen alle vijf de niveaus van automatisering. De eigenaar of houder van de zelfrijdende auto is dan zoals gezegd aansprakelijk voor de door de (deels) zelfrijdende auto toegebrachte schade aan een ongemotoriseerde verkeersdeelnemer, tenzij er sprake is van

[95] De Bruin, *European Journal of Risk Regulation* 2016, p. 493.
[96] HR 22 mei 1992, ECLI:NL:HR:1992:ZC0616; HR 16 februari 1996, ECLI:NL:HR:1996:ZC1991; HR 17 november 2000, ECLI:NL:HR:2000:AA8737 en HR 4 mei 2001, ECLI:NL:HR:2001:AB1426.
[97] Hierbij weegt, op grond van art. 6:101 BW, de mate van eigen schuld van het slachtoffer mee.
[98] HR 2 juni 1995, ECLI:NL:HR:1995:ZC1740.
[99] Zie HR 1 juni 1990, ECLI:NL:HR:1990:AB7631; HR 31 mei 1991, ECLI:NL:HR:1991:ZC0253.
[100] Tjong Tjin Tai & Boesten, *NJB* 2016, p. 657; De Bruin, *European Journal of Risk Regulation* 2016, p. 493.
[101] *Kamerstukken II* 1997/98, 25759, nr. 3, p. 6.
[102] Art. 1 lid 1 onder c WVW.
[103] *Kamerstukken II* 2017/18, 34838, nr. 3, p. 3.

overmacht. Technische gebreken in het voertuig zelf, en dus ook gebreken die verband houden met de (in de software opgeslagen) algoritmen, zullen niet gauw als grond voor overmacht worden aangemerkt, zo leert de ervaring.[104] Bovendien kunnen de 100%-regel en de 50%-regel ook de eigenaar of de houder van een (deels) zelfrijdende auto worden tegengeworpen.[105] Het ongemotoriseerde slachtoffer van een ongeval met een (deels) zelfrijdende auto kan zijn schade in beginsel dus verhalen op de eigenaar of houder van desbetreffende auto. Aandachtspunt voor wat betreft de positie van het slachtoffer is wel dat het overmachtscriterium bij een ongeval met een (deels) zelfrijdende auto niet *ruimer* wordt ingevuld dan wanneer sprake is van een niet-zelfrijdende auto.[106]

Hieronder zal in het kader van gemotoriseerde slachtoffers aan de orde komen dat bij auto's met automatiseringsniveau 4 of 5, waarbij de auto de rijtaak in feite zelfstandig uitvoert, aansprakelijkheid van de 'bestuurder' wellicht niet wenselijk of reëel is. Hetzelfde kan worden opgemerkt in het kader van de aansprakelijkheid van de eigenaar of houder van een auto met automatiseringsniveau 4 of 5 bij *on*gemotoriseerde slachtoffers. Het is dan ook denkbaar dat in die gevallen wordt aangesloten bij aansprakelijkheid van de producent (welke hieronder uiteen wordt gezet), en niet bij het huidige regime van art. 185 WVW.

Art. 6:162 BW (Verkeersaansprakelijkheid voor gemotoriseerde slachtoffers)

Gemotoriseerde verkeersdeelnemers die schade hebben geleden door een ander motorrijtuig vallen buiten het bereik van art. 185 WVW. Zij zullen, indien zij hun schade willen verhalen op de bestuurder van de auto die schade heeft veroorzaakt, een beroep moeten doen op art. 6:162 BW. Bij een actie op grond van art. 6:162 BW zal moeten worden aangetoond dat de bestuurder onrechtmatig heeft gehandeld. Voor zover er niet een duidelijke wettelijke plicht (lees: een verkeersregel) is geschonden, zal er aan de zorgvuldigheidsnorm worden getoetst.[107] De bestuurder moet een fout hebben gemaakt, oftewel anders hebben gehandeld dan hij had behoren te doen. Er worden in dit verband hoge eisen gesteld aan de bestuurder,[108] hetgeen ook hier wordt gerechtvaardigd door het *Betriebsgefahr*.[109]

Een geslaagd beroep op art. 6:162 BW vereist een toerekenbaar onrechtmatig doen of nalaten. Wanneer de auto de rijtaak zelfstandig uitvoert, kan men de vraag stellen of hiervan überhaupt sprake kan zijn. Met andere woorden: kan een schadeveroorzakende rijactie van de auto worden

[104] Vellinga, *VR* 2014, p. 372; Schreuder, *AV&S* 2014, p. 134; Tjong Tjin Tai & Boesten, *NJB* 2016, p. 659; Engelhard, *AA* 2017, p. 233.
[105] Lavrijssen & Weitering, *VR* 2019, p. 169 (met verwijzingen).
[106] Zie in dit verband Vellinga, *VR* 2014, p. 373.
[107] Al dan niet in samenhang met art. 6 WVW. Zie Tjong Tjin Tai & Boesten, *NJB* 2016, p. 657.
[108] Van Wees, *AV&S* 2015, p. 174 (met verwijzing).
[109] Tjong Tjin Tai & Boesten, *NJB* 2016, p. 657.

toegerekend aan de menselijk bestuurder wegens de in het verkeer geldende opvattingen?[110] Zoals gezegd worden hoge eisen gesteld aan de kennis en kunde van de bestuurder. Een perfecte chauffeur vormt veelal de norm, met name wanneer het slachtoffer zelf geen verwijt kan worden gemaakt; een beroep op het ontbreken van toerekenbaarheid wordt aldus niet snel gehonoreerd.[111] Indien de auto de rijtaak (deels) zelfstandig uitvoerde, maar de bestuurder nog steeds een toezichthoudende rol had en indien nodig had moeten ingrijpen (te denken valt aan de de lagere automatiseringsniveaus), dan zal de bestuurder die te laat is met ingrijpen zich naar alle waarschijnlijkheid niet kunnen disculperen en zal hij aansprakelijk zijn.[112]

Art. 6:185 e.v. BW (Productaansprakelijkheid)

Hierboven is geconstateerd dat in gevallen waarin de bestuurder een toezichthoudende rol heeft en indien nodig de rijtaak moet overnemen, de chauffeur op grond van art. 6:162 BW waarschijnlijk aansprakelijk zal zijn voor schade van een gemotoriseerd slachtoffer. Wanneer de bestuurder geen toezichthoudende rol heeft, bijvoorbeeld bij automatiseringsniveau 4 of 5, is het minder overtuigend om vast te houden aan aansprakelijkheid van de 'bestuurder'.[113] Een meer voor de hand liggende kandidaat is in dat geval de producent van de (deels) zelfrijdende auto.[114]

Op grond van art. 6:185 BW is de producent aansprakelijk voor schade die wordt veroorzaakt door een gebrek in zijn product. Een product is gebrekkig indien het product (in dit geval de zelfrijdende auto) niet de veiligheid biedt die men daarvan, alle omstandigheden in aanmerking genomen, redelijkerwijs mag verwachten.[115] Een voorvraag die rijst in het licht van zelfrijdende auto's en productaansprakelijkheid, is of software als een product kan worden aangemerkt.[116] Het antwoord op deze vraag is van belang, aangezien het schadeveroorzakende potentieel van (deels) zelfrijdende auto's met name verband houdt met de (in de software opgeslagen) algoritmen.[117] In de literatuur wordt veelal aangenomen dat software die geïncorporeerd is in een stoffelijk product en dienstbaar is aan het functioneren van dat product, onder het productaansprakelijkheidsregime valt.[118] Daarmee zijn echter nog niet alle denkbare vragen over het productbegrip en de reikwijdte van het productaansprakelijkheidsregime ten aanzien van software in zelfrijdende auto's

[110] Art. 6:162 lid 3 BW.
[111] Van Wees, *AV&S* 2015, p. 174.
[112] Van Wees, *AV&S* 2015, p. 174; Tjong Tjin Tai & Boesten, *NJB* 2016, p. 659. Vgl. Tichelaar, *TAV* 2018, p. 31.
[113] Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 35 en 42.
[114] Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 35 en 42.
[115] Art. 6:186 BW. Hierbij dienen de presentatie van het product, het redelijkerwijs te verwachten gebruik van het product en het tijdstip waarop het product in het verkeer werd gebracht in het bijzonder in aanmerking genomen te worden. En ook andere omstandigheden, zoals bijvoorbeeld de beschikbaarheid van alternatieven, de ernst van het gevaar en de waarschijnlijkheid dat het gevaar zich zal voordoen, en een afweging van de voor- en nadelen van een product kunnen een rol spelen.
[116] Zie in dit verband: Schreuder, *AV&S* 2014, p. 132; Van Wees, *AV&S* 2015, p. 172; Vellinga & Vellinga, *VR* 2015, p. 87; Engelhard, *AA* 2017, p. 231; Van Wees, *Maandblad voor Vermogensrecht* 2018, p. 115; Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 28.
[117] Van Wees, *Maandblad voor Vermogensrecht* 2018, p. 115.
[118] Immers, de software vormt in dat geval een onlosmakelijk onderdeel van het product. Zo constateert Van Wees, *Maandblad voor Vermogensrecht* 2018, p. 115.

beantwoord.[119] De software die bepalend is voor het functioneren van een zelfrijdende auto kan immers ook door de producent in het verkeer worden gebracht zonder dat deze is geïncorporeerd in het voertuig.[120] Daarbij kan worden gedacht aan *non-embedded software* of *over-the-air updates*. De New Technologies Formation (EU Expert Group on Liability and New Technologies)[121] heeft deze onduidelijkheid gesignaleerd en is van mening dat ook wanneer (essentiële componenten van) een product een *digital form* aannemen, het productaansprakelijkheidsregime van toepassing zou moeten zijn.[122]

Wanneer is een zelfrijdende auto gebrekkig? Van Wees merkt op dat beperkingen van de techniek kunnen meebrengen dat een zelfrijdende auto – die in het algemeen verkeersongevallen helpt voorkomen – een enkele keer een ongeval veroorzaakt dat een bestuurder van vlees en bloed had kunnen (en moeten) vermijden. Met andere woorden: bij de huidige stand van de techniek zal een zelfrijdende auto niet onder alle omstandigheden aan de menselijke-bestuurdermaatstaf kunnen voldoen. De vraag is dan of dit een gebrek in de zin van art. 6:186 BW oplevert, waarvoor de producent aansprakelijk is.[123] Een in de literatuur veelgemaakt onderscheid is de onderverdeling tussen productiegebreken, informatiegebreken en ontwerpgebreken. Deze kunnen zich in de context van (deels) zelfrijdende auto's alledrie voordoen. Indien schade ontstaat als gevolg van bijvoorbeeld een ondeugdelijke laserscanner of camera, dan zal er over het algemeen sprake zijn van een productiegebrek.[124] Daarnaast is denkbaar dat de producent onjuiste of ontoereikende informatie aan de inzittende geeft,[125] bijvoorbeeld wanneer de bestuurder de rijtaak alleen aan de auto mag overlaten op de snelweg, maar dit niet (duidelijk) wordt medegedeeld. In zo'n situatie is naar alle waarschijnlijkheid sprake van een informatiegebrek.[126] Bij ontwerpgebreken is meer discussie mogelijk.[127] In dat verband kan bijvoorbeeld worden gedacht aan een zelfrijdende auto die op de snelweg onterecht remt voor een opwaaiende krant. Kwalificeert het feit dat de algoritmen de krant niet konden onderscheiden van een daadwerkelijk gevaar (zoals een blok beton) als een ontwerpgebrek? Een complicatie wordt in dit verband gevormd door het feit dat (deels) zelfrijdende auto's steeds meer gebruik zullen maken van *machine learning*-algoritmen, hetgeen hierboven reeds naar voren is gekomen. De producent weet daardoor eigenlijk ook niet

[119] Van Wees, *Maandblad voor Vermogensrecht* 2018, p. 115.
[120] Van Wees, *Maandblad voor Vermogensrecht* 2018, p. 115.
[121] Hieronder zal (nog) een aantal keer worden verwezen naar het recente rapport *Liability for Artificial Intelligence and other emerging digital technologies* van de New Technologies Formation - EU Expert Group on Liability and New Technologies. In dit verband is het relevant om te vermelden dat de verwachting bestaat dat de tweede tak van deze Expert Group (zijnde de Product Liability Directive Formation), op termijn met een publicatie zal komen waarin expliciet(er) aandacht wordt besteed aan de vraag of de Richtlijn productaansprakelijkheid al dan niet moet worden gewijzigd met het oog op de komst van '*artificial intelligence and other emerging digital technologies*'.
[122] Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 6, 28 en 42-43.
[123] Van Wees, *Maandblad voor Vermogensrecht* 2018, p. 118 (met verwijzingen).
[124] Schreuder, *AV&S* 2014, p. 132; De Vey Mestdagh & Lubbers, *AA* 2015, p. 275; Van Wees, *AV&S* 2015, p. 172; Tjong Tjin Tai & Boesten, *NJB* 2016, p. 660.
[125] Het belang van toereikende productbegeleidende informatie is extra groot bij nieuwe producten zoals de zelfrijdende auto. Zie Van Wees, *Maandblad voor Vermogensrecht* 2018, p. 116.
[126] Tjong Tjin Tai & Boesten, *NJB* 2016, p. 660.
[127] Zie bijvoorbeeld Van Wees, *AV&S* 2015, p. 173.

precies wat de auto geleerd heeft en heeft daardoor geen volledig zicht meer op hoe de auto zal reageren in een concrete situatie.[128] Of dit de kwalificatie als een (ontwerp)gebrek in de weg staat, is de vraag.[129] Tjong Tjin Tai en Boesten wijzen in dit verband op het feit dat de keuze voor dergelijke algoritmen ook een ontwerpkeuze is. Als er te beperkte leerscenario's of testscenario's zijn gebruikt is dat volgens hen eveneens een ontwerpgebrek.[130]

Het is vervolgens aan het slachtoffer om het gebrek aan te tonen, alsmede om aan te tonen dat dit de oorzaak is van zijn schade.[131] De complexiteit en ondoorzichtigheid van (de werking van) de algoritmen kunnen er echter voor zorgen dat het onevenredig ingewikkeld en kostbaar is voor het slachtoffer om het gebrek en/of causaliteit te bewijzen.[132] De New Technologies Formation stelt in dat verband voor om, wanneer is bewezen dat de technologie schade heeft veroorzaakt, de bewijslast voor wat betreft het gebrek om te draaien wanneer sprake is van dergelijke onevenredige '*difficulties or costs*'.[133] Voor wat betreft causaliteit, merkt zij op dat een verlichting van de bewijslast voor het slachtoffer denkbaar is.[134] Tevens voorstelbaar is het (verplicht) gebruik van een zogenaamde *Event Data Recorder*.[135] Doordat de *Event Data Recorder* relevante gegevens over het functioneren van de zelfrijdende auto registreert, kunnen bewijsproblemen mogelijk worden voorkomen.[136]

Aandachtspunt voor wat betreft de positie van het slachtoffer is voorts de uitleg van het ontwikkelingsrisicoverweer, op grond waarvan een producent zich van aansprakelijkheid kan bevrijden.[137] Voor een geslaagd beroep op deze bepaling moet de producent aantonen dat het gezien de stand van de wetenschappelijke en technische kennis op het tijdstip dat hij het product in het verkeer bracht, onmogelijk was het bestaan van het gebrek te ontdekken.[138] Een voor de producent gunstige uitleg van ontwikkelingsrisicoverweer zou erop neerkomen dat de woorden 'wetenschappelijke en technische kennis' zo geïnterpreteerd dienen te worden dat dit verweer niet slechts ziet op de vraag of het gebrek (in abstracto) bekend was, maar ook op de vraag of het op basis van de stand van de techniek mogelijk was het gebrek in het concrete geval te ontdekken.[139]

---

[128] Zie Tjong Tjin Tai & Boesten, *NJB* 2016, p. 660; Van Wees, *Maandblad voor Vermogensrecht* 2018, p. 115.
[129] Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 28: *'Sophisticated AI autonomous systems with self-learning capabilities also raise the question of whether unpredictable deviations in the decision-making path can be treated as defects. Even if they constitute a defect, the state-of-the-art defence may apply.'*
[130] Zie Tjong Tjin Tai & Boesten, *NJB* 2016, p. 660; Van Wees, *Maandblad voor Vermogensrecht* 2018, p. 115.
[131] Art. 6:188 BW. Zie De Bruin, *European Journal of Risk Regulation* 2016, p. 491.
[132] De Bruin, *European Journal of Risk Regulation* 2016, p. 491 (met verwijzing); Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 28.
[133] Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 6 en 42.
[134] Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 3, 44 en 49-52.
[135] Van Wees, *AV&S* 2015, p. 176.
[136] Van Wees, *AV&S* 2015, p. 176; De Bruin, *European Journal of Risk Regulation* 2016, p. 495. Ook de New Technologies Formation wijst op de mogelijkheid om producenten te verplichten om technologie uit te rusten met zogenaamde *logging-by-design* functies. Zie Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 47-48.
[137] Art. 6:185 lid 1 onder e BW. Zie in dit verband: De Bruin, *European Journal of Risk Regulation* 2016, p. 490.
[138] Van Wees, *Maandblad voor Vermogensrecht* 2018, p. 119.
[139] Van Wees, *Maandblad voor Vermogensrecht* 2018, p. 119.

Een beperktere, slachtoffervriendelijke uitleg komt erop neer dat onvoorspelbare en praktisch onvermijdelijke 'softwarefouten' in generieke zin nu eenmaal bekend zijn en er daarom geen beroep op het ontwikkelingsrisicoverweer mogelijk is.[140] De New Technologies Formation erkent dat, met de komst van geavanceerde en op ingewikkelde *machine learning*-algoritmen gebaseerde producten, het ontwikkelingsrisicoverweer mogelijk een grotere rol gaat spelen.[141] Dit zou in haar ogen onwenselijk zijn.**[142]** Vellinga wijst in dit verband op de mogelijkheid voor EU-lidstaten om het ontwikkelingsrisicoverweer met betrekking tot zelfrijdende auto's uit te sluiten.[143]

Ten slotte is nog van belang dat fabrikanten zich van aansprakelijkheid kunnen bevrijden door te stellen dat de schade is ontstaan door risico's die pas zijn ontstaan na productie.[144] Immers, het kantelpunt voor aansprakelijkheid is het moment dat de producent het product in het verkeer heeft gebracht.[145] Zelfrijdende auto's zijn echter per definitie producten die niet 'af' zijn wanneer deze op de markt worden gebracht.[146] Producenten zullen in meer of mindere mate grip houden op door hen op de markt gebrachte auto's in de vorm van, bijvoorbeeld, software-updates.[147] Zij moeten zich in zulke gevallen niet kunnen bevrijden van aansprakelijkheid wanneer het gebrek het gevolg is van veranderingen aan het product die de producent heeft doorgevoerd nadat deze op de markt is gebracht (bijvoorbeeld in de vorm van een software-update).[148]

Art. 6:173 BW (Aansprakelijkheid voor gebrekkige roerende zaken)

Op grond van art. 6:173 BW is de bezitter van een roerende zaak waarvan bekend is dat zij, zo zij niet voldoet aan de eisen die men in de gegeven omstandigheden aan de zaak mag stellen, een bijzonder gevaar voor personen of zaken oplevert en dit gevaar zich verwezenlijkt, aansprakelijk. Enig schuldfeit in de vorm van onvoldoende onderhoud of onzorgvuldig gebruik is niet vereist en de bezitter zal zich niet kunnen verweren met het argument dat hij het gebrek niet kende.[149] Het gebreksbegrip vertoont grote gelijkenis met het gebreksbegrip in het kader van de productaansprakelijkheid.[150]

Voor de bezitter ligt een (vermoedelijk) belangrijke disculpatiemogelijkheid besloten in het tweede lid van art. 6:173 BW. Daarin is bepaald dat de bezitter niet aansprakelijk is voor de door de gebrekkige zaak veroorzaakte schade wanneer een gebrekkige zaak tevens als een gebrekkig

[140] Van Wees, *Maandblad voor Vermogensrecht* 2018, p. 119.
[141] Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 29.
[142] Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 6 en 42-43.
[143] Artt. 7 onder e jo. 5 lid 1 onder b Productaansprakelijkheidsrichtlijn. Zie Vellinga, *VR* 2020, p. 70.
[144] Art. 6:185 lid 1 onder b BW. Zie in dit verband: Engelhard, *AA* 2017, p. 232.
[145] Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 28.
[146] Vgl. Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 33.
[147] Van Wees, *Maandblad voor Vermogensrecht* 2018, p. 115.
[148] Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 3, 6, 42-43. Vgl. Vellinga, *VR* 2020, p. 70.
[149] Van Wees, *AV&S* 2015, p. 174.
[150] Van Wees, *AV&S* 2015, p. 174 (met verwijzing).

product in de zin van art. 6:185 e.v. BW kwalificeert. Het gaat hier dus om een exclusieve kanalisatie naar de producent; indien het gebrek al bestond toen deze door de producent in het verkeer werd gebracht, rust de aansprakelijkheid op de producent (en niet tevens op de bezitter).[151] Aangezien een gebrekkige zaak in de zin van art. 6:173 BW ook steeds een gebrekkig product in de zin van art. 6:186 BW zal opleveren, is het goed denkbaar dat aansprakelijkheid van de producent in het licht van de (deels) zelfrijdende auto de hoofdregel zal worden.[152]

5.5.3 Tussenconclusie

Het regime van verkeersaansprakelijkheid voor ongemotoriseerde slachtoffers (art. 185 WVW) laat zich in haar huidige vorm ook toepassen op de situatie waarin zich in een ongeval met een (deels) zelfrijdende auto voordoet. Als gevolg hiervan zal het ongemotoriseerde slachtoffer zijn schade in beginsel kunnen verhalen op de eigenaar of houder van desbetreffende auto (ongeacht het automatiseringsniveau van de auto).[153] Een gemotoriseerd slachtoffer, dat een beroep moet doen op art. 6:162 BW, zal daar in de regel in slagen wanneer sprake is van een auto tot en met automatiseringsniveau 3. Indien sprake is van een auto met automatiseringsniveau 4 of 5, is het minder voor de hand liggend om vast te houden aan aansprakelijkheid voor de 'bestuurder'. Hierbij kan overigens hetzelfde worden opgemerkt in het kader van de aansprakelijkheid van de eigenaar of houder van een auto met automatiseringsniveau 4 of 5 bij *on*gemotoriseerde slachtoffers, ongeacht de 'toepasbaarheid' van art. 185 WVW. Een overtuigender kandidaat is in beide gevallen de producent van de zelfrijdende auto. In dit verband kan worden gesteld dat het regime zich niet steeds eenvoudig laat toepassen op (deels) zelfrijdende auto's, maar dat het gros van de potentiële problemen middels interpretatie van het bestaande juridisch kader kan worden opgelost. De enige mogelijke uitzondering hierop ziet op de bewijsproblematiek voor het slachtoffer (ten aanzien van zowel het gebrek als de causaliteit), in het licht waarvan een omkering en/of verlichting van de bewijslast valt te overwegen.[154] Ten slotte is art. 6:173 BW besproken, op grond waarvan de bezitter van een gebrekkige roerende zaak door het slachtoffer aansprakelijk kan worden gesteld. Op grond van de exclusieve kanalisatiebepaling van art. 6:173 lid 2 BW is echter goed denkbaar dat aansprakelijkheid van de producent in het licht van de zelfrijdende auto de hoofdregel zal worden.

5.6 Duurzaamheid

In aanvulling op de drie overkoepelende waarden en belangen, vormt duurzaamheid een eerste specifieke publieke waarde voor deze casestudy. Drie centrale elementen binnen het begrip

[151] Dit geldt ook als er sprake is van een gebrekkige zaak in de zin van art. 6:173 BW. Zie in dit verband: Van Wees, *AV&S* 2015, p. 174; Engelhard, *AA* 2017, p. 233.
[152] Schreuder, *AV&S* 2014, p. 135; Van Wees, *AV&S* 2015, p. 174 (beide met verwijzingen).
[153] Op grond van art. 185 WVW.
[154] Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 3, 6, 42, 44 en 49-52.

'duurzaamheid' zijn het voorkomen van milieudegradatie, het faciliteren van duurzame economische groei en het streven naar sociale rechtvaardigheid.[155]

5.6.1 Duurzaamheid en (deels) zelfrijdende auto's

Innovatie op het gebied van (deels) zelfrijdende auto's wordt in de literatuur vaak in verband gebracht met duurzaamheid. Rijstijlverbetering kan immers leiden tot optimalisatie van brandstofgebruik.[156] De vraag of de introductie van steeds autonomer vervoer daadwerkelijk leidt tot duurzamer vervoer, is echter afhankelijk van allerlei factoren, zoals de al dan niet gecombineerde overgang naar elektrisch rijden. Dit zal onder 5.6.2 nader worden toegelicht. In deze subparagraaf staat het beleids- en regelgevingskader dat van belang is voor duurzaamheid centraal. Dit beleids- en regelgevingskader is volop in ontwikkeling op internationaal, Europees en nationaal niveau.

Op internationaal niveau valt allereerst te wijzen op de *Sustainable Development Goals* of Duurzame Ontwikkelingsdoelen die in 2015 zijn aangenomen door alle lidstaten van de Verenigde Naties als onderdeel van de *2030 Agenda for Sustainable Development*.[157] Volgens Lim en Taeihagh kan de wijdverspreide adoptie van autonome voertuigen bijdragen aan ontwikkelingsdoel 11: *Sustainable cities and communities*.[158] Hoewel de *Sustainable Development Goals* geen wetgeving als zodanig vormen, hebben zij wel effect op nationale kaders. Zo blijkt uit het regeerakkoord een focus op het stimuleren van CO2-neutrale oplossingen zoals zelfrijdende auto's oftewel 'intelligente transportsystemen'.[159]

Ook het Europees beleids- en regelgevingskader is in ontwikkeling. De Europese Commissie erkent dat de automatisering van vervoer kan bijdragen aan duurzamer vervoer.[160] Om deze ontwikkeling te stimuleren, is volgens de Europese Commissie een combinatie van *supportive measures* en *key policy and regulatory initiatives* van belang.[161] Verder wordt in dit verband in het rapport *GEAR 2030* de aanbeveling gedaan om door samenwerking tussen publieke en private actoren te komen tot verdere *regulation, standards and their enforcement*.[162]

Op nationaal niveau valt te wijzen op het *Wetgevingsoverzicht Klimaatakkoord* dat is opgesteld in het kader van het Klimaatakkoord van 3 juli 2019.[163] In het Wetgevingsoverzicht ligt voor wat betreft

[155] Vgl. onder andere Noy & Givoni, *Sustainability* 2018.
[156] Gawron e.a., *Environmental Science & Technology* 2018 p. 3249-3256.
[157] Zie sustainabledevelopment.un.org/sdgs.
[158] Lim & Taeihagh, *Energies* 2018, p. 1062. Daarnaast is ontwikkelingsdoel 9 relevant: *Industry, innovation and infrastructure*. Voor wat betreft de overgang naar elektrisch rijden is ten slotte ook ontwikkelingsdoel 7 van belang: *Affordable and clean energy*.
[159] Regeerakkoord 2017-2021, 'Vertrouwen in de toekomst', p. 39. Vgl. ook *Kamerstukken II* 2016/17, 26485, nr. 232.
[160] Europese Commissie 2018.
[161] Europese Commissie 2018.
[162] High Level Group on the Competitiveness and Sustainable Growth of the Automotive Industry in the European Union 2017, p. 10 en 50.
[163] *Kamerstukken II* 2019/20, 32813, nr. 394.

het stimuleren van elektrische auto's en elektrisch vervoer de nadruk op fiscale maatregelen en subsidies. Bovendien verdient de (hieronder in het kader van verkeersveiligheid nader te bespreken) Experimenteerwet voor zelfrijdende auto's vermelding. Een van de centrale gedachten achter deze wet is de bijdrage die zelfrijdend vervoer kan leveren aan duurzaamheids-doelstellingen,[164] welke onder andere afhangt van de mate waarin zelfrijdende auto's uiteindelijk op de weg kunnen komen.

5.6.2 Kansen, risico's en bestendigheid juridisch kader

De betekenis van duurzaamheid voor de verkeer- en vervoerssector ligt in de transitie naar, onder andere, geautomatiseerd rijden.[165] Communicatie tussen auto's onderling en met de infrastructuur (door middel van de besproken V2X-communicatietechnologie) zou kunnen zorgen voor beslissingen die leiden tot, onder meer, efficiënter rijgedrag en verbeterde doorstroom door onderlinge afstemming van afstand en snelheid. Dit kan op haar beurt resulteren in geoptimaliseerd weggebruik en verminderde uitstoot en energieverbruik.[166]

Op zowel Europees als nationaal niveau onderkennen beleidsmakers dat het vergroten van de efficiëntie en duurzaamheid van het *huidige* transportsysteem onvoldoende is om doelen op het gebied van CO2-reductie en luchtkwaliteit te halen.[167] Technologische innovaties moeten daarom gefaciliteerd worden, waarbij onder andere wordt nagedacht over 'toekomstscenario's voor slimme en duurzame mobiliteit'.[168] Suggesties uit de literatuur in dat opzicht zijn autonoom en elektrisch rijden, evenals gedeeld gebruik van voertuigen (als dienst in plaats van als eigendom), gedeelde ritten en *mobility as a service*.[169] Dit laatste concept trekt de 'deelcultuur' een stap verder en biedt mobiliteit op maat die de noodzaak voor het bezit van particuliere wijzen van vervoer zou kunnen vervangen.[170] Het Ministerie van Infrastructuur en Waterstaat wil de werking en effecten van *mobility as a service* testen met pilots.[171]

Van belang is wel dat de technische mogelijkheden van verschillende vormen van mobiliteit worden afgewogen tegen de maatschappelijke wenselijkheid, tegen de achtergrond van uitdagingen van veiligheid (hetgeen centraal staat in de volgende paragraaf), duurzaamheid en bereikbaarheid.[172] De zojuist geschetste kansen voor duurzaamheid kunnen namelijk omslaan in

[164] *Kamerstukken II* 2017/18, 34838, nr. 3, p. 1-2 en 19.
[165] Vgl. *Kamerstukken II* 2015/16, 21501-33, nr. 592, p. 1.
[166] Zie in deze zin onder andere: Lim & Taeihagh, *Energies* 2018, p. 1062; *Kamerstukken II* 2013/14, 31305, nr. 210, p. 1.
[167] *Kamerstukken II* 2015/16, 21501-33, nr. 592, p. 2.
[168] *Kamerstukken II* 2015/16, 21501-33, nr. 592, p. 2.
[169] Zie onder andere Nikitas e.a., *Urban Science* 2017; Lastdrager, tno.nl 27 februari 2019.
[170] Nikitas e.a., *Urban Science* 2017.
[171] *Kamerstukken II* 2018/19, 31305, nr. 294.
[172] Hiernaar wordt onderzoek gedaan door, onder andere, TNO (Wilmink, tno.nl 17 juli 2018). Vgl. ook Van de Weijer, *De Volkskrant* 27 december 2019.

risico's als ons gedrag verandert door de ontwikkelende mogelijkheden in mobiliteit.[173] Dit is bijvoorbeeld het geval wanneer zelfrijdende auto's alom tegenwoordig en toegankelijk zouden worden en mensen daardoor steeds minder vaak voor het OV zouden kiezen. Als de wegen vollopen met zelfrijdende auto's, worden de gunstige effecten voor doorstroming en brandstofverbruik immers mogelijk teniet gedaan (tenzij de auto's niet alleen autonoom worden, maar ook elektrisch en gedeeld).[174] De kansen worden echter ook niet gerealiseerd als de introductie van zelfrijdende auto's in het verkeersbeeld onvoldoende van de grond komt. Er zal dus een balans moeten worden gevonden.[175]

Al deze ontwikkelingen vragen om een passend juridisch kader. Zoals duidelijk is geworden, is het beleids- en regelgevingskader voor duurzaamheid in ontwikkeling en is er recent nieuwe regelgeving met relevantie voor zelfrijdende auto's geïntroduceerd. De vraag naar de bestendigheid van het huidige juridisch kader om kansen op dit gebied te verwezenlijken en risico's te mitigeren, moet dan ook tegen de achtergrond van die ontwikkelingen worden bezien. Om de hierboven besproken balans te vinden in de context van regulering, is empirische data over de technologie van zelfrijdende auto's (met name wat betreft automatiseringsniveau 3 en hoger) nodig. Omdat deze technologie nog in de kinderschoenen staat, is toereikende empirische data over het gebruik van deze voertuigen nog niet ruimschoots voorhanden.[176] Het feit dat er in Nederland wetgeving wordt aangenomen die experimenten met zelfrijdende auto's mogelijk maakt, is in dat opzicht een positieve ontwikkeling.

5.6.3 Tussenconclusie

De introductie van zelfrijdende auto's kan bijdragen aan duurzaamheidsdoelstellingen, mede door efficiënter rijgedrag dankzij de inzet van V2X-communicatietechnologie. Deze kansen voor duurzaamheid kunnen echter omslaan in risico's, bijvoorbeeld als men door de komst van de zelfrijdende auto minder vaak voor het OV kiest en de wegen vollopen. De kansen worden daarentegen ook niet gerealiseerd als er onvoldoende ruimte wordt gegeven aan de introductie van zelfrijdende auto's op de weg. De ontwikkelingen vragen dan ook om een passend juridisch kader. Hiervoor is empirische data over de technologie van zelfrijdende auto's nodig. Met het verzamelen van deze data kan als gevolg van de Experimenteerwet een start worden gemaakt, hetgeen positief te noemen is.

---

[173] Hetgeen afhangt van de concrete kenmerken van de zelfrijdende auto's en de mate waarin deze op de weg verschijnen. Vgl. ook Van de Weijer & Van der Bent, *De Volkskrant* 26 juni 2015.
[174] Vgl. Milakis, Van Arem & Van Wee, *Journal of Intelligent Transportation Systems* 2017, p. 324; Gryffroy, Fiten & Surinx, *Nieuw Juridisch Weekblad* 2019, p. 544.
[175] Vgl. Lim & Taeihagh, *Energies* 2018, p. 1062. De minister en staatssecretaris van Infrastructuur en Milieu hebben in elk geval aangegeven dat het toekomstige transportsysteem multimodaal moet zijn; naast (zelfrijdende) auto's moeten ook de fiets en het OV er een grote rol in spelen. Zie *Kamerstukken II* 2015/16, 21501-33, nr. 592, p. 3.
[176] Vgl. Milakis, Van Arem & Van Wee, *Journal of Intelligent Transportation Systems* 2017, p. 342.

5.7 Verkeersveiligheid

Verkeersveiligheid is een tweede casusspecifieke waarde, die nauw samenhangt met grondrechtelijke waarden als het recht op leven.[177] Volgens schattingen wordt zo'n 90% van verkeersongevallen – met als gevolg jaarlijks 1,3 miljoen doden en 50 miljoen ernstig gewonden wereldwijd – veroorzaakt door menselijke fouten. Deze fouten kunnen mogelijk voorkomen worden door de introductie van zelfrijdende auto's.[178] Bij de bespreking van het begrip verkeersveiligheid in de context van zelfrijdende auto's zijn drie terugkerende factoren het voertuig, de mens en de infrastructuur.

5.7.1 Verkeersveiligheid en (deels) zelfrijdende auto's

Verkeersveiligheid wordt vaak genoemd als een verwacht positief effect van zelfrijdende auto's omdat zij de *human error* minimaliseren.[179] In deze lijn heeft de EU een *Vision Zero*-doelstelling geformuleerd: geen verkeersdoden op de Europese wegen in 2050.[180] Uit de Kamerbrief *Grootschalige testen van zelfrijdende auto's* (2014) blijkt in dit verband het streven naar (inter)nationale regelgeving die de introductie van 'zelfrijdende voertuigtechnologie' mogelijk maakt.[181] In dezelfde Kamerbrief werd geconcludeerd dat het toenmalig regelgevend kader te beperkt was voor situaties waarin de auto de rijtaak overneemt.[182] De ontwikkeling van het juridisch kader wordt hieronder verder geanalyseerd in het licht van kansen en risico's.

5.7.2 Kansen, risico's en bestendigheid juridisch kader

Zoals hierboven reeds naar voren is gekomen, zijn drie terugkerende factoren in het licht van zelfrijdende auto's en verkeersveiligheid de aspecten voertuig, mens en infrastructuur. Wat de eerste factor betreft, is de mate waarin experimenten met zelfrijdende auto's zijn toegestaan een belangrijke voorvraag. Experimenten zijn immers belangrijk om de functionele veiligheid van de voertuigen, die sterk afhangt van elektronische systemen zoals de sensoren en de communicatietechnologie, te testen.[183] In dit verband gaf toenmalig Minister van Infrastructuur en Milieu in 2014 al aan tests op de openbare weg mogelijk te willen maken. Veiligheid werd de belangrijkste randvoorwaarde genoemd.[184] Hieruit is de Experimenteerwet voor zelfrijdende auto's voortgekomen, die beoogt zulke tests op een veilige wijze te faciliteren.[185]

[177] Het recht op leven is onder andere vastgelegd in art. 2 EVRM.
[178] Zie onder andere: Walker Smith, cyberlaw.stanford.edu/blog 18 november 2013; Europese Commissie 2016; De Bruin, *European Journal of Risk Regulation* 2016, p. 487.
[179] In dit verband moet onderkend worden dat eerst de 'kinderziekten' opgelost moeten worden en dat ongelukken naar verwachting ook in de toekomst zullen blijven gebeuren.
[180] Europese Commissie 2011.
[181] *Kamerstukken II* 2013/14, 31305, nr. 210, p. 3.
[182] *Kamerstukken II* 2013/14, 31305, nr. 210, p. 2.
[183] Vgl. *Kamerstukken II* 2018/19, 31305, nr. 264.
[184] *Kamerstukken II* 2013/14, 31305, nr. 210.
[185] *Kamerstukken II* 2017/18, 34838, nr. 3, p. 1.

Als vergaand geautomatiseerde auto's uiteindelijk op de weg komen, biedt dit mogelijk kansen voor het vergroten van verkeersveiligheid. Zo kan worden gewezen op het feit dat zelfrijdende auto's in beginsel alerter zijn dan mensen en bovendien in staat zijn dingen waar te nemen die een mens (nog) niet kan zien. Wanneer de auto taken overneemt die eerder uitsluitend voorwerp van menselijk handelen waren, kan dit de verkeersveiligheid dus vergroten.[186] Bovendien kan het gebruik van V2X-communicatietechnologie bijdragen aan het creëren van veilige rijomstandigheden.[187] Een risico dat samenhangt met deze communicatietechnologie is de (bewuste) verspreiding van verkeerde informatie. Een gerelateerd risico is de onder 5.3.2 al kort aangestipte vatbaarheid voor cyberaanvallen, waardoor verkeersveiligheid in het geding komt.[188]

Binnen de factor mens kan onderscheid worden gemaakt tussen de bestuurder van de (deels) zelfrijdende auto en andere menselijke verkeersdeelnemers. Wat de bestuurder betreft, is zoals gezegd een verschuiving van controle zichtbaar, namelijk van de bestuurder naar het voertuig. De mate waarin dat gebeurt, is afhankelijk van het automatiseringsniveau van het voertuig. Aan de ene kant impliceert die verschuiving een kans: uit de literatuur volgt dat geavanceerde rijhulpsystemen ongelukken kunnen verminderen, en dat voertuigen met automatiseringsniveau 3 en hoger de verkeersveiligheid verder kunnen vergroten.[189] Aan de andere kant bestaat het risico dat bestuurders te veel gaan vertrouwen op het systeem en daardoor riskanter gedrag vertonen of te laat ingrijpen.[190] De vraag is dan steeds wat verwacht kan worden van bestuurders in een concreet geval, bijvoorbeeld als het gaat om aandachtsspanne, perceptie en besluitvorming.[191] Voor wat betreft andere menselijke verkeersdeelnemers bestaat het risico dat zij bepaalde verwachtingen van zelfrijdende auto's hebben en hun gedrag daarop aanpassen, met mogelijk misbruik van de technologie en gevaarlijke verkeerssituaties tot gevolg.[192] Tegelijkertijd is het voor andere menselijke verkeersdeelnemers niet altijd makkelijk om het gedrag van zelfrijdende auto's te voorspellen, hetgeen op haar beurt risico's voor verkeersveiligheid kan opleveren.[193] In dat opzicht is een voordeel van het toestaan van experimenten dat overige verkeersdeelnemers daarmee kunnen wennen aan zelfrijdende voertuigen in het verkeer.[194] In de literatuur wordt bovendien gesuggereerd dat een mogelijke oplossing is gelegen in het maken van *deliberate technological design decisions* die het gedrag van de zelfrijdende auto voorspelbaarder maken

[186] In dit verband kan worden gewezen op het feit dat de auto door de komst van rijhulpsystemen, zoals onder 5.2.2 reeds naar voren is gekomen, al steeds meer taken van de menselijk bestuurder overneemt. Deze rijhulpsystemen wil de Europese Commissie onder omstandigheden verplicht stellen. Zie daarover het persbericht van 16 april 2019 op: europarl.europa.eu/news/nl/press-room/20190410IPR37528.
[187] Hierbij kan worden gedacht aan een veilige snelheid, een veilige afstand en veilige besturing. Zie in dit verband Weiß, *Computer Networks* 2011, p. 3105.
[188] Milakis, Van Arem & Van Wee, *Journal of Intelligent Transportation Systems* 2017, p. 342.
[189] Vgl. Milakis, Van Arem & Van Wee, *Journal of Intelligent Transportation Systems* 2017, p. 337 en 342.
[190] Martens 2014.
[191] Vgl. Calvert e.a., *IEEE Intelligent Transportation Systems Magazine* 2020.
[192] Surden & Williams, *Cardozo Law Review* 2016, p. 171. Vgl. Milakis, Van Arem & Van Wee, *Journal of Intelligent Transportation Systems* 2017, p. 324-348.
[193] Surden & Williams, *Cardozo Law Review* 2016, p. 151-163.
[194] *Kamerstukken II* 2013/14, 31305, nr. 210.

voor andere (menselijke) verkeersdeelnemers. Het voertuig kan bijvoorbeeld zo worden ontworpen dat het op een intuïtievere manier met mensen kan interacteren. Een centraal principe hierbij is verhoogde communicatie, zoals extern communiceren aan andere verkeersdeelnemers dat zij gedetecteerd zijn door de auto en duidelijk maken wat de verdere intenties van de auto zijn.[195] Ten aanzien van de menselijke factor is, zowel voor de bestuurder van de (deels) zelfrijdende auto als voor overige menselijke verkeersdeelnemers, verder winst te behalen op het gebied van maatschappelijke bewustwording omtrent wat (deels) zelfrijdende auto's wel en niet kunnen. Op die manier kan mogelijk worden voorkomen dat zij de capaciteiten van (deels) zelfrijdende auto's overschatten.[196]

Ten slotte is de factor infrastructuur van belang. Volgens de Europese Commissie roept de steeds verdergaande automatisering van voertuigen vragen op, bijvoorbeeld over de mate waarin de infrastructuur hierop is afgestemd en de interactie tussen de infrastructuur en de voertuigen.[197] In de literatuur wordt in dit licht opgemerkt dat autofabrikanten bij het ontwerpen van (deels) zelfrijdende auto's daarom rekening moeten houden met de 'stand' van de infrastructuur. Tegelijkertijd is het in het licht van verkeersveiligheid van belang dat de infrastructuur op termijn wordt aangepast met het oog op deze ontwikkelingen.[198]

De bespreking van verkeersveiligheid als publieke waarde laat zien dat de kansen, net als bij de casusspecifieke waarde duurzaamheid, mogelijk kunnen omslaan in risico's. In hoeverre is het juridisch kader in staat om de risico's te vermijden of te mitigeren en de kansen te verwezenlijken? Waar het huidige kader de menselijke bestuurder als uitgangspunt lijkt te nemen, is een verschuiving zichtbaar richting meer aandacht voor (deels) zelfrijdende auto's. Hiermee wordt, onder andere, gezocht naar een balans tussen kansen en risico's. Twee voorbeelden in dit verband zijn de volgende. Ten eerste is in juli 2019 de eerdergenoemde Experimenteerwet voor zelfrijdende auto's van kracht geworden. Deze wet operationaliseert het voornemen om innovatiebevorderende wetgeving te ontwikkelen en tests met zelfrijdende auto's juridisch te faciliteren.[199] De Wegenverkeerswet 1994 is met de komst van de Experimenteerwet aangepast in die zin dat nu een vergunning kan worden verleend voor 'experimenten met motorrijtuigen waarin zich geen bestuurder bevindt'.[200] Ten tweede zijn de plannen van RDW, Rijkswaterstaat en CBR om naast persoonsgebonden rijbewijzen een rijbewijs voor (de software van) zelfrijdende

[195] Surden & Williams, *Cardozo Law Review* 2016, p. 163-170. Auteurs wijzen (vanaf p. 174) tevens op de rol die in dit verband mogelijk is weggelegd voor het recht.
[196] Surden & Williams, *Cardozo Law Review* 2016, p. 127 en 170-173.
[197] Europese Commissie 2018, p. 2. Vgl. ERTRAC Working Group "Connectivity and Automated Driving" 2019, p. 46.
[198] Vgl. Calvert e.a., *Theoretical Issues in Ergonomics Science* 2019, p. 7-8 (met verwijzing).
[199] *Kamerstukken II* 2013/14, 31305, nr. 210. Naast experimenten met zelfrijdende auto's op de openbare weg wordt in de literatuur gewezen op de validatie van beveiliging en veiligheid van zelfrijdende auto's via 'digital twins': 'virtual counterpart[s] to actual physical devices' die o.a. de toekomstige toestand van de machine kunnen voorspellen en verschillende omstandigheden kunnen simuleren die onuitvoerbaar zouden zijn in de echte wereld. Vgl. Veledar, Damjanovic-Behrendt & Macher 2019, p. 415-426.
[200] *Kamerstukken II* 2017/18, 34838, nr. 3, p. 1-2.

auto's te introduceren noemenswaardig. Bij het verlenen van het rijbewijs wordt gekeken naar de 'huidige staat' van de software. Daarbij staat centraal of de auto veilig rijgedrag vertoont en of dit rijgedrag voorspelbaar is voor andere menselijke verkeersdeelnemers.[201] Na de verlening wordt bovendien regelmatig gecontroleerd hoe (de software van) de zelfrijdende auto zich ontwikkelt (bijvoorbeeld na updates) en of het rijbewijs als gevolg daarvan al dan niet moet worden ingetrokken.[202] Het rijbewijs voor de zelfrijdende auto bevindt zich momenteel in de testfase.[203] Beoogd wordt het examen zodanig te automatiseren dat de software wordt getest door software.[204]

5.7.3 Tussenconclusie

Innovatie op het gebied van zelfrijdende auto's leidt in theorie tot kansen voor verkeersveiligheid doordat menselijke fouten worden geminimaliseerd. Voorwaarde in dit verband is dat de voertuigen op een veilige manier in het verkeer worden geïntroduceerd. Risico's zijn er echter alsnog, bijvoorbeeld als bestuurders en andere verkeersdeelnemers riskanter gedrag gaan vertonen door een te groot vertrouwen in de zelfstandige besluitvorming van het systeem, of als de beslissingen van de zelfrijdende auto onvoldoende voorspelbaar zijn voor andere verkeersdeelnemers. Recentelijk is regelgeving uitgevaardigd voor een veilige introductie van zelfrijdende auto's, waarbij oog is voor het mogelijke omslagpunt tussen kansen en risico's voor de publieke waarde verkeersveiligheid. Los van het juridisch kader, is mogelijk nog winst te behalen op het gebied van de technologie van de zelfrijdende auto alsmede publieke bewustwording omtrent haar capaciteiten en tekortkomingen.

5.8 Conclusie

In deze casestudy stond de ontwikkeling van (de algoritmische besluitvorming in) zelfrijdende auto's centraal. Meer specifiek is ingezoomd op de werking van de beslissingsalgoritmen (nu en in de toekomst); de kansen en risico's voor publieke waarden en belangen; en de bestendigheid van het juridisch kader in dat verband. Enkele conclusies zijn de volgende.

Algoritmische technologie

De mate waarin voertuigen autonoom kunnen rijden, is afhankelijk van de ontwikkeling van twee onderliggende technieken die het beslissingsalgoritme (dat de beslissing neemt ten aanzien van de rijactie) van invoer voorzien. De eerste en belangrijkste technologie is de sensoriek in de auto, welke kan worden onderverdeeld in drie typen, zijnde i) laserscanners of lidars, ii) camera's en iii) gewone radars. De tweede technologie is de V2X-communicatietechnologie welke voorziet in de communicatie met i) gemotoriseerde weggebruikers (V2V), ii) niet-gemotoriseerde weggebruikers

[201] KPMG 2019, p. 14.
[202] Zie 'Het voertuigrijbewijs', cbr.nl/web/show/id=376508/langid=42.
[203] Kok, automobielmanagement.nl 4 september 2019.
[204] 'Rijexamen voor zelfrijdende auto is afgelast', rijschoolpro.nl/autonieuws 8 oktober 2019.

(V2P) en iii) de infrastructuur (V2I). De huidige toepassing van deze technologie is met name gelegen in zogenaamde *advanced driver assistance systems* in de eerste twee niveaus van automatisering, waarbinnen de beslissingsalgoritmen doorgaans regelgebaseerd zijn. Naarmate de ontwikkeling vordert, zullen de regelgebaseerde algoritmen steeds meer plaats maken voor *machine learning* (beslissings)algoritmen. De voorspelbaarheid en uitlegbaarheid van deze algoritmen zal in beginsel relatief laag zijn. Tegelijkertijd zal de zelfstandigheid relatief hoog zijn.

Publieke waarden en belangen

De centrale onderzoeksvraag van deze casestudy was die naar de relatie tussen de algoritmische besluitvorming in (deels) zelfrijdende auto's en de onderzochte publieke waarden en belangen, te weten bescherming van persoonsgegevens, rechtsbescherming, non-discriminatie, duurzaamheid en verkeersveiligheid. Meer specifiek ging het hierbij om de kansen en risico's die de algoritmische besluitvorming voor deze publieke waarden en belangen in het leven roept. Geconcludeerd kan worden dat het *gebruik* van sensortechnologie en communicatietechnologie gevolgen heeft voor de genoemde publieke waarden en belangen. Hetzelfde geldt voor de omstandigheid dat de besluiten van (deels) zelfrijdende auto's steeds zelfstandiger en tegelijkertijd minder voorspelbaar en uitlegbaar zullen worden.

Te beginnen met het gebruik van sensoriek en V2X-communicatietechnologie voor het leveren van *input* aan het beslissingsalgoritme. Bij de publieke waarde bescherming van persoonsgegevens bestaat een verband tussen het in kaart gebrachte risico en het gebruik van V2X-communicatietechnologie, nu dit een van de factoren is die de auto in staat stelt om persoonsgegevens te verwerken. In aanvulling daarop hangt het gesignaleerde risico voor non-discriminatie direct samen met de effecten van het gebruik van sensortechnologie: worden voetgangers in voldoende mate herkend, en geldt dat voor iedereen in dezelfde mate ongeacht (onveranderlijke) persoonskenmerken? Een kans is daarentegen zichtbaar bij duurzaamheid, waar de inzet van V2X-communicatietechnologie mogelijk leidt tot efficiënter weggebruik en doorstroming, afhankelijk van de concrete modaliteiten. Ook voor verkeersveiligheid bestaan in beginsel kansen voor zover zelfrijdende auto's de *human error* minimaliseren.

Ten tweede is het van belang aandacht te besteden aan de mate van zelfstandigheid, voorspelbaarheid en uitlegbaarheid van het beslissingsalgoritme. De afnemende mate van uitlegbaarheid van het beslissingsalgoritme leidt (al dan niet onbewust) tot risico's voor de publieke waarde non-discriminatie. Immers, lage uitlegbaarheid heeft tot gevolg dat het moeilijk is om te ontdekken en/of bewijzen dat er sprake is van ongelijke behandeling door de beslissingen van de zelfrijdende auto. Ook in het licht van rechtsbescherming zal de verminderde uitlegbaarheid, binnen de context van productaansprakelijkheid, duidelijk negatieve gevolgen hebben voor de bewijspositie van het slachtoffer. Ten slotte levert de toenemende zelfstandigheid van het

beslissingsalgoritme, voor wat betreft de publieke waarde verkeersveiligheid, een potentieel risico op als zowel de bestuurder als andere verkeersdeelnemers riskant gedrag gaan vertonen als gevolg van een te groot vertrouwen in het systeem. Een tweede risico in dit verband hangt samen met de vraag of de beslissingen van zelfrijdende auto's voldoende voorspelbaar zijn voor andere verkeersdeelnemers.

Juridisch kader

Tot slot is in deze casestudy onderzoek gedaan naar het juridisch kader en de vraag of dat de geschetste kansen en risico's helpt verwezenlijken, respectievelijk te vermijden of te mitigeren. Daarbij staat allereerst centraal of de huidige wet- en regelgeving voldoende ruimte biedt om de gesignaleerde risico's door middel van interpretatie onder het bereik van het juridisch kader te brengen. In het licht van de bescherming van persoonsgegevens is vooralsnog onduidelijk of dit het geval is, nu dit een brede interpretatie van het begrip 'verwerkingsverantwoordelijke' uit art. 4(7) AVG zou betekenen. Voor non-discriminatie is wel geconcludeerd dat het huidige juridisch kader een aantal relevante gronden bevat waar het optreden van ongelijke behandeling in de context van zelfrijdende auto's onder kan vallen. Dergelijke ongelijke behandeling kan echter moeilijk aan te tonen zijn door het niet-transparante karakter van de algoritmische besluitvorming, waardoor niet aan de juridische toetsing wordt toegekomen. Iets vergelijkbaars doet zich voor in het licht van rechtsbescherming. Allereerst is geconstateerd dat de (deels) zelfrijdende auto door middel interpretatie onder het bereik van het relevante juridisch kader kan worden gebracht. Tevens kan het gros van de 'scherpe randjes' die zich in dit verband voor kunnen doen, middels interpretatie worden gemitigeerd. De enige mogelijke uitzondering hierop houdt verband met de bewijsproblematiek voor het slachtoffer binnen het productaansprakelijkheidsregime (ten aanzien van zowel het gebrek als de causaliteit), in het licht waarvan een omkering en/of verlichting van de bewijslast valt te overwegen.[205]

Met betrekking tot duurzaamheid en verkeersveiligheid, worden op dit moment (tijdelijke) wet- en regelgevingsinitiatieven ontplooid met het oog op de kansen en risico's van zelfrijdende auto's. Mocht op deze gebieden op langere termijn eveneens worden ingezet op wetgeving, dan vraagt de voortdurend ontwikkelende technologie ook hier om een flexibel en technologieneutraal kader met oog voor het omslagpunt tussen kansen en risico's.

[205] Expert Group on Liability and New Technologies - New Technologies Formation 2019, p. 3, 6, 42, 44 en 49-52.

Hoofdstuk 6. Casestudy De rechtspraak

Stijn van Deursen & Stefan Philipsen

6.1 Introductie

Het belang van een goed functionerende rechtspraak kan moeilijk worden overschat. De rechtspraak is een wezensbestanddeel van iedere rechtsstaat. Binnen de rechtsstaat ziet de rechter toe op de grenzen die het recht stelt aan het handelen van burgers en de overheid. Daarmee draagt hij niet alleen bij aan de instandhouding van de rechtsstaat, maar ook aan het vertrouwen van de burger in het recht, rechtvaardigheid, veiligheid en gerechtigheid.[1]

Sinds een aantal jaar wordt bij de uitoefening van overheidsgezag in toenemende mate gebruikgemaakt van algoritmen. Daarbij komt steeds nadrukkelijker de vraag op of ook de rechtspraak kan profiteren van de kansen die het gebruik van algoritmen zou kunnen bieden. De inzet van algoritmen zou dan onder andere moeten leiden tot een betere en efficiëntere rechtspraak.[2] De aandacht voor de beloften en kansen die gepaard kunnen gaan met de inzet van algoritmen, kan mede worden bezien in het licht van de aanhoudende aandacht binnen de rechtspraak voor een hoge werkdruk en ontoereikende budgetten.[3]

Voor de rechtspraak is een samenstel van publieke waarden en belangen relevant.[4] Die publieke waarden – zoals onafhankelijkheid en onpartijdigheid, maar ook in meer algemene zin de bescherming van persoonsgegevens en het gelijkheidsbeginsel – geven richting aan de wijze waarop de rechtspraak functioneert en zou moeten functioneren. Bij een beslissing over de inzet van algoritmen in de rechtspraak zullen deze waarden dan ook een (beslissende) rol moeten spelen. Daarbij geldt dat de inzet van algoritmen zowel positieve als negatieve effecten kan hebben op de mogelijkheid om de betrokken publieke waarden te verwezenlijken. In deze casestudy onderzoeken wij de kansen en risico's die gepaard gaan met de inzet van algoritmen in de rechtspraak voor de verwezenlijking van de publieke waarden die in hoofdstuk 3 van dit onderzoek zijn toegelicht, in het bijzonder de manier waarop die in het domein van de rechtspraak verder gestalte hebben gekregen.

[1] *NVvR-rechterscode* 2011, p. 1.
[2] Prakken, *NJB* 2018, p. 271; Van Ettekoven & Prins 2018; 'In 2030 zullen computers rechtspreken', mr-online.nl 31 oktober 2016.
[3] Zie bijvoorbeeld Van Rhee 2018.
[4] Hierna wordt alleen het begrip 'publieke waarden' gebruikt. Voor zover het mogelijk is een onderscheid aan te brengen wordt daarmee steeds gedoeld op zowel publieke waarden als op publieke belangen.

6.1.1 Verantwoording en aanpak

Hierna identificeren en beschrijven wij eerst verschillende mogelijke vormen van algoritmische besluitvorming in de rechtspraak. Daarbij bespreken wij algoritmen die op dit moment al worden gebruikt in de rechtspraak, maar vooral ook toepassingen waarvan de ontwikkeling en inzet in de komende vijf tot tien jaar voorzienbaar is. Nadat wij deze (mogelijke) toepassingen in kaart hebben gebracht, zetten wij uiteen welke kansen en risico's uit deze toepassingen voortvloeien voor de realisatie van de publieke waarden die in dit onderzoek centraal staan. Deze waarden liggen voor de rechtspraak onder meer vast in constitutionele waarborgen, zoals de grondrechtelijke eis dat rechtspraak onafhankelijk moet zijn. Daarnaast komen zij tot uitdrukking in regels van procesrecht die bepalen onder welke voorwaarden en op welke wijze rechtzoekenden toegang tot de rechter hebben, en in richtlijnen die door de rechterlijke macht zelf zijn opgesteld.

Er is inmiddels eerste (verkennende) literatuur beschikbaar over de inzet van algoritmen in de rechtspraak. Ook is over dit onderwerp vanuit de rechterlijke macht een aantal (beleids)documenten gepubliceerd. Deze literatuur en (beleids)documenten vormden het startpunt van deze casestudy. In aanvulling daarop zijn acht interviews afgenomen.[5] Deze interviews zijn gebruikt om te toetsen of de bevindingen uit het literatuuronderzoek aansloten bij de wijze waarop door betrokkenen bij de rechtspraak in Nederland over de inzet van algoritmen wordt nagedacht. Vaak verdiepten de interviews de informatie die voortkwam uit de literatuurstudie. Zij boden in het bijzonder een goed beeld van de context waarbinnen de inzet van algoritmen in de Nederlandse rechtspraak vorm zou kunnen krijgen.

Om een inschatting te kunnen maken van de wijze waarop de inzet van algoritmen in de rechtspraak zich in de nabije toekomst kan ontwikkelen, zijn in dit onderzoek ook enkele buitenlandse toepassingen van algoritmen in de rechtspraak betrokken, alsmede toepassingen die zich buiten het domein van de overheidsrechtspraak afspelen. Ook in de interviews is aandacht besteed aan dergelijke toepassingen. Daarbij moet in het bijzonder worden gedacht aan het nog altijd toenemende gebruik van buitengerechtelijke geschilbeslechtings-mechanismen.[6] Vooral vormen van Online Dispute Resolution (ODR) bieden inzicht in de wijze waarop geschilbeslechting met behulp van algoritmen in een digitale omgeving vorm zou kunnen krijgen.[7] Omdat de focus in deze casestudy ligt op de inzet van algoritmen in de overheidsrechtspraak, komen verwijzingen naar ODR-mechanismen in het verslag van ons onderzoek uitsluitend expliciet aan bod wanneer die verwijzing rechtstreeks betekenis heeft, of kan hebben voor (mogelijke) ontwikkelingen binnen de overheidsrechtspraak.

[5] Voor een overzicht van geïnterviewde personen, zie Bijlage 4.
[6] Kramer 2016.
[7] Zie bijvoorbeeld het CREA Project (crea-project.eu/), The Resolver (theresolver.com/) en Modria (tylertech.com/products/modria).

Tot besluit moet nog een tweetal algemene opmerkingen worden gemaakt over de reikwijdte van deze casestudy. Ten eerste dient een onderscheid gemaakt te worden tussen discussies over digitalisering van de rechtspraak enerzijds en discussies over de inzet van algoritmen anderzijds. Deze casestudy richt zich uitsluitend op de laatste ontwikkelingen. Voor veel – maar zeker niet alle – van de door ons beschreven algoritmen in de rechtspraak geldt daarbij wel dat zij pas echt goed van de grond kunnen komen als de rechterlijke procedure digitaal gevoerd wordt.[8] Ten aanzien van die digitalisering werden de afgelopen jaren de eerste voorzichtige stappen gezet.[9] Overigens leidt de digitalisering van rechterlijke procedures ook op zichzelf tot kansen en risico's voor de publieke waarden die in dit onderzoek centraal staan. Die kansen en risico's zullen evenwel niet expliciet in dit onderzoek aan bod komen, omdat dit onderzoek zich in het bijzonder richt op de betekenis van algoritmen voor de besluitvorming als zodanig. Ten tweede merken wij, in het verlengde van het voorgaande, expliciet op dat de toepassing van algoritmen in de rechtspraak nog in de kinderschoenen staat. Een groot deel van deze casestudy is dan ook gericht op de toekomstige inzet van algoritmen in de rechtspraak.[10] Op basis van de literatuurstudie en de interviews hebben wij een inschatting gemaakt van algoritmische toepassingen die in de komende vijf tot tien jaar voorstelbaar zijn. Of die toepassingen ook technisch realiseerbaar zullen blijken, vormt in veel gevallen onderwerp van debat. In deze casestudy hebben wij geprobeerd om dit debat en de verschillende nog te nemen horden op hoofdlijnen zo inzichtelijk mogelijk weer te geven. In veel gevallen geldt namelijk voor de in deze casestudy beschreven toepassingen dat nog aanzienlijke ontwikkelingsslagen moeten worden gemaakt alvorens de algoritmen daadwerkelijk inzetbaar zijn.

6.2 De inzet van algoritmen in de rechtspraak

Eind 2018 heeft de minister voor Rechtsbescherming in een brief aan de Eerste Kamer de verschillende mogelijke algoritmische toepassingen in de rechtspraak uiteengezet. Daarbij zijn de mogelijke toepassingen op de hieronder weergegeven wijze gevisualiseerd.[11] Op de horizontale

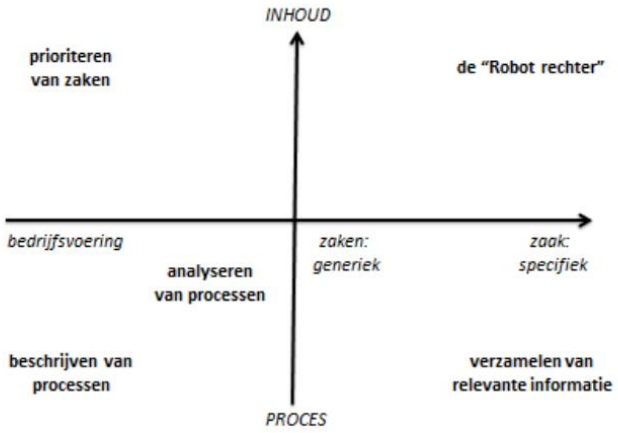
[8] Zie over dit onderwerp: Susskind 2019; Reiling, *Computerrecht* 2020, p. 41 en 44. Voor zover algoritmen gebruikt worden op het niveau van de bedrijfvoering is het niet nodig dat procedures digitaal worden gevoerd. De rechtspraak kan op dat niveau gebruikmaken van digitale data die de rechtspraak ook nu al zelf produceert.

[9] Zie o.a. 'Plannen digitale toegankelijkheid Rechtspraak naar Bureau ICT-toetsing', rechtspraak.nl 15 januari 2020; 'Digitaal procederen nu in alle strafzaken mogelijk', rechtspraak.nl 3 februari 2020. In het buitenland bestaan voorbeelden van (overheids)rechtbanken die wel al gebruikmaken van volledig gedigitaliseerde procedures. Vaak gebruikte voorbeelden in dit kader zijn de internetrechtbanken in Hangzhou, Beijing en Guangzhou. Zie uitvoerig op dit punt Du & Yu, *China Justice Observer* 16 december 2018. Zie daarnaast de Canadese Civil Resolution Tribunal. Andere voorbeelden buiten de sfeer van de overheid kunnen worden gevonden in alternatieve geschilbeslechtingsmechanismen. Een bekend voorbeeld in Nederland is 'e-court'. Zie over dit onderwerp uitvoeriger: Van Duin, *Sdu Blog* 5 februari 2018; zie over het verloop van de procedure bij e-court het procesreglement (*Procesreglement e-court 2017,* 1 februari 2018, e-court.nl/juridisch). Zie verder Spronken, *NJB* 2018, p. 791; Bauw, *AA* 2018, p. 890-893.

[10] Zie voor een overzicht van de technische (on)mogelijkheden: Bex & Prakken, *AA* 2020, p. 255-259; Prakken, *NJB* 2018, p. 269-274; Ashley 2017; Branting, *Artificial Intelligence and Law* 2017, p. 5-27; Surden, *Georgia State University Law Review* 2019, p. 1305-1337.

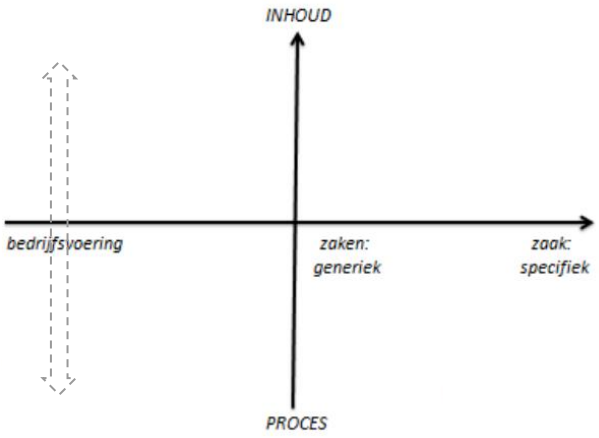
[11] *Kamerstukken I* 2018/19, 34755 VI, nr. AH.

as wordt onderscheid gemaakt tussen algoritmen die gericht zijn op de versterking van de bedrijfsvoering, dat wil zeggen het management en de organisatie van de rechtspraak, en algoritmen die een bijdrage leveren aan de rechterlijke oordeelsvorming in een specifieke procedure. De verticale as differentieert tussen algoritmen die kunnen worden ingezet voor het nemen van procedurele beslissingen en algoritmen die kunnen worden ingezet voor inhoudelijke beslissingen.[12] In het navolgende zullen wij nader ingaan op een aantal van de in onderstaande figuur genoemde toepassingen die op dit moment of in de komende vijf tot tien jaar ontwikkeld en toegepast zouden kunnen worden.[13]



6.2.1 Organisatie, bedrijfsvoering en management

In de linkerhelft van het assenstelsel bevinden zich de algoritmen die ondersteuning bieden bij de bedrijfsvoering. Een aantal mogelijke toepassingen laat zich daarbij voorstellen.



Ten eerste kunnen binnen deze categorie de algoritmen worden geplaatst die gebruikt kunnen worden om zaaksstromen te managen en te monitoren. Hierbij kan worden gedacht aan algoritmen

[12] Prins & Van der Roest, *NJB* 2018, p. 263. Zie voor een vergelijkbaar onderscheid: Philipsen & Themeli, *Rechtstreeks* 2019, p. 46-49.
[13] Deze figuur is afkomstig uit *Kamerstukken I* 2018/19, 34755 VI, nr. AH.

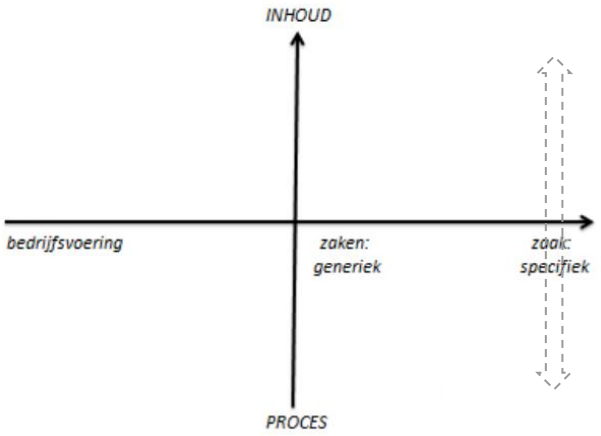
die een rol kunnen spelen in de zaakstoedeling en zaaksverdeling, het roosteren en het plannen van zittingen.[14] Wanneer algoritmen op deze wijze worden ingezet dan kunnen zij helpen om bijvoorbeeld zaken aselect toe te delen, of juist – als daaraan behoefte is – zaken op een evenwichtige wijze aan rechters toe te delen en daarbij rekening te houden met de mogelijkheid zaken over hetzelfde soort geschil samen te voegen. Algoritmen zouden ook kunnen helpen bij het identificeren van procedures waarin eenzelfde persoon een rol speelt en daar handelingen aan kunnen verbinden. Dit laatste lijkt vooral interessant als die procedures op hetzelfde moment aanhangig zijn. Zaken zouden dan kunnen worden samengevoegd, of de rechters in de verschillende procedures zouden op de hoogte gesteld kunnen worden van de uitkomst van die procedures.[15] Uit de interviews is gebleken dat algoritmen nog niet op deze wijze in de rechtspraak in Nederland worden gebruikt, maar volgens sommige geïnterviewden behoort de introductie van dit soort algoritmen in de komende vijf tot tien jaar wel tot de mogelijkheden.

Een tweede mogelijke toepassing van algoritmen binnen de categorie bedrijfsvoering heeft betrekking op een zeer specifiek aspect van die bedrijfsvoering, namelijk het gebruik van algoritmen bij het anonimiseren van uitspraken.[16] Uit de interviews volgt dat de rechtspraak op dit moment tests uitvoert met algoritmen die in staat zijn geautomatiseerd uitspraken te anonimiseren. Het gaat daarbij om *supervised machine learning*-algoritmen die uiteindelijk de menselijke bemoeienis met het anonimiseren moeten verminderen. Deze algoritmen zouden in de toekomst zelfs volledig zelfstandig moeten kunnen functioneren, in die zin dat er geen menselijke interventie meer nodig is in het anonimiseringsproces.[17]

6.2.2 Rechterlijke oordeelsvorming

Een tweede toepassing van algoritmen in de rechtspraak, naast het gebruik van algoritmen in de bedrijfsvoering, is het gebruik van algoritmen bij het nemen van rechterlijke beslissingen.[18] Dit zijn de algoritmen die aan de rechterzijde van het assenstelsel zijn gepositioneerd. Dit type algoritme komt in twee varianten voor: algoritmen die de rechter ondersteunen bij het nemen van een inhoudelijke beslissing, en algoritmen die de rechterlijke beslissing vervangen.

[14] Prins & Van der Roest, *NJB* 2018, p. 263.
[15] Zie over het clusteren van zaken in het bestuursrecht ook Schuurmans 2015, p. 17.
[16] Wij hebben er in deze studie voor gekozen het anonimiseren als onderdeel van de bedrijfsvoering te beschouwen. Aan die keuze ligt een inhoudelijke beoordeling ten grondslag. Een andere visie is het anonimiseren te beschouwen als het sluitstuk van de rechterlijke procedure en de anonimiseringsalgoritmen dan ook aan te merken als algoritmen die rechters ondersteunen. De inhoudelijke beoordeling van de toelaatbaarheid van anonimiseringsalgoritmen vindt evenwel vooral plaats aan de hand van normen die op het niveau van de bedrijfsvoering een nadrukkelijke rol spelen.
[17] Zie in dat kader ook Bakker, *NJBlog* 17 augustus 2017.
[18] Philipsen & Themeli 2020.



Beslisondersteuningsalgoritmen leveren informatie die de rechter gebruikt bij het nemen van zijn beslissing in een concrete zaak. Het kan daarbij gaan om juridische informatie zoals relevante wetgeving en jurisprudentie, maar bijvoorbeeld ook om informatie die betrekking heeft op een voorliggend feitencomplex. Voorstelbaar is ook dat algoritmen in de toekomst gebruikt kunnen worden om rechtstreeks de uitkomst van een concrete procedure vast te stellen. Algoritmen in deze laatste categorie worden ook wel aangeduid als robotrechters.

6.2.2.1 Voorbereiding van een rechterlijke beslissing

Algoritmen beschikken over een aantal specifieke eigenschappen die maken dat zij vooral ter ondersteuning van of in de voorbereiding op de rechterlijke oordeelsvorming goed bruikbaar zijn. De toepassing van algoritmen maakt het ten eerste mogelijk om grote hoeveelheden informatie te doorzoeken en daaruit potentieel relevante gegevens te filteren. Daarnaast kunnen algoritmen tekst analyseren en aan de hand daarvan relevante documenten in bijvoorbeeld een dossier snel identificeren.[19] De advocatuur maakt voor dit soort taken op dit moment al gebruik van algoritmische systemen.[20] In de rechtspraak is dat nog niet het geval.[21] Algoritmen functioneren wanneer zij op deze wijze in de rechtspraak zouden worden gebruikt als een soort deskundige.[22]

Bij de Rechtbank Oost-Brabant wordt op dit moment een pilot uitgevoerd rondom een algoritme dat rechters kan ondersteunen bij de voorbereiding van rechterlijke beslissingen. Daartoe wordt gebruik gemaakt van een zelflerend algoritme dat wordt toegepast op zogenaamde Mulder-zaken. Dit zijn procedures over verkeersovertredingen die vallen binnen de reikwijdte van de Wet administratiefrechtelijke handhaving verkeersvoorschriften (Wahv of Wet Mulder). Het experiment

[19] Reiling, *Computerrecht* 2020, p. 41.
[20] Zie bijvoorbeeld de diensten van LexIQ (lexiq.nl/). Daarover: 'Lexiq lanceert met lexalyse baanbrekende AI-software voor juristen', mr-online.nl 30 juli 2019.
[21] Zie hierover ook: Kauffman & Knowlton 2018. Voor een voorbeeld van een in de private sector ontwikkeld systeem, zie: equivant.com/jworks-for-courts/.
[22] In vergelijkbare zin: Bex & Prakken, *AA* 2020, p. 255-259.

maakt gebruik van een algoritme dat met behulp van *topic modelling* relevante verbanden identificeert. Zo wordt geprobeerd op basis van het digitale dossier van het OM aan de rechter een overzicht van relevante eerdere uitspraken en wet- en regelgeving te presenteren. Met behulp van deze informatie zou de rechter in de toekomst dan weer een uitspraak moeten kunnen opstellen.[23]

6.2.2.2 Een blik op de toekomst: 'De Robotrechter'?

De meest futuristische toepassing van algoritmen in de rechtspraak betreft algoritmen die in staat zijn zelfstandig uitspraak te doen. De vraag of en in hoeverre dit toekomstbeeld zich ook daadwerkelijk zal verwezenlijken is onderwerp van debat. Zo stelt Van den Herik dat er tussen 2030 en 2040 algoritmen ontwikkeld zullen zijn die in staat zijn om een individueel geschil te beslechten.[24] Ook Frits Bakker, voormalig voorzitter van de Raad voor de rechtspraak, stelde in zijn toespraak op de Dag van de Rechtspraak in 2017 dat hij potentie zag in computers die rechtspreken.[25] In de media wordt ook veelvuldig over dergelijke toepassingen gerapporteerd.[26]

Prakken lijkt het daarentegen minder waarschijnlijk dat robotrechters in de nabije toekomst daadwerkelijk het werk van menselijke rechters zullen kunnen overnemen. Volgens hem berusten voornoemde standpunten over de inzet van algoritmen voor individuele geschilbeslechting veelal op een verwarring van de begrippen 'voorspellen' en 'beslissen'.[27] Van voorspellen is sprake als een systeem een inschatting kan maken van de uitkomst van een rechtszaak. Dit type algoritme speelt meestal een centrale rol in berichtgeving in de media over de opkomst van de robotrechter. Een voorbeeld van een voorspelalgoritme is het algoritme dat in staat is om met een nauwkeurigheid van 79% te voorspellen of het EHRM een schending van het EVRM zal vaststellen,[28] of het algoritme dat 70% van de uitspraken van het Amerikaanse Hooggerechtshof kan voorspellen.[29] Prakken merkt daarover op dat dit soort algoritmen veelal slechts een binair (ja/nee) antwoord geven op de vraag of een bepaald grondwets- of verdragsartikel is geschonden.

Bex en Prakken maken op dit punt onderscheid tussen drie typen voorspelalgoritmen.[30] Ten eerste algoritmen die voorspellen op basis van niet-inhoudelijke kenmerken van een procedure. Een voorbeeld daarvan is het algoritme dat de uitspraken van het Amerikaanse Hooggerechtshof voorspelde en daarbij onder ander gebruikmaakte van de datum waarop de zaak diende en de lagere rechtbank die eerder over de zaak had beslist.[31] Ten tweede verwijzen Bex en Prakken naar algoritmen die voorspellen op basis van een tekstuele beschrijving van de procedure. Dit

[23] Zie over dit experiment meer uitvoerig: Van der Put, *Rechtstreeks* 2019, p. 50-60.
[24] 'In 2030 zullen computers rechtspreken', mr-online.nl 31 oktober 2016.
[25] Bakker 2017.
[26] 'In 2030 zullen computers rechtspreken', mr-online.nl 31 oktober 2016.
[27] Zie ook Prakken, *NJB* 2018, p. 269-274.
[28] Zie daarover Aletras e.a., *PeerJ Computer Science* 2016.
[29] Katz, Bommarito & Blackman 2017, *PloS ONE* 2017.
[30] Bex & Prakken, *AA* 2020, p. 255-259.
[31] Prakken, *NJB* 2018, p. 271.

soort algoritmen legt statistische verbanden tussen woorden. Om de uitkomst van een concrete procedure te kunnen voorspellen, hebben deze algoritmen al een groot deel van de tekst van de uiteindelijke uitspraak nodig. Beide algoritmen zijn principieel ongeschikt om inhoudelijke rechterlijke beslissingen te nemen. Een dergelijke beslissing vergt immers een motivering op basis van juridisch relevante kenmerken van een casus.[32] Bij het eerste algoritme spelen die kenmerken geen rol, en ook het tweede algoritme is niet in staat in een dergelijke motivering te voorzien.

Ten derde zijn er algoritmen die wél op basis van juridische relevante factoren kunnen voorspellen. Dan is het evenwel nodig dat eerst voor zowel de relevante jurisprudentie als de procedure waarvoor een voorspelling moet worden gedaan de juridisch relevante factoren worden aangewezen, of gecodeerd zodat zij als zodanig voor het algoritme herkenbaar zijn. Dit vergt dus nog altijd een aanzienlijke menselijke inspanning. Een zelflerend algoritme is vervolgens in staat om verbanden te leggen tussen de juridisch relevante factoren. Bex en Prakken wijzen erop dat het coderen van de juridisch relevante data arbeidsintensief is en dat het onderzoek naar taalanalyse-algoritmen die dit automatisch zouden kunnen doen nog in de kinderschoenen staat.[33] Daarnaast zijn deze voorspelalgoritmen nog niet in staat om voorspellingen te doen die 100% accuraat zijn. Naarmate de procedure complexer wordt, laat dit bezwaar zich sterker voelen. Ook deze nog gebrekkige kwaliteit zet de bruikbaarheid van voorspelalgoritmen onder druk.

Voor zover er in de tijdspanne waarop dit onderzoek betrekking heeft een robotrechter het licht zou zien, is de kans groot dat een dergelijke robotrechter eerst met name zal worden gebruikt om zeer eenvoudige rechtsvragen te beantwoorden. Daarbij kan bijvoorbeeld worden gedacht aan de vraag of griffierecht is betaald, of dat een beroepschrift binnen de daarvoor geldende termijnen is ingediend, of misschien zelfs over inhoudelijke vragen in overzichtelijke en routinematige procedures.[34] Als wij in het vervolg van dit onderzoek over een robotrechter spreken, dan hebben wij telkens het oog op deze eenvoudige, regelgebaseerde toepassing. Overigens moet uit het voorgaande niet worden afgeleid dat de beantwoording van de vraag of een robotrechter in de toekomst zal worden ingezet, volledig wordt bepaald door de technische (on)mogelijkheid een dergelijke rechter te bouwen. De beslissing tot de inzet van robotrechters is vooral ook een ethisch-morele beslissing over de wenselijkheid van geschilbeslechting door algoritmen. Dat onderwerp valt evenwel buiten de reikwijdte van dit onderzoek.

Verder benadrukken wij hier dat in de praktijk lang niet altijd mogelijk is om een scherp onderscheid te maken tussen algoritmen die een beslisondersteuning bieden en algoritmen die functioneren als een robotrechter. Een algoritme dat ontwikkeld is om rechters te ondersteunen bij het nemen

[32] Daarover ook: AARvS 2018, p. 13.
[33] Bex & Prakken, *AA* 2020, p. 255-259.
[34] Reiling, *Computerrecht* 2020, p. 41.

van een beslissing kan immers in de praktijk een grote invloed hebben op de uitkomst van een concrete procedure. Zo is voorstelbaar dat een regelgebaseerd algoritme wordt ingezet om bijvoorbeeld de hoogte van de alimentatie op basis van wettelijke criteria te berekenen. Als dit soort algoritmen in de toekomst gebruikt zouden worden dan is in sommige gevallen de door het algoritme aangereikte informatie, die in eerste instantie vooral bedoeld was als ondersteuning voor de rechters, op in ieder geval een onderdeel van het geschil bepalend voor de uitkomst van dat geschil.[35] Verder zal de invloed van ondersteunende algoritmen op de uitkomst van een concrete procedure groter zijn naarmate rechters de suggesties van die algoritmen makkelijk overnemen.

6.2.3 Ontwikkelingen buiten de rechtspraak en implicaties voor de rechtspraak

Om een zo volledig mogelijk beeld te schetsen van de mogelijke (toekomstige) inzet van algoritmische systemen in de rechtspraak, is het nuttig om ook een blik te werpen op de ontwikkelingen buiten het domein van de (Nederlandse) overheidsrechtspraak. In de afgelopen jaren zijn op het terrein van de *mediation* en ODR verschillende tools ontwikkeld die partijen in de gelegenheid stellen om op efficiënte wijze een oplossing voor hun geschil te vinden.[36] Door sommige geïnterviewden is gesuggereerd dat dergelijke vormen van alternatieve geschilbeslechting in de toekomst ook binnen het stelsel van overheidsrechtspraak ingezet kunnen worden. Volgens deze geïnterviewden moet dit dan leiden tot een efficiëntere geschilbeslechting. Voor het onderhavige onderzoek is vooral van belang te constateren dat als de rechtspraak in de toekomst zou besluiten andere wijzen van geschilbeslechting binnen het publieke domein te faciliteren, algoritmen ook via die weg een intrede kunnen doen in de rechtspraak.

Een van de manieren waarop de rechtspraak daaraan gestalte zou kunnen geven, is door partijen in de fase voorafgaand aan de rechtszaak door een voorportaal, ook wel *solution explorer* genoemd, te leiden. In een dergelijk voorportaal krijgen partijen basale juridische informatie over hun rechten en worden partijen op de mogelijkheid van alternatieve geschilbeslechting gewezen. Verder kan een *solution explorer* partijen stapsgewijs helpen om een geschilbeslechtings-procedure te starten. Een *solution explorer* helpt partijen om een geschil op een voor hen efficiënte wijze te beslechten. Daarmee zou een deel van de conflicten al kunnen worden opgelost voordat de rechter in beeld komt.[37] Het Canadese Civil Resolutions Tribunal (CRT) biedt een goede illustratie van hoe dit zou kunnen werken. Het CRT is een online platform dat partijen, alvorens zij een bindende beslissing krijgen, in de gelegenheid stelt om het geschil op minnelijke wijze te beslechten door gebruik te maken van een door het CRT beschikbaar gesteld programma.[38]

[35] Waarschijnlijk maakte ook e-court gebruik van eenvoudige regelgebaseerde algoritme (Van Duin, *Sdu Blog* 5 februari 2018).
[36] Zie bijvoorbeeld het CREA Project (crea-project.eu/), The Resolver (theresolver.com/) en Modria (tylertech.com/products/modria).
[37] Voor een uitgebreider beschrijving van verschillende modaliteiten, zie Barendrecht e.a. 2016.
[38] Daarover ook Reiling, *Computerrecht* 2020, p. 42.

6.2.4 Tussenconclusie

Hiervoor zijn twee verschillende categorieën van algoritmische toepassingen in de rechtspraak onderscheiden. Ten eerste kunnen algoritmen worden gebruikt voor de bedrijfsvoering van de rechterlijke macht. Ten tweede kunnen algoritmen een rol spelen bij de rechterlijke oordeelsvorming. Binnen die laatste categorie hebben wij vervolgens een nader onderscheid aangebracht tussen algoritmen die dienen ter ondersteuning van de rechterlijke beslissing en algoritmen die functioneren als een zelfstandige geschilbeslechter, ook wel een robotrechter genoemd.

Voor zowel het gebruik van algoritmen in de bedrijfsvoering, als bij de ondersteuning van rechters, geldt dat er op dit moment slechts enkele (in omvang) zeer bescheiden experimenten plaatsvinden. Zo volgt uit de door ons gehouden interviews dat wordt geëxperimenteerd met algoritmen die helpen bij het anonimiseren van uitspraken en met een systeem datde rechter in een specifieke procedure relevante informatie, zoals jurisprudentie en toepasselijke wetgeving, moeten aanreiken. In beide gevallen gaat het om *machine learning*-algoritmen. Over de resultaten van beide experimenten valt op dit moment nog weinig op te merken.

Algoritmen spelen vandaag de dag een zeer beperkte rol in de rechtspraak. Voor de meeste algoritmische toepassingen die hiervoor werden geïdentificeerd geldt dan ook dat onzeker is of zij in de komende vijf tot tien jaar het licht zullen zien. Wel geldt voor alle toepassingen dat uit de literatuur of uit de door ons afgenomen interviews volgt dat de ontwikkeling van die algoritmen voorstelbaar wordt geacht. Daarbij is het aannemelijk dat taken die door een regelgebaseerd (expert)systeem uitgevoerd kunnen worden als eerste in aanmerking komen om met behulp van algoritmen geautomatiseerd te worden. Om dat mogelijk te maken geldt evenwel als voorwaarde dat de rechterlijke procedure verder gedigitaliseerd zal moeten worden.

6.3 Bescherming van persoonsgegevens

6.3.1 Bescherming van persoonsgegevens en algoritmen in de rechtspraak

De verwerking van persoonsgegevens binnen de rechtspraak wordt gereguleerd door de Algemene Verordening Gegevensbescherming (AVG) en enkele specifieke wetten.[39] De AVG geldt in beginsel niet voor de verwerking van persoonsgegevens in het kader van de strafvervolging. De verwerking van dergelijke gegevens wordt bestreken door de Richtlijn

[39] Zie overweging 20 AVG.

gegevensbescherming opsporing en vervolging.[40] In Nederland is die Richtlijn geïmplementeerd in de Wet politiegegevens (Wpg) en de Wet justitiële en strafvorderlijke gegevens (Wjsg).[41]

Uitgangspunt is dat de verwerking van persoonsgegevens dient plaats te vinden in overeenstemming met de beginselen die zijn neergelegd in art. 5 AVG. Voor de verwerking van persoonsgegevens in het kader van een strafvervolging zijn die beginselen terug te vinden in art. 3 Wjsg. Op grond van art. 1 sub k sub 4 Wjsg, zijn de gerechten verantwoordelijk voor de verwerking van gerechtelijke strafgegevens. Voor de verwerking van andere persoonsgegevens in het kader van de uitoefening van gerechtelijke taken zijn dat de volgende partijen: bij de rechtbanken en de gerechtshoven: het gerechtsbestuur; bij de Hoge Raad: de president van de Hoge Raad en bij het parket bij de Hoge Raad: de procureur-generaal bij de Hoge Raad.[42] Krachtens bepaling 1.1 van de Regeling verwerking persoonsgegevens bestuursrechtelijke colleges berust de verantwoordelijkheid bij de voorzitter van de Afdeling bestuursrechtspraak van de Raad van State (ABRvS) en de besturen van de Centrale Raad van Beroep (CRvB) en het College van Beroep voor het bedrijfsleven (CBb).

In het licht van de vereiste rechterlijke onafhankelijkheid valt het toezicht op de verwerking van persoonsgegevens door rechterlijke instanties voor zover het gaat om de uitoefening van rechterlijke taken buiten de bevoegdheden van de Autoriteit Persoonsgevens (AP), die in algemene zin belast is met het toezicht op naleving van de AVG.[43] De verantwoordelijkheid voor het toezicht op de verwerking van persoonsgegevens in het kader van de uitoefening van rechterlijke taken ligt op grond van de AVG bij ‘specifieke instanties binnen de rechterlijke organisatie van de lidstaat, die met name de naleving van de regels van deze verordening moeten garanderen en klachten met betrekking tot die gegevensverwerkingen moeten behandelen’.[44] Toezicht op de verwerking van persoonsgegevens door de gerechten en de Hoge Raad vindt daarbij plaats door respectievelijk de aangewezen Functionarissen Gegevensverwerking en door de Procureur-Generaal bij de Hoge Raad.[45] Bij de bestuursrechtelijke colleges is dit toezicht opgedragen aan de zogenaamde AVG-Commissie.[46]

---

[40] Richtlijn (EU) 2016/680 van het Europees Parlement en de Raad van 27 april 2016 betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens door bevoegde autoriteiten met het oog op de voorkoming, het onderzoek, de opsporing en de vervolging van strafbare feiten of de tenuitvoerlegging van straffen, en betreffende het vrije verkeer van die gegevens en tot intrekking van Kaderbesluit 2008/977/JBZ van de Raad (*PbEU* 2016, L 119). Zie ook art. 3 lid 7 van de Richtlijn en de *Wegwijzer Europese richtlijn gegevensbescherming opsporing en vervolging (EU) 2016/680* van het ministerie van Justitie en Veiligheid, p. 4-5.

[41] Volgens art. 2 lid 2 onder d AVG is de AVG niet van toepassing in de gevallen waarin de Richtlijn van toepassing is.

[42] Bepaling 3 van de Regeling toezicht verwerking persoonsgegevens door gerechten en het parket bij de Hoge Raad.

[43] Art. 55 lid 3 en overweging 20 AVG; overweging 80 Richtlijn (EU) 2016/680. Laatstgenoemde overweging bepaalt verder dat die vrijstelling beperkt dient te blijven tot gerechtelijke activiteiten in het kader van rechtszaken en niet dient te gelden voor andere activiteiten die rechters overeenkomstig het lidstatelijke recht verrichten. Verder bepaalt deze overweging dat het toezicht dient te voldoen aan de voorwaarden zoals gesteld in art. 8 lid 3 Handvest. Zie voor de invulling van het begrip ‘rechterlijke taken’ onder meer AVG Commissie Bestuursrechtelijke Colleges 2019.

[44] Overweging 20 AVG.

[45] Regeling toezicht verwerking persoonsgegevens door gerechten en het parket bij de Hoge Raad.

[46] Regeling verwerking persoonsgegevens bestuursrechtelijke colleges.

Ten aanzien van de omvang van het toezicht laat de zinsnede ‘uitoefening van de rechterlijke taken’ ruimte voor interpretatie. Duidelijk is in ieder geval dat het daarbij dient te gaan om meer dan enkel ‘*genuine judicial activities in court cases*’, zoals werd voorgesteld in het ontwerp voor de AVG, en dat sprake dient te zijn van ‘*courts (...) acting in their judicial capacity*’.[47] Hiervan is sprake bij ‘alle gerechtelijke activiteiten in het kader van rechtszaken’.[48] De AVG-Commissie, die naast zijn toezichthoudende taak de bestuursrechtelijke rechtscolleges adviseert over de toepassing van de AVG,[49] stelt zich op het standpunt dat het daarbij niet enkel gaat om de vraag of de verwerking directe invloed heeft op de oordeelsvorming van de rechter, maar dat daarbij ook belang toekomt aan de aard en het doel van de verwerking.

Tot slot kan niet onvermeld blijven dat een bijzonder deel van de bescherming van persoonsgegevens in de rechtspraak terug te vinden is in de anonimiseringsrichtlijn. Deze richtlijn bepaalt welke gegevens geanonimiseerd moeten worden bij publicatie van een rechterlijke beslissing om op die manier de privacy van de betrokken personen te waarborgen.[50]

6.3.2 Kansen, risico’s en bestendigheid juridisch kader

Bij het gebruik van de in par. 6.2 beschreven algoritmische systemen worden vrijwel altijd persoonsgegevens verwerkt. Dat geldt onder andere voor de verwerking van persoonsgegevens van rechters en andere medewerkers van de rechtspraak wanneer algoritmen worden toegepast op het niveau van de bedrijfsvoering, en voor de verwerking van gegevens van personen die bij specifieke procedures zijn betrokken in het geval waarin algoritmen een rol spelen in het rechterlijke beslissingsproces. Op dit moment gebruikt de rechtspraak in Nederland algoritmen nog niet op de wijze die in dit onderzoek centraal staat. Het zou dan ook speculatief zijn om specifieke algoritmische toepassingen op overeenstemming met bijvoorbeeld het gegevensbeschermingsrecht te beoordelen. Voor die beoordeling is immers vereist dat het specifieke functioneren van een bepaald algoritme duidelijk is. Daarmee is niet gezegd dat zich niet nu al risico’s voordoen voor de publieke waarde van gegevensbescherming. Hoewel de meeste geïnterviewden van mening waren dat het toezicht op de gegevensbescherming binnen de rechtspraak op dit moment naar behoren functioneert, bestaan er ten aanzien van de toekomstige toepassing van algoritmen wel zekere institutionele risico’s. Een beslissing over de inzet van algoritmen in de rechtspraak zal altijd acht moeten slaan op de hierna te bespreken institutionele risico’s.

[47] Voorstel voor een verordening van het Europees Parlement en de Raad betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens en betreffende het vrije verkeer van die gegevens (Algemene Verordening Gegevensbescherming), COM(2012) 11 final, p. 33. Zie ook overweging 20 AVG.
[48] Behandeling van de wijziging van de Wpg en Wjsg: *Kamerstukken II* 2017/18, 34889, nr. 6, p. 19.
[49] AVG Commissie Bestuursrechtelijke Colleges 2019.
[50] Zie daarover rechtspraak.nl/Uitspraken/paginas/anonimiseringsrichtlijnen.aspx#totstandkoming.

Een eerste risico voor de waarde van de bescherming van persoonsgegevens is de onduidelijkheid die er bestaat ten aanzien van het toezicht op de verwerking van persoonsgegevens in de rechtspraak. Het toezicht op de verwerking van persoonsgegevens *in het kader van de rechterlijke taak* valt niet binnen de bevoegdheden van de AP. Daarvoor zijn de rechterlijke instanties zelf verantwoordelijk. Zij hebben zich van die verantwoordelijkheid op de hierboven beschreven wijzen gekweten. Door inherente vaagheid van de term 'rechterlijke taak' is het op voorhand evenwel niet altijd duidelijk wie toezicht moet houden op de verwerking van persoonsgegevens voor doelen van management, bedrijfsvoering en organisatie. Als op het niveau van de bedrijfsvoering tot de inzet van algoritmen wordt besloten zal dit, ook als die algoritmen regelgebaseerd zijn, leiden tot meer, maar vooral efficiëntere verwerking van persoonsgegevens. Het is daarom in het bijzonder relevant om bij een besluit over de inzet van algoritmen op het niveau van de bedrijfsvoering ook duidelijk te maken wie de verantwoordelijkheid draagt voor de verwerking van persoonsgegevens door die algoritmen. De noodzaak daartoe is groter als gebruik wordt gemaakt van zelflerende algoritmen. In de huidige praktijk wordt in twijfelgevallen met betrekking tot de competentie over een bepaald dossier overleg gevoerd tussen de AP en de met het toezicht op de gegevensbescherming belaste organen binnen de rechtspraak. Hoewel dit volgens verschillende geïnterviewden niet tot problemen leidt, kan een helderder verdeling van bevoegdheden bevorderlijk zijn voor de bescherming van persoonsgegevens, zeker als gewerkt wordt met algoritmen.

Ten tweede kan de versplintering van het toezicht op de verwerking van persoonsgegevens ook op zichzelf een risico zijn. Hiervoor werd opgemerkt dat de verwerking van persoonsgegevens die plaatsvindt in het kader van de rechterlijke taak niet onder het toezicht van de AP valt. Als uitvloeisel daarvan zijn binnen de rechtspraak verschillende partijen belast met toezicht op de verwerking van persoonsgegevens. Deze versplintering mag er uiteraard niet toe leiden dat er binnen de rechterlijke macht uiteenlopende toezichtsregimes ontstaan. Uiteindelijk wordt de publieke waarde van gegevensbescherming door een uniform juridisch kader genormeerd en dat kader moet dan ook op uniforme wijze worden toegepast. Bij de ontwikkeling van algoritmen voor de rechtspraak zal dit risico, dat inherent is aan de wijze waarop het toezicht nu is vormgegeven, in acht moeten worden genomen. Daar komt bij dat kleine toezichthoudende organen binnen de rechtspraak niet zonder meer over de kennis beschikken om de verwerking van persoonsgegevens door algoritmen te controleren.

Naast de hiervoor gedefinieerde institutionele risico's zou, in het licht van de experimenten die op dit moment plaatsvinden, één specifieke algoritmische toepassing ook tot een potentiële kans kunnen leiden voor de verwezenlijking van de publieke waarde gegevensbescherming. Daarbij hebben wij het oog op het gebruik van algoritmen om rechterlijke uitspraken te anonimiseren. Dit soort algoritmen leveren vooral een kans op voor de gegevensbescherming, als de anonimisering

met behulp van algoritmen niet alleen efficiënter, maar ook beter zou geschieden. Het is evenwel maar zeer de vraag of het binnen afzienbare tijd technisch mogelijk zal zijn dit soort algoritmen tot stand te brengen. In beginsel zijn de algemene regels uit de anonimiseringsrichtlijn deterministisch en zij laten in de meeste gevallen weinig ruimte voor interpretatie. Zo dienen persoonsnamen van een procespartij vervangen te worden door bijvoorbeeld [eiser], [verweerder], [geïntimeerde], [verzoeker], [gedaagde], [verdachte], [belanghebbende], etc. Dergelijke regels zijn dan ook goed inpasbaar in een regelgebaseerd algoritme. Moeilijker wordt het evenwel met betrekking tot anonimiseringscriteria waarvoor meer interpretatie in het licht van de omstandigheden van het geval nodig is. Dit soort anonimiseringsbeslissingen vereist complexere zelflerende systemen, maar de techniek achter dit soort algoritmen is op dit moment nog niet in staat om nauwkeuriger dan mensen uitspraken te anonimiseren.

Omdat algoritmen nu nog onvoldoende nauwkeurig anonimiseren is door sommige geïnterviewden voorgesteld een hybride aanpak te hanteren. Daarbij structureren mensen de informatie op een wijze die door het algoritme eenvoudig kan worden herkend. Dat kan bijvoorbeeld door te werken met een standaard-format voor uitspraken, waarbij persoonsgegevens (in bepaalde zaaktypen) altijd slechts in een bepaalde paragraaf worden opgenomen of waarbij rechters informatie die van het standaardformat afwijkt, markeren.

Door de grotere efficiëntie waarmee rechterlijke uitspraken door de inzet van algoritmen kunnen worden geanonimiseerd, ontstaan echter wel kansen van andere aard. Enerzijds kan er voor medewerkers van de rechtspraak meer tijd overblijven voor andere (meer inhoudelijke) werkzaamheden en anderzijds zou de openbaarheid van rechtspraak kunnen toenemen. Een toename in de openbaarheid van jurisprudentie kan de publieke waarde van rechtsbescherming ten goede komen.

6.3.3 Tussenconclusie

De inzet van algoritmen in de rechtspraak leidt tot een aantal specifieke risico's voor de bescherming van persoonsgegevens. Die risico's zijn op de eerste plaats van institutionele aard. Zij vloeien voort uit de versplinterde wijze waarop het toezicht op de bescherming van persoonsgegevens binnen de rechtspraak is vormgegeven. Het is daarom zaak bij de verdere ontwikkeling van algoritmen voor de rechtspraak nadrukkelijk stil te staan bij de vraag wie toezicht houdt op de toepassing van algoritmen en of er bij die toezichthouders voldoende capaciteit bestaat om invulling te geven aan die toezichthoudende taak. Dat geldt in het bijzonder voor zover algoritmen toegepast gaan worden in de bedrijfsvoering. Ten tweede volgt uit de versplintering van het toezicht ook de waarschuwing dat voorkomen moet worden dat toezichtregimes uit elkaar gaan lopen.

De inzet van algoritmen biedt mogelijk ook een kans voor de bescherming van persoonsgegevens, vooral wanneer algoritmen ingezet zouden kunnen worden om uitspraken met een grotere nauwkeurigheidsgraad dan mensen dat doen te anonimiseren. Op dit moment zijn algoritmen daartoe evenwel nog niet in staat. De inzet van algoritmen ten behoeve van anonimisering kan echter ook op dit moment een belangrijke efficiëntieslag betekenen en er daarmee toe leiden dat meer tijd overblijft voor andere (rechterlijke) taken.

6.4 Non-discriminatie

6.4.1 Non-discriminatie en algoritmen in de rechtspraak

Een volgende waarde die in dit onderzoek centraal staat, is de waarde van non-discriminatie. Op de rechtspraak zijn vanzelfsprekend de algemene regels van non-discriminatierecht van toepassing. Voor een verdere uitwerking van deze waarde en die regels verwijzen wij terug naar par. 3.2 van dit rapport.

6.4.2 Kansen, risico's en bestendigheid juridisch kader

Het is zinvol om ook bij de beoordeling van de kansen en risico's op het terrein van non-discriminatie onderscheid te maken tussen het gebruik van algoritmen in de bedrijfsvoering en het gebruik van algoritmen in de oordeelsvorming. Wanneer algoritmen worden ingezet om de bedrijfsvoering te verbeteren, lijkt het risico op directe of indirecte discriminatie op verboden of verdachte gronden van onderscheid minimaal. Dat komt vooral doordat algoritmen op het niveau van de bedrijfsvoering de grootste toegevoegde waarde lijken te hebben in toepassingen waarbij persoonskenmerken geen directe of indirecte rol spelen. Het gaat dan onder meer om algoritmische beslissingen die gericht zijn op aantallen zaken, doorlooptijden van procedures, en het onderwerp van een procedure. Voor zover er risico's bestaan van ongerechtvaardigd onderscheid op andere gronden, hangen die risico's zo nauw samen met de wijze waarop een concreet algoritme is vormgegeven dat wij daar in deze studie niet op in gaan.

De risico's voor de waarde van non-discriminatie worden groter wanneer algoritmen worden gebruikt in de rechterlijke oordeelsvorming. Bij regelgebaseerde algoritmen houdt dit verband met het feit dat discriminatoire vooronderstellingen direct of indirect een rol kunnen spelen bij de selectie van variabelen die tot een beslissing leiden. Voor zelflerende algoritmen geldt daarbij dat zij moeten worden getraind op basis van gegevenssets. Het selecteren van volledig objectieve data of het 'wegfilteren' van irrelevant of ontoelaatbaar geachte gegevens (bijv. gegevens over etnische afkomst of geslacht) blijkt daarbij zeer moeilijk; dit laatste kan namelijk ook de functionaliteit van het algoritme raken. Daarnaast bewijst het experiment met het geautomatiseerd afhandelen van 'Mulderzaken' bij de Rechtbank Oost-Brabant dat er technische redenen kunnen

bestaan waardoor data aan objectiviteit kan inboeten.[51] In Mulderzaken worden namelijk geen vormvereisten gesteld aan de manier waarop het beroep bij de rechter wordt ingediend. Daardoor zijn veel processtukken handgeschreven. Het is technisch heel moeilijk om handgeschreven processtukken om te zetten in relevante invoer voor een algoritme. Omdat de huidige technologieën nog niet (goed) in staat blijken te zijn om dergelijke stukken te ontcijferen en in een voor computers leesbare vorm om te zetten, wordt in de pilot van de rechtbank Oost-Brabant gebruikgemaakt van door het openbaar ministerie opgestelde (proces)stukken.

Ook als aan alle ontwerpeisen is voldaan, kan de waarde van non-discriminatie onder druk komen te staan. Dat geldt zowel bij de toepassing van zelflerende algoritmen als beslisondersteuner als wanneer zij worden gebruikt als robotrechters. De oorzaak is gelegen in de omstandigheid dat zelflerende algoritmen tot hun uitvoer komen door eigenstandig statische verbanden te leggen. Daarbij kan de invoer van ogenschijnlijk neutrale variabelen en data ertoe leiden dat er verbanden worden gelegd die vooral negatieve effecten hebben voor personen met bepaalde persoonskenmerken (bijvoorbeeld mensen met een migratieachtergrond of met een handicap). Of dat gebeurt hangt voornamelijk af van de wijze waarop de variabelen door het algoritme ten opzichte van elkaar worden gewogen. Een voorbeeld waarnaar in dit kader vaak wordt verwezen is het in de Amerikaanse rechtspraak gebruikte algoritme COMPAS.[52] Dat systeem wordt onder andere gebruikt om een beoordeling te maken van de kans op recidive van een verdachte. Volgens tegenstanders van dit systeem is uit door dit systeem genomen beslissingen gebleken dat Afro-Amerikaanse burgers een grotere kans hebben om ten onrechte in een hoog-risico categorie geplaatst te worden dan andere bevolkingsgroepen.[53] Hoewel de etniciteit van een verdachte niet direct of indirect in het systeem werd ingevoerd, waren de waarden voor Afro-Amerikaanse verdachten vaker fout-positief en minder vaak fout-negatief dan voor verdachten met een andere etniciteit.[54] Hoewel het onderzoek waarin dit werd vastgesteld later van verschillende kanten is bestreden,[55] kan daaraan wel de voorzichtige gevolgtrekking worden verbonden dat wanneer complexe zelflerende algoritmen worden gebruikt, uit het feit dat de ingevoerde variabelen en data neutraal waren en ogenschijnlijk niet direct of indirect aan persoonskenmerken gerelateerd kunnen worden, niet automatisch kan worden afgeleid dat de uitkomst van het algoritme daarmee juridisch aanvaardbaar is. De uitkomst – die vooral wordt bepaald door de wijze waarop het algoritme

[51] Zie Van der Put, *Rechtstreeks* 2019, p. 50-60.
[52] Washington, *Colorado Technology Law Journal* 2018, p. 148; Fenton & Neil 2017. Het staat echter niet zonder meer vast dat COMPAS daadwerkelijk van een zelflerend algoritme gebruikmaakt. Die onduidelijkheid komt onder meer doordat het algoritme in particuliere handen is. Zie ook Rudin, *Nature Machine Intelligence* 2019, p. 209.
[53] Dieterich, Mendoza & Brennan 2016.
[54] Angwin e.a., propublica.org 23 mei 2016; Chouldechova 2017, p. 2. Afro-Amerikaanse verdachten die niet recidiveerden werden door COMPAS in 44,9% van de gevallen, naar later bleek, ten onrechte aangemerkt als recidivisten. Ten aanzien van witte verdachten was dit percentage 23,5. Witte verdachten daarentegen die *wel* recidiveerden werden in 47,7% van de gevallen door COMPAS aangemerkt als niet-recidivisten. Voor Afro-Amerikaanse verdachten was dat 28,0%. Daarmee lijkt COMPAS dus in het nadeel te werken van Afro-Amerikaanse verdachten, door het relatief hoge aantal fout-positieve kwalificaties van Afro-Amerikaanse verdachten tegenover een relatief hoog aantal fout-negatieve kwalificaties van witte verdachten. Zie Dressel & Farid, *Science Advances* 2018, p. 1-5.
[55] Zie Rudin, *Nature Machine Intelligence* 2019, p. 209.

variabelen weegt en met elkaar in verband brengt – zal uiteindelijk maatgevend zijn voor het oordeel of er in juridische zin sprake is van indirect onderscheid. Of dat indirecte onderscheid gerechtvaardigd kan worden hangt niet alleen af van de kwaliteit van de ingevoerde data, de validiteit en betrouwbaarheid van het algoritme, maar is uiteindelijk vooral een (juridisch-)ethische vraag over eerlijkheid en rechtvaardigheid en dus niet primair op basis van een technische analyse van het algoritme te beantwoorden.[56]

Het is in het kader van de kansen en risico's van belang op deze plaats ook op te merken dat algoritmen op het gebied van non-discriminatie niet altijd slechter presteren dan mensen. Ook menselijke rechters zijn immers niet perfect en ook zij kunnen in sommige gevallen beslissingen nemen die mogelijk gebaseerd zijn op diepgewortelde maatschappelijke vooroordelen of stereotypen.[57] Daar komt bij dat ook het menselijk beslissingsproces net als de algoritmische beslissing vaak is te karakteriseren als een *black box*.[58] Het voordeel dat een algoritme heeft ten opzichte van menselijke beslissingen is dat áls uiteindelijk wordt vastgesteld dat een reeks van beslissingen indirect discriminerende uitkomsten bevat, het meestal mogelijk zal zijn om het algoritme technisch bij te stellen (hoewel dat in technische zin lang niet altijd eenvoudig zal zijn). Menselijk gedrag bijsturen is in dat licht vaak moeilijker. De enige aanvullende juridische complexiteit die daarbij kan ontstaan, is dat het soms voor een algoritme alleen mogelijk zal zijn om te corrigeren voor discriminatoire uitkomsten als ook bepaalde persoonskenmerken in het algoritme worden ingevoerd.[59] Dergelijke gegevens zijn echter vaak bijzondere categorieën persoonsgegevens die op grond van art. 9 AVG in beginsel niet verwerkt mogen worden. In art. 22 UAVG is een aantal uitzonderingen op dit verbod neergelegd, die in dit soort situaties in de toekomst mogelijk uitkomst kunnen bieden.

Voorgaande laat al zien dat het toepassen van algoritmen op individuele rechterlijke procedures in de vorm van beslisondersteuning of als robotrechter, niet alleen tot risico's voor de waarde van non-discriminatie leidt, maar ook kansen kan opleveren. Daar komt bij dat de toepassing van algoritmen kan bijdragen aan een grotere mate van rechtsgelijkheid of rechtseenheid. Beslisondersteuning van rechters door algoritmen kan er in potentie bijvoorbeeld toe leiden dat in gelijke gevallen op basis van analyses van dezelfde relevante informatie (bijvoorbeeld jurisprudentie) beslissingen worden genomen. Die informatie wordt dan niet meer op basis van zelfgeformuleerde zoekcriteria gevonden en de verbanden ertussen worden niet meer per geval door een individuele medewerker of rechter geanalyseerd, maar in gelijksoortige procedures worden de analyses op eenzelfde wijze door het algoritme aangereikt. Daarbij dient evenwel te

[56] Bijlsma, Meynen & Bex, *NJB* 2019, p. 3313-3319; Barocas, Hardt & Narayanan 2015, in het bijzonder hoofdstuk 2; Kleinberg, Mullainathan & Raghavan 2016; Chouldechova 2017; Pleis e.a. 2017.
[57] Fulda, *Artificial Intelligence and Law* 2012, p. 321-333.
[58] Kleinberg e.a., *Journal of Legal Analysis* 2018, p. 113-174.
[59] Daarover ook Brief van de Minister voor Rechtsbescherming van 8 oktober 2019, *Kamerstukken II* 2019/20, 26643 en 32761, nr. 641, p. 5 en 11.

worden aangetekend dat de nu bestaande algoritmen technisch gezien nog niet in staat zijn om deze belofte te verwezenlijken. Een andere kans die algoritmen bieden, is dat zij ertoe bij zouden kunnen dragen dat eventuele inconsistenties, bijvoorbeeld ten gevolge van (onbewuste) vooroordelen, in rechterlijke redeneringen worden gesignaleerd. Rechters en de organisatie waarin zij werken kunnen hun gedrag en beleid naar aanleiding daarvan dan aanpassen.[60]

Ook op het gebied van organisatie, bedrijfsvoering en management kan de inzet van algoritmen een kans bieden voor de verwezenlijking van rechtsgelijkheid en daarmee samenhangende rechtseenheid. Zo volgt uit de afgenomen interviews dat het gebruik van algoritmen in de bedrijfsvoering ertoe zou kunnen leiden dat zaken met betrekking tot gelijksoortige rechtsvragen meer gecoördineerd worden behandeld dan nu het geval is. Dit zou de voorspelbaarheid van uitspraken ten goede kunnen komen, omdat het kan leiden tot een consistentere aanpak van gelijke rechtsvragen.

6.4.3 Tussenconclusie

De toepassing van algoritmen in de rechtspraak brengt het risico met zich dat ongeoorloofde stereotypen en vooroordelen structureel een rol gaan spelen bij de rechterlijke besluitvorming. Weliswaar kunnen ook de beslissingen van menselijke rechters beïnvloed zijn door (maatschappelijke) vooroordelen of stereotypen, maar algoritmen hebben vaak een groter toepassingsbereik (in die zin dat hetzelfde algoritme voor een zeer groot aantal gevallen patroonanalyses kan maken en voorspellingen kan doen) en de vooroordelen die in deze systemen besloten liggen kunnen daarmee bredere gevolgen hebben. Het is dus zaak op dit punt extra waakzaam te zijn. Dit betekent dat bij het ontwerp van de algoritmen – dat geldt zowel voor regelgebaseerde algoritmen als voor zelflerende algoritmen – voldoende acht moet worden geslagen op de waarde van non-discriminatie. Daartoe is onder meer van belang dat voldaan wordt aan bepaalde ontwerpeisen, zoals het gebruik van neutrale data. Daarbij geldt dat vooral bij zelflerende algoritmen altijd het risico blijft bestaan dat het algoritme op basis van objectieve of op het oog neutrale variabelen, verbanden legt die in juridische zin als indirecte discriminatie kwalificeren.

De verwezenlijking van dit soort risico's lijkt op het eerste gezicht minder groot wanneer algoritmen worden gebruikt in de bedrijfsvoering van de rechtspraak. Dat komt vooral doordat de variabelen die de algoritmische besluitvorming op het niveau van de bedrijfsvoering sturen, in beginsel niet gerelateerd zijn aan persoonskenmerken.

De grootste kans die het gebruik van algoritmen biedt voor het verwezenlijken van de waarde van non-discriminatie, is dat het recht door de inzet van algoritmen gelijkvormiger zal worden

[60] Daarover ook Lodder e.a. 2014, p. 49.

toegepast. Dat geldt zeker als algoritmen worden gebruikt op het niveau van de bedrijfsvoering bij de toedeling van zaken opdat gelijksoortige rechtsvragen door dezelfde rechters worden behandeld. Mogelijk kan deze kans zich ook voordoen als algoritmen worden ingezet ter ondersteuning van het beslisproces van de rechter, omdat algoritmen dan de voor een procedure relevante informatiebronnen op meer eenvormigere wijze zouden kunnen ontsluiten.

6.5 Rechtsbescherming

6.5.1 Rechtsbescherming en algoritmen in de rechtspraak

De laatste publieke waarde die in deze casestudy centraal staat, is de waarde van rechtsbescherming. Deze waarde kan worden geconcretiseerd aan de hand van het recht op een eerlijk proces en het recht op een effectief rechtsmiddel.[61] Om te komen tot een gestructureerde bespreking is het zinvol nader onderscheid te maken tussen verschillende aspecten van deze waarde. Het gaat daarbij om de eisen die worden gesteld aan de rechterlijke procedure; de eisen die worden gesteld aan de beslissing en het beslissingsproces van de rechter; en de eisen die worden gesteld aan de persoon van de rechter.

6.5.1.1 Eisen aan de rechterlijke procedure als zodanig

Uit het recht op een eerlijk proces volgt een aantal eisen ten aanzien van de rechterlijke procedure als zodanig. Ten eerste hebben procespartijen recht op een gelijke procespositie. Dit uitgangspunt van *equality of arms* brengt onder andere met zich dat partijen in gelijke mate het recht en de mogelijkheid moeten hebben om voor de procedure relevant materiaal aan te dragen. Daarnaast moeten zij door andere partijen aangedragen materiaal kunnen inzien en betwisten.[62]

Ten tweede volgt uit het recht op een eerlijk proces dat er binnen een redelijke termijn uitspraak moet worden gedaan.[63] De beoordeling of sprake is van een redelijke termijn, vindt daarbij plaats in het licht van de omstandigheden van een geval, waarbij in ieder geval belang toekomt aan de complexiteit van de voorliggende zaak, alsmede aan het gedrag van de rechtzoekende en de betrokken autoriteiten.[64] In dit kader moet worden opgemerkt dat de rechtspraak ernaar streeft om in de periode 2015-2020 rechtszaken 40% korter te laten duren dan in 2014.[65]

[61] Zie in het bijzonder art. 17 Gw, artt. 6 en 13 EVRM en art. 47 Handvest.
[62] Zie in deze zin Vetzo, Gerards & Nehmelman 2018, p. 170 e.v.; Gerards 2019, p. 168; Philipsen & Themeli 2020.
[63] Zie voor een wettelijke verankering van die eis onder meer art. 6 EVRM.
[64] EHRM 28 oktober 1999, ECLI 24846/94 34165/96 34166/96 (*Zielinski en Pradal & Gonzalez/Frankrijk*), r.o. 65.
[65] *Agenda van de Rechtspraak 2015-2018*, uitgave van de Rechtspraak van mei 2014. Deze agenda is verlengd van 2018 tot 2020.

6.5.1.2 Eisen aan de rechterlijke beslissing

Als afzonderlijk aspect van de publieke waarde van rechtsbescherming moeten de eisen worden genoemd die worden gesteld aan de rechterlijke beslissing. Rechters dienen, als uitvloeisel van het recht op een eerlijk proces, hun uitspraken te motiveren en in het openbaar te doen. Volgens art. 121 Gw houdt deze motiveringsplicht in dat een uitspraak de gronden bevat waarop zij rust.[66]

De eis dat uitspraken een dragende motivering moeten bevatten en in het openbaar moeten worden gedaan, is in het bijzonder van belang omdat procespartijen alleen op basis van een motivering die aan hen kenbaar is gemaakt rechtsmiddelen, zoals hoger beroep, kunnen aanwenden.[67] De door de wetgever opengestelde mogelijkheid van hoger beroep biedt rechtszoekende de gelegenheid om uitspraken door een andere rechter te laten toetsen. Daarmee kunnen fouten in de eerdere procedure worden hersteld. Op zichzelf wordt door het openstellen van de mogelijkheid van hoger beroep, welke mogelijkheid niet in alle gevallen dwingend uit het recht op een eerlijk proces voortvloeit, de publieke waarde van rechtsbescherming ook gediend.[68]

6.5.1.3 Eisen aan de rechter

De waarde van rechtsbescherming vertaalt zich ook in eisen die moeten worden gesteld aan de rechter. Deze eisen zijn in de meest concrete zin terug te vinden in de kernwaarden van de rechtspraak. Zij komen ook tot uitdrukking in het grondrecht op een eerlijk proces. De belangrijkste kernwaarden zijn daarbij onafhankelijkheid, onpartijdigheid en autonomie van de rechter.[69] Essentieel voor de onafhankelijkheid van de rechterlijke macht is dat de rechter bij de uitoefening van de rechtsprekende taak wordt beschermd tegen invloeden van buitenaf.[70] Een kernmerk daarvan is dat een rechter autonoom handelt, hetgeen betekent dat hij zelfstandig optreedt en zelfstandig tot een oordeel te komt.[71] Waar onafhankelijkheid van de rechter vooral betrekking heeft op de relatie tussen de rechter en andere overheidsorganen, ziet de onpartijdigheid op de verhouding tussen de rechter en de procespartijen.[72] Deze onpartijdigheid moet zowel subjectief van aard zijn (de rechter mag geen persoonlijke betrokkenheid hebben bij de zaak of bij de partijen) als objectief (er mag geen schijn van partijdigheid of vooringenomenheid van de rechter bestaan).

---

[66] Zie onder meer ook artt. 30, 230 en 287 lid 1 Rv, artt. 358 en 359 Sv, art. 3:46 Awb en art. 5 RO.
[67] Zie bijvoorbeeld artt. 332 en 398 Rv, art. 427 Sv en art. 8:1 Awb.
[68] Op grond van art. 26 Wet RO dienen de gerechten ook te voorzien in een klachtenregeling, maar die klachtenregelingen bieden geen bescherming tegen (veronderstelde) onjuiste uitspraken. Ook kan volgens art. 2 van de modelklachtenregeling niet geklaagd worden over de inhoud en de motivering van een beslissing van een met rechtspraak belaste rechterlijke ambtenaar, noch over de totstandkoming van een rechterlijke beslissing met inbegrip van de in dat kader genomen beslissingen van procedurele aard.
[69] De overige kernwaarden van integriteit, professionaliteit en deskundigheid spelen in dit kader dan ook geen zelfstandige rol. Zie de navolgende literatuur voor uitgebreider beschouwingen over de kernwaarden van de rechter: Van Emmerik, Loof & Schuurmans 2014; *Visie op de rechtspraak* 2010; *NVvR-rechterscode* 2011; Van Ettekoven & Prins 2018, p. 442. Voor de eisen met betrekking tot onafhankelijkheid en onpartijdigheid, zie art. 6 EVRM, art. 47 Handvest en art. 14 IVBPR. Deze eisen zijn niet als zodanig expliciet in de Nederlandse Grondwet opgenomen. Zie daarover ook Van Emmerik, Loof & Schuurmans 2014, p. 15.
[70] Van den Eijnden 2011, p. 4.
[71] *NVvR-rechterscode* 2011, p. 4 (punt 2.2).
[72] Van den Eijnden 2011, p. 7; Van Orshoven 2001, p. 79.

Overigens merken wij hier ten overvloede op dat het onder bijzondere omstandigheden mogelijk is de overheid aansprakelijk te stellen voor fouten die hebben geleid tot een schending van fundamentele rechtsbeginselen waardoor van een eerlijke en onpartijdige behandeling van de zaak niet meer kan worden gesproken.[73] De mogelijkheid van een aansprakelijkheidsstelling wegens onrechtmatige rechtspraak is het laatste instrument dat aan rechtszoekenden ter beschikking staat om de waarde van rechtsbescherming veilig te stellen.[74] Voorwaarde voor het instellen van een vordering op grond van onrechtmatige rechtspraak is wel dat alle andere rechtsmiddelen zijn uitgeput.

6.5.2 Kansen, risico's en bestendigheid juridisch kader

De inzet van algoritmen in de rechtspraak leidt zowel tot kansen als tot risico's voor de publieke waarde van rechtsbescherming. Om die kansen en risico's gestructureerd te bespreken, maken wij ook hier een onderscheid tussen de inzet van algoritmen in de bedrijfsvoering en organisatie, en de inzet van algoritmen ten behoeve van de rechterlijke oordeelsvorming. Vooraf merken wij op dat de inzet van algoritmen in ieder geval altijd wordt gemotiveerd door de wens om (beslissings)processen efficiënter te laten verlopen, zonder daarbij te hoeven inboeten op nauwkeurigheid. Uit de door ons afgenomen interviews volgt dat dit ook voor de rechtspraak geldt. Als in de toekomst tot de inzet van algoritmen wordt besloten op een wijze die in dit onderzoek is beschreven en als die inzet tot tijdswinst leidt bij personen die rechtstreeks bij de rechtspraak zijn betrokken, kan dat ten goede komen aan de rechtsbescherming. De gedacht is dan dat deze personen meer tijd overhouden om zich met de beantwoording van (belangrijke) rechtsvragen bezig te houden. Hoewel deze algemene kans in onze interviews met een zekere regelmaat werd genoemd, is het in het licht van de huidige stand van de techniek op dit moment lastig in te schatten in hoeverre zij realiseerbaar is.

6.5.2.1 Het gebruik van algoritmen in de bedrijfsvoering en organisatie

Algoritmen in de bedrijfsvoering van de rechtspraak lijken met name gebruikt te kunnen worden bij de toedeling van zaken en het monitoren van zaaksstromen. Volgens het EHRM moet de toedeling van zaken aan gerechten en aan specifieke rechters gebeuren aan de hand van vooraf vastgelegde objectieve, duidelijke en transparante factoren.[75] In februari 2020 is daaraan door de

[73] Zie daarover: HR 17 maart 1978, ECLI:NL:PHR:1978:AC6217, *NJ* 1979/204 m.nt. M. Scheltema; Lindenbergh, in: *T&C BW* 2019.
[74] Zie in dat kader, ook voor verdere literatuurverwijzingen: Uzman & Boogaard, *Overheid & Aansprakelijkheid* 2017, p. 65.
[75] EHRM 12 januari 2016, ECLI:CE:ECHR:2016:0112JUD005777413 (*Miracle Europe Kft/Hongarije*) en EHRM 5 oktober 2010, *EHRC* 2011, 3, m.nt. E. Mak (*OMO Group/Slowakije*). De Venetië Commissie overweegt hierover: 'A predictable and transparent predetermined plan for case assignment is important not only to guarantee an independent and impartial tribunal established by law, as required under Article 6 of the European Convention, but also to guarantee the internal independence of the judges' (Venice Commission 2019, par. 108). Zie in dit kader ook een eerder rapport van de Venetië Commissie, waarin wordt overwogen dat: 'In order to enhance impartiality and independence of the judiciary it is highly recommended that the order in which judges deal with the cases be determined on the basis of *general criteria*. This can

rechtspraak invulling gegeven door de publicatie van een Code Zaakstoedeling. Ook daarin wordt expliciet verwezen naar de mogelijkheid om zaken op geautomatiseerde wijze aselect toe te wijzen.[76] Daartoe zou gebruik kunnen worden gemaakt van regelgebaseerde systemen, wat kan leiden tot een grotere uniformiteit dan het geval was tot februari 2020.[77] Dat komt vooral doordat de eerste systemen die hiervoor gebruikt kunnen worden waarschijnlijk regelgebaseerd zijn. Zij zullen de criteria op basis waarvan zaken worden verdeeld en toegewezen op eenduidige wijze toepassen.

Naast de aselecte toewijzing aan de hand van algoritmen wordt in de interviews ook genoemd dat de inzet van dergelijke systemen een kans in zich kan dragen voor de rechtseenheid. Zo kan de inzet van algoritmen eraan bijdragen dat procedures die gaan over eenzelfde persoon, onderwerp of rechtsvraag bij dezelfde rechters terechtkomen. Op deze wijze kan voorkomen worden dat verschillende rechters binnen hetzelfde rechtscollege uiteenlopende uitspraken doen over dezelfde rechtsvraag. Indien gewenst kan dan bovendien rekening worden gehouden met de omstandigheid dat een rechter zich over een zaak moet buigen waarbij hij al eerder betrokken is geweest. Dit is een kans vanuit het oogpunt van onafhankelijkheid en onpartijdigheid.

Tegelijkertijd draagt gerichte toewijzing aan de hand van een algoritmisch systeem een risico in zich. Rechters die door de toebedeling via een algoritme heel vaak met eenzelfde type rechtsvraag geconfronteerd worden, krijgen weinig tegenspraak meer als het gaat om hun eigen standpunten en kunnen bovendien een grote invloed krijgen op de rechtspraak over een bepaald onderwerp.[78]

De inzet van algoritmen kan verder als voordeel hebben dat op het niveau van de bedrijfsvoering beter inzicht kan ontstaan in allerlei kwaliteitsmaatstaven en productienormen, zoals die ten aanzien van doorlooptijden.[79] Uit de door ons afgenomen interviews blijkt dat dit overzicht nu maar heel beperkt bestaat. Voor zover het inzicht in dit soort cijfers wel bestaat, wezen enkele geïnterviewden erop dat die cijfers niet buiten de specifieke gerechten kenbaar zijn. Meer en een

be done for example on the basis of alphabetical order, on the basis of a computerised system or on the basis of objective criteria such as categories of cases. The general rules (including exceptions) should be *formulated by the law or by special regulations on the basis of the law*, eg in court regulations laid down by the presidium or president. It may not always be possible to establish a fully comprehensive abstract system that operates for all cases, leaving no room to decisions regarding allocation in individual cases. There may be circumstances requiring a need to take into account the *workload* or the *specialisation* of judges. Especially complex legal issues may require the participation of judges who are expert in that area. (...) The criteria for making such decisions by the court president or presidium should, however, be defined in advance.' (Venice Commission 2012, par. 91). Ook het European Network of Councils for the Judiciary beveelt aan dat er een methode voor het toedelen van zaken moet worden vastgesteld die toegankelijk is voor het publiek (European Network of Councils for the Judiciary 2014, aanbeveling 2).

[76] Zie preambule Code Zaakstoedeling, vastgesteld door het Overleg van presidenten van de gerechten en leden van de Raad voor de rechtspraak op 27 januari 2020.

[77] Voor meer daarover in kritische zin: Grimmelikhuijsen, *Rechtstreeks* 2018, p. 21; Van Emmerik & Schuurmans, *NJB* 2016, p. 795-799; Ingelse 2010, p. 192.

[78] Daarbij is van belang dat sinds februari 2020 als uitgangspunt de aselecte toewijzing geldt.

[79] Zie over de inzet van *big data* in het kader van de doelmatigheid van de rechtspraak ook Lodder e.a. 2014, p. 47-49.

betere (geautomatiseerde) uitwisseling van die informatie kan vermoedelijk bijdragen aan de efficiëntie van de rechtspraak.

Een risico dat verbonden is aan beter inzicht in het functioneren van de rechtspraak is dat de cijfers over het functioneren van gerechten en de rechters daarbinnen ook tegen individuele rechters gebruikt kunnen worden. Als tot de inzet van algoritmen in de bedrijfsvoering wordt besloten, zal daarom onder andere nagedacht moeten worden over de vraag in hoeverre rechters verantwoordelijk kunnen worden gehouden voor hun productie (ten opzichte van hun collega's of vooraf vastgestelde productienomen) en welke analyses de besturen binnen de rechterlijke macht met die cijfers mogen uitvoeren, bijvoorbeeld ten aanzien van de vraag hoe vaak individuele rechters bepaalde vorderingen toe- of afwijzen, of hoe vaak zij bepaalde procesrechtelijke instrumenten – zoals het stellen van prejudiciële vragen – inzetten.

Andere risico's van het gebruik van algoritmen op het niveau van de bedrijfsvoering voor de waarde rechtsbescherming kunnen ontstaan wanneer gegevens over het presteren en het functioneren van gerechten niet alleen dienen voor intern gebruik, maar ook beschikbaar zijn voor externe partijen, zoals de procespartijen. Informatie over aantallen toe- of afwijzingen kan dan bijvoorbeeld *forumshopping* in de hand werken.[80] Voor een belangrijk deel is dit ook nu al niet te voorkomen; soortgelijke informatie kan bijvoorbeeld ook al uit analyses van de gepubliceerde jurisprudentie worden gehaald. Sommige – vooral grote – advocatenkantoren beschikken dan ook al over dergelijke analyses. Als het aantal gepubliceerde uitspraken door de inzet van algoritmen toeneemt, en als het bovendien mogelijk wordt om veel meer uitspraken veel sneller en systematischer te analyseren, kan ook het risico toenemen dat procespartijen hun keuzes gaan afstemmen op cijfers over het functioneren van de rechtspraak. Waar sommige advocatenkantoren wel al over deze technologie beschikken en er gebruik van maken (zoals uit de interviews is gebleken) en andere niet, kan dit bovendien ongelijkheid tussen de procespartijen in de hand werken en kan hierdoor ook de *equality of arms* nadelig worden beïnvloed.

Een laatste kans die het gebruik van algoritmen op het niveau van de bedrijfsvoering biedt, is gelegen in de inzet van algoritmen voor de anonimisering van uitspraken, zoals ook al is besproken ten aanzien van het recht op gegevensbescherming. De inzet van algoritmen ten behoeve van de anonimisering van rechterlijke uitspraken moet het in de toekomst mogelijk maken om sneller te anonimiseren. Op dit moment vindt dat proces nog handmatig plaats. Dit arbeidsintensieve werk leidt ertoe dat slechts een zeer beperkt aandeel (2 tot 3%) van de gerechtelijke uitspraken online

[80] Zie bijvoorbeeld Kreulen, *Trouw* 26 augustus 2019; Dunk, *Advocatenblad* 28 augustus 2019. Hier kan melding worden gemaakt van de Franse wet die het gebruik van de persoonsgegevens van rechters en griffiers met als doel om hun (veronderstelde) professionele activiteiten te evalueren, te analyseren, te vergelijken of te voorspellen verbiedt. Zie Art. 33 LOI n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice. Over de implicaties voor Nederland, zie ook de antwoorden van 27 maart 2020 van de Minister voor Rechtsbescherming op de Kamervragen inzake algoritmische analyse van vonnissen, *Aanhangsel Handelingen II* 2019/20, nr. 2255.

wordt gepubliceerd.[81] Door algoritmen in te zetten voor de anonimisering zouden in de toekomst meer uitspraken gepubliceerd kunnen worden. Daardoor zou het voor advocaten en procespartijen gemakkelijker kunnen worden om een realistische inschatting van hun positie te maken. Door meer uitspraken te publiceren kan bovendien de voorspelbaarheid van de rechtspraak potentieel worden vergroot, zeker wanneer die uitspraken op hun beurt weer met behulp van algoritmen kunnen worden geanalyseerd op de aanwezigheid van patronen.[82]

Hier komt bij dat momenteel vooral de bijzondere en nieuwswaardige uitspraken worden gepubliceerd. De grotere beschikbaarheid van uitspraken, waaronder vooral ook de meer standaarduitspraken, maakt het mogelijk om beter (rechts)wetenschappelijk onderzoek naar de jurisprudentie te doen. In het verlengde daarvan ontstaat voor de markt een kans om data-gedreven initiatieven, zoals de website magontslag.nl, te ontplooien. Op deze website kunnen individuen aan de hand van verschillende tools een inschatting maken van bijvoorbeeld de rechtmatigheid van hun ontslag of hun recht op een ontslagvergoeding. Dit soort tools is afhankelijk van de door de rechtspraak beschikbaar gestelde informatie. Als het gemakkelijker wordt voor de ontwikkelaars van dit soort tools om aan informatie te komen, of de kwaliteit van de beschikbare informatie toeneemt, kan dat toekomstige procespartijen een kans bieden om een realistischer beeld te krijgen in hun procespositie. Daarmee komen ook tools als magontslag.nl in potentie de rechtszekerheid en de effectieve toegang tot rechtsmiddelen ten goede.[83]

6.5.2.2 Het gebruik van beslisondersteuningsalgoritmen

Hiervoor is er al op gewezen dat de inzet van beslisondersteuningsalgoritmen bepaalde voordelen kan hebben vanuit het perspectief van efficiëntie en rechtseenheid. Het gebruik van algoritmen als beslisondersteuning gaat echter gepaard met een aantal risico's voor de waarde van rechtsbescherming. Ten eerste kwam in de door ons afgenomen interviews naar voren dat sommige geïnterviewden vrezen dat het gebruik van algoritmen in de ondersteuning van rechters de rechterlijke onafhankelijkheid raakt. De vrees is in het bijzonder dat softwareontwikkelaars via de ontwikkeling en training van algoritmen indirect invloed krijgen op de wijze waarop rechters hun taak uitoefenen, of dat rechters afhankelijk worden van softwareontwikkelaars.[84] Hoewel dit

[81] Zie ook *Aanhangsel Handelingen II* 2019/20, nr. 2255, p. 3. Zie op dit punt onder meer Mommers, Zwenne & Schermer, *NJB* 2010, p. 2072-2078. In een recente uitspraak in een vreemdelingenzaak van de ABRvS kwam dit aspect ook naar voren: 'Het arbeidsintensieve proces van anonimisering in combinatie met het grote aantal uitspraken maakt integrale publicatie van alle uitspraken op dit moment onmogelijk. De Raad voor de rechtspraak heeft verklaard wel te streven naar integrale publicatie van alle uitspraken. Een belangstellende die op zoek is naar een specifieke uitspraak, kan zich daarnaast wenden tot de griffie van de rechtbank'. Zie ABRvS 9 oktober 2019, ECLI:NL:RVS:2019:3410, r.o. 4.3.

[82] De Meij e.a. 2006, *Mediaforum* 2006, p. 121-142.

[83] Ook Frits Bakker, voormalig voorzitter van de Raad voor de rechtspraak, heeft aangegeven dat de rechtsstaat gediend zou zijn met een grotere openbaarheid van uitspraken. Zie Bakker, *NJBlog* 17 augustus 2017. Een belangrijke nuancering daarbij is wel dat hij aangeeft dat het van belang blijft dat uitspraken niet alleen te bestuderen zijn door middel van 'big data-technieken', maar dat ook rechtstreekse bestudering van uitspraken mogelijk dient te blijven.

[84] De rol van technologie als derde partij en van de leverancier daarvan als vierde partij is uitvoeriger beschreven in het domein van de alternatieve geschilbeslechting. Zie daarover: Katsh & Rifkin 2001; Lodder, *Information and Communications Technology Law* 2006, p. 143-155; Carneiro e.a., *Artificial Intelligence Review* 2014, p. 214.

probleem zich ook kan voordoen bij de toepassing van algoritmen op het niveau van de bedrijfsvoering, zoals hierboven omschreven, wordt de mogelijke impact groter als rechters beslissingen in individuele zaken baseren op algoritmen. De kans dat dit gevaar wordt verwezenlijkt, neemt toe wanneer binnen de rechtspraak zelf onvoldoende kennis bestaat om het functioneren van het algoritme te controleren en om als kritische gesprekspartner (klant) van ontwikkelaars te kunnen fungeren. Het wordt daarom verstandig gevonden voor de rechterlijke macht om te voorkomen dat zij gebonden raakt aan een specifieke toeleverancier van algoritmen en niet meer op ieder gewenst moment vrijelijk naar een andere leverancier – lees: een ander algoritme – kan overstappen. Het voorkomen van een dergelijke *vendor lock-in* betekent dat de rechtspraak minder kwetsbaar wordt voor beïnvloeding van buitenaf. Voor zover de rechterlijke macht besluit gebruik te maken van algoritmen die zij niet zelf heeft ontwikkeld, wordt het daarnaast door veel geïnterviewden van belang gevonden dat de rechterlijke macht zelf eigenaar is van de ontwikkelde algoritmen en dat alle betrokkenen (met name de gebruikers, zoals rechters) voldoende informatie krijgen over de werking van de systemen (technische transparantie).[85] Om te voorkomen dat ontwikkelaars van algoritmen via het ontwerp van het algoritme inhoudelijke beslissingen nemen, die later door een rechter worden overgenomen, is het verder van belang dat rechters een voldoende kritische houding aannemen ten opzichte van suggesties die door een beslisondersteuningsalgoritme worden gedaan. Deze eis wordt onderstreept door de eis van rechterlijke onafhankelijkheid. Het is ook in dat licht in ieder geval van belang dat rechters voldoende kennis en begrip hebben over de werking van het systeem dat zij ter ondersteuning van hun beslissing gebruiken. In het bijzonder moeten zij kunnen begrijpen aan de hand van welk soort criteria, parameters, factoren en data het systeem een bepaalde suggestie doet.[86] Dit kan bijvoorbeeld worden vormgegeven door een rechter als gebruiker inzicht te bieden in filters die een algoritmisch systeem gebruikt om tot een informatieselectie te komen, en hem daar ook zelf keuzes in te bieden. Dit laat echter ook al zien dat de risico's voor de rechterlijke onafhankelijkheid groter worden naarmate de algoritmen die in de ondersteuning van rechters worden gebruikt complexer zijn; dat maakt hun toepassing immers ook voor de rechters minder inzichtelijk en minder controleerbaar. In de door ons afgenomen interviews werd erop gewezen dat het daarom in ieder geval van belang is om bij de inzet van algoritmen in de rechtspraak rechters te wijzen op hun rechterlijke autonomie als deelaspect van de rechterlijke onafhankelijkheid. Een onderdeel van die autonomie is dat een rechter zelf verantwoordelijk is voor zijn uitspraak, ook als hij een

[85] Zie in dat kader ook het principe van 'user control' als zelfstandige beoordelingsfactor over de wenselijkheid van het gebruik van algoritmen in de rechtspraak in de *Ethical Charter on the use of artificial intelligence in judicial systems,* (Adopted at the 31st plenary meeting of the CEPEJ (Strasbourg, 3-4 December 2018)), Council of Europe, februari 2019, Principle 5.

[86] Alhoewel dergelijke zorgen vanzelfsprekend ook kunnen bestaan bij de inzet van meer klassieke juridische zoekmachines door rechters. Daar is immers ook lang niet altijd duidelijk waarom een bepaald resultaat gepresenteerd wordt.

algoritme heeft geraadpleegd.[87] Als een rechter niet voor de inhoud van door algoritmen aangereikte informatie kan instaan, zou hij deze dan ook niet moeten gebruiken.

Een ander risico van de inzet van beslisondersteuningsalgoritmen betreft de betekenis van (hoger) beroepsmogelijkheden. Zolang algoritmen niet feilloos functioneren, biedt de mogelijkheid om in hoger beroep te gaan een belangrijke waarborg tegen fouten die door algoritmen kunnen worden gemaakt; de hogerberoepsrechter kan dan immers de inhoudelijke en juridische kwaliteit van de rechterlijke uitspraak zelfstandig toetsen. Om op betekenisvolle wijze invulling te kunnen geven aan deze waarborg moet aan ten minste twee voorwaarden zijn voldaan. In de eerste plaats moet voorkomen worden dat hoger beroepsrechters gebruikmaken van dezelfde beslisondersteuningsalgoritmen als rechters in eerste aanleg. Anders zou immers het risico bestaan dat fouten van het algoritme in hoger beroep worden herhaald. Daarnaast moet de manier van functioneren van het algoritme dat door de lagere rechter gebruikt is bij het nemen van de beslissing voldoende duidelijk zijn. Alleen dan kan tegen het onderdeel van de uitspraak dat steunt op of mede gebaseerd is op een algoritme, zinvol worden opgekomen. Zolang gewerkt wordt met een regelgebaseerd systeem dat het karakter heeft van een beslisboom, is het vaak mogelijk om te herleiden hoe een systeem tot een bepaalde conclusie kwam. Bij zelflerende algoritmen is dat veel moeilijker en soms zelfs principieel onmogelijk. In dat geval verzet de waarde van rechtsbescherming zich tegen het gebruik van algoritmen.

Overigens staat op dit moment niet in alle procedures hoger beroep open. Vooral in eenvoudige procedures met een klein monetair belang is de mogelijkheid van hoger beroep niet gegarandeerd. Het lijkt verstandig om als algoritmische systemen in die procedures een rol gaan spelen in de besluitvorming, hoger beroep steeds mogelijk te maken.

6.5.2.3 Het gebruik van algoritmen als robotrechter

Mochten robotrechters technisch gezien inzetbaar worden, dan kunnen zij als voordeel hebben dat procedures daardoor in potentie veel sneller zouden kunnen worden afgedaan. Dit geldt vooral voor de eenvoudige geschillen waarop de toepassing van algoritmen als robotrechter het best voorstelbaar is. De beschikbare rechters zouden zich dan op andere, meer complexe zaken en rechtsvragen kunnen richten.

De inzet van algoritmen als robotrechter zou tegelijkertijd echter leiden tot aanzienlijke risico's voor de publieke waarde van rechtsbescherming. De omvang van die risico's hangt samen met het

[87] Dit onderzoek heeft uitsluitend betrekking op kansen en risico's die samenhangen met de inzet van algoritmen als zodanig. Op deze plaats wordt evenwel opgemerkt dat sommige risico's die samenhangen met de inzet van algoritmen door technologische ingrepen kunnen worden gemitigeerd. Dat kan bijvoorbeeld door goed na te denken op welke wijze – via welke interface – het algoritme aan de rechter wordt gepresenteerd. Denk bijvoorbeeld aan het inbouwen van de eis dat een bepaald aantal wijzigingen in de automatisch gegenereerde uitspraak altijd noodzakelijk is.

soort algoritme dat wordt gebruikt. Zoals reeds werd opgemerkt is het aannemelijk dat, voor zover het al mogelijk zal zijn dat algoritmen in de komende vijf tot tien jaar als robotrechter kunnen functioneren, het waarschijnlijk zal gaan om regelgebaseerde algoritmen die worden toegepast om eenvoudige rechtsvragen te beantwoorden. Als voorbeeld kan daarbij worden gewezen op simpele ontvankelijkheidsvragen; zoals de vraag of het griffierecht wel is betaald of de vragen naar de beroepstermijnen. De rechterlijke procedure is in dit soort gevallen in hoge mate iteratief van karakter. Volgens sommige geïnterviewden zouden de meest eenvoudige procedures zich daarmee kunnen lenen voor de inzet van regelgebaseerde algoritmen, zonder dat daarbij al te grote risico's bestaan voor – bijvoorbeeld – de kwaliteit van rechtspraak of de procedurele rechten van procespartijen.

Als sprake is van meer discretionaire ruimte in het beslissingsproces en dus ruimte bestaat voor verschillende verdedigbare uitkomsten, bereiken regelgebaseerde systemen echter al snel de grenzen van het toepassingsbereik. Complexere algoritmen zijn op dit moment technisch nog niet in staat om zelfstandig een oordeel te geven in rechterlijke procedures waarin regels met discretionaire ruimte moeten worden toegepast. Of dat in de toekomst wel mogelijk zal zijn, valt nu niet te voorspellen en over het antwoord op die vraag bestaat zoals reeds beschreven debat.

Als op enig moment besloten zou worden tot de inzet van regelgebaseerde algoritmen ten behoeve van het geautomatiseerd afdoen van duidelijk omlijnde rechtsvragen, stelt de publieke waarde van rechtsbescherming wel een aantal eisen. Die eisen zijn voor een belangrijk deel gelijk aan de eisen die worden gesteld aan beslisondersteuningsalgoritmen. Ten eerste moeten de beslissingen uitlegbaar zijn. Een beslissing zonder verdere onderbouwing, of een beslissing waarbij alleen wordt verwezen naar de inzet van het onderliggende algoritme, kan in ieder geval niet als een voldoende motivering in de zin van de publieke waarde van rechtsbescherming en art. 6 EVRM gelden. Robotrechters die niet in staat zijn hun beslissing uit te leggen, bieden op grond van het geldend recht dan ook geen effectief rechtsmiddel en geen eerlijk proces. Mede om deze reden moet worden aangenomen dat de toepassing van algoritmen die zelfstandig rechterlijke uitspraken doen alleen mogelijk is als rechtzoekenden toegang hebben tot de gronden waarop het rechterlijk oordeel rust. Deze voorwaarde volgt ook uit art. 22 AVG.[88] In deze bepaling is een uitgangspunt vastgelegd dat beslissingen in beginsel niet volledig geautomatiseerd genomen mogen worden. Voor zover volledig geautomatiseerde beslissingen in uitzonderlijke gevallen toch mogelijk worden gemaakt, moet steeds in passende waarborgen worden voorzien.[89] Zo zal de betrokkene zijn

[88] Zie voor dit verbod in strafrechtelijke zaken ook art. 7e Wjsg. Zie voor het verbod op geautomatiseerde besluitvorming in algemene zin: Groep gegevensbescherming artikel 29 2017b; Zwenne, in: *T&C Privacy- en telecommunicatierecht* 2018.

[89] De uitwerking daarvan is te vinden in art. 40 UAVG. Dit artikel geldt voor geautomatiseerde besluitvorming anders dan profilering, voor zover dat noodzakelijk is ten behoeve van het voldoen aan een wettelijke verplichting die op de verwerkingsverantwoordelijke rust of noodzakelijk is voor de vervulling van een taak van algemeen belang. Het gaat daarbij om de uitoefening van gebonden bevoegdheden waarbij geen of weinig discretionaire ruimte is, waardoor de

standpunt kenbaar moeten kunnen maken en moet hij, om het besluit effectief te kunnen aanvechten, uitleg kunnen krijgen over het genomen besluit.[90] Bij regelgebaseerde algoritmen die de vorm van een beslisboom aannemen kan in beginsel veelal in deze uitleg worden voorzien door de onderliggende keuzes zichtbaar te maken. Voor zelflerende algoritmen ligt dat anders. Dergelijke algoritmen werken immers op basis van statistische verbanden en dus niet op basis van causale of juridisch redengevende verbanden. Voor zover het al mogelijk is om die verbanden en de wijze waarop die verbanden worden gelegd, te achterhalen, is het maar zeer de vraag of een overzicht van de door het algoritme in de data gelegde verbanden kan gelden als een dragende motivering. Rechters worden immers geacht niet naar statistische correlaties te zoeken, maar juist naar redengevende verbanden te verwijzen.[91] Dat betekent dat een motivering inhoudelijke argumenten bevat die daadwerkelijk tot de beslissing hebben geleid en die beslissing ook kunnen dragen. Dat is wezenlijk anders dan een statistisch verband.

In sommige van de door ons gehouden interviews werd door respondenten aangegeven dat soms soepeler omgesprongen zou kunnen worden met de eisen die aan de motivering worden gesteld. Zij vonden, met andere woorden, dat een statistische onderbouwing soms wel als een voldoende motivering zou moeten kunnen gelden. Dat zou volgens hen bijvoorbeeld het geval kunnen zijn voor procedures waarin partijen het eens zijn over de gewenste uitkomst, of waarin partijen vooral behoefte hebben aan een snelle beslissing en de motivering van die beslissing voor hen veel minder waarde heeft. Deze geïnterviewden wierpen de vraag op of partijen geen afstand van hun recht op een motivering zouden moeten kunnen doen. Zou dit in de toekomst worden overwogen, dan moet overigens wel worden opgemerkt dat partijen zich ervan bewust moeten zijn dat hun keuze voor een robotrechter ook gevolgen kan hebben voor hun recht op hoger beroep, omdat dat alleen verwezenlijkt kan worden als voorzien is in een dragende motivering.

Een tweede eis die vanuit het perspectief van rechtsbescherming moet worden gesteld in verband het gebruik van algoritmen als robotrechters is dat de *equality of arms* wordt gerespecteerd. Die eis kan onder druk komen te staan doordat de procedure voor een robotrechter noodzakelijkerwijs digitaal gevoerd moet worden en er niet zonder meer vanuit kan worden gegaan dat iedereen voldoende digitaal geletterd is om effectief gebruik te kunnen maken van zo'n procedure.[92] Ook om die reden is het belangrijk om te (blijven) voorzien in een volwaardig alternatief voor een robotrechter. Ook moet worden voorkomen dat *repeat players* een voordeel kunnen hebben

betekenis hiervan voor de rechtspraak waarschijnlijk beperkt is. Zie ook Jak & Bastiaans, *NJB* 2018, p. 3018-3025. Art. 7e Wjsg lijkt iets meer ruimte te creëren, door te bepalen dat geautomatiseerde besluitvorming verboden is, tenzij wordt voorzien in voorafgaande menselijke tussenkomst door of namens de verwerkingsverantwoordelijke en in specifieke voorlichting aan de betrokkene.

[90] Overweging 71 AVG. Uitwerking daarvan is onder meer te vinden in de informatierechten die zijn neergelegd in artt. 13 lid 2 onder f, 14 lid 2 onder g en 15 lid 1 onder h AVG.

[91] In deze zin Bex & Prakken, *AA* 2020, p. 255-259; AARvS 2018.

[92] Ter illustratie wordt hier gewezen op de problemen rond e-court: Gerards 2019, p. 168; Van Duin, *Sdu Blog* 5 februari 2018. Zie over het verloop van de procedure bij e-court het procesreglement (*Procesreglement e-court 2017,* 1 februari 2018, e-court.nl/juridisch); zie verder Spronken, *NJB* 2018, p. 791; Bauw, *AA* 2018, p. 890-893.

doordat zij beter in staat zijn de werking van het algoritme te begrijpen. Dat gevaar wordt groter naarmate dergelijke *repeat players* financiële middelen hebben om de uitkomsten van de procedures voor de robotrechter te analyseren en op basis daarvan hun processtrategie aan te passen. Van belang is dan ook dat bij de inzet van algoritmen als robotrechter oog wordt gehouden voor de (machts)verhoudingen tussen procespartijen.

Een volgend punt dat ten aanzien van de potentiële inzet van robotrechters uit de interviews naar voren kwam, is dat het systeem van rechtspraak voldoende flexibel dient te blijven en ruimte moet blijven bieden aan rechtsontwikkeling door rechters. Zou een robotrechter leiden tot verstarring, dan zou daarmee een belangrijke rechterlijke functie verloren gaan, namelijk het bieden van effectieve rechtsbescherming in het licht van de omstandigheden van een individueel geval, ook als die bescherming een verandering van de interpretatie van rechtsregels vereist. Met dit punt moet rekening worden gehouden bij de inzet van zowel regelgebaseerde als zelflerende algoritmen. Voor regelgebaseerde systemen geldt dat zij niet kunnen omgaan met omstandigheden die niet van tevoren zijn voorzien; ze zijn immers helemaal voorgeprogrammeerd. Zelflerende systemen kunnen zich wel aanpassen en ontwikkelen, maar zij zullen dat altijd doen op basis van de data die zij gedurende hun trainingsproces gepresenteerd hebben gekregen. Die data zijn altijd gebaseerd op het verleden. Door deze modellen toe te passen op nieuwe feitencomplexen, wordt de toekomst altijd in zekere zin door het verleden gekleurd, waarmee algoritmen kunnen leiden tot een *self fulfilling prophecy*. Ook dit pleit er vanuit de publieke waarde van rechtsbescherming voor om algoritmen vooralsnog alleen in te zetten als robotrechter (voor zover dat technisch al mogelijk is) in procedures die gaan over (zeer) eenvoudige juridische vragen. Gezien de stand van de techniek zal het bijna noodzakelijkerwijs moeten gaan om geschillen en procedures die geen diepgravende analyses van het recht of het feitencomplex behoeven, en waarbij minimale interpretatie van rechtsregels of minimale toespitsing op een individueel geval is vereist.[93] In dit soort procedures is immers minder behoefte aan rechtsvorming, maar vooral aan een (snelle) beslissing. Voor zover dat anders ligt, is het wel van belang dat er altijd hoger beroep openstaat bij een rechter, en dat betekent dat ook weer moet worden voldaan aan de hiervoor besproken eisen van voldoende uitlegbaarheid en motivering.

6.5.3 Tussenconclusie

De inzet van algoritmen in de rechtspraak wordt, net als in veel andere domeinen, gedreven door de wens (beslissings)processen efficiënter in te richten. De inzet van algoritmen biedt daarnaast in meer specifieke zin een aantal kansen voor de publieke waarde van rechtsbescherming. In het bijzonder kan de inzet van algoritmen op het niveau van de bedrijfsvoering leiden tot een grotere transparantie over het functioneren van de rechtspraak. Zo kunnen algoritmen eraan bijdragen dat

[93] In het civiele recht maakt dit soort procedures een groot deel van de werklast uit: Reiling 2009, p. 121-122 en 126; Reiling, *Computerrecht* 2020, p. 41.

meer uitspraken worden gepubliceerd, hetgeen partijen beter inzicht kan geven in hun proceskansen en dus de rechtszekerheid ten goede kan komen. Voor zover algoritmen worden ingezet ter ondersteuning van rechters kunnen zij bovendien leiden tot snellere en efficiëntere procedures. Als het lukt om dergelijke doelen te bereiken en als door het bereiken van die doelen personen die rechtstreeks bij de rechtspraak zijn betrokken meer tijd hebben om hun werkzaamheden te verrichten, kan de inzet van algoritmen een positief effect hebben op de verwezenlijking van de waarde van rechtsbescherming. Op basis van de nu beschikbare gegevens en ervaringen is echter moeilijk te voorspellen of, hoe en op welke wijze dit effect kan intreden.

Bezien in het licht van de publieke waarde van rechtsbescherming, is in het voorgaande ook een aantal belangrijke risico's geïdentificeerd. Een algemeen risico van de inzet van algoritmen in de rechtspraak is dat de rechtspraak of individuele rechters afhankelijk worden van algoritmeontwikkelaars en dat die ontwikkelaars bij het ontwerp van de algoritmen inhoudelijke beslissingen nemen die doorwerken in de rechtspraak. Die invloed kan de onafhankelijkheid van de rechtspraak en rechters onder druk zetten. Het is daarom van belang dat, als de rechtspraak de door haar in de toekomst te gebruiken algoritmen niet zelf ontwikkelt, zij altijd (product)eigenaar is van de gebruikte algoritmen en dat er inzicht en kennis bestaat over de werking daarvan. Tevens eist de onafhankelijkheid van de rechter en de rechterlijke macht dat keuzes die door een algoritme worden gemaakt altijd transparant en uitlegbaar zijn. Als niet aan die voorwaarden wordt voldaan, zijn algoritmen ongeschikt om in de rechtspraak gebruikt te worden.

Wanneer algoritmen gebruikt worden ter ondersteuning van de bedrijfsvoering, kunnen zij mogelijk leiden tot beter inzicht in het functioneren van de rechtspraak. Algoritmen bieden dan de mogelijkheid om efficiënter te sturen op doorlooptijden en productienormen. Het gevaar daarvan is evenwel dat die informatie ook tegen rechters kan worden gebruikt. Het lijkt dan ook waarschijnlijk dat het toepassen van algoritmen in de bedrijfsvoering explicieter de vraag op tafel zal leggen in hoeverre rechters kunnen worden afgerekend op hun prestaties.

Bij het gebruik van algoritmen in de ondersteuning van rechters is vooral van belang dat rechters begrijpen op welke wijze het algoritme tot een advies is gekomen. Rechters zijn namelijk onder alle situaties verantwoordelijk voor de beslissingen die zij nemen, ook als die door algoritmen worden ingegeven. Een andere vraag is of ook procespartijen toegang dienen te krijgen tot die informatie, zodat zij verweer kunnen voeren tegen de invloed die algoritmen op de beslissing door de rechter zouden kunnen hebben. Op dit punt lijkt het juridisch kader op dit moment geen eenduidig antwoord te bieden.

De grootste problemen met de inzet van algoritmen in de rechtspraak in het licht van de rechtsbescherming als publieke waarde doen zich voor wanneer algoritmen worden gebruikt om

zelfstandig geschillen te beslechten. Zo volgt uit de publieke waarde rechtsbescherming dat uitspraken op een draagkrachtige en inhoudelijke manier moeten worden gemotiveerd. Lang niet alle algoritmen zijn in staat in een dergelijke motivering te voorzien. Dat kan een gevolg zijn van het gebrek aan transparantie van het algoritme, maar ook van de omstandigheid dat algoritmen niet in staat zijn om relevante juridische redengevende argumenten te geven die aan een bepaalde uitkomst ten grondslag liggen. Zelflerende algoritmen, en de modellen op basis waarvan zij werken, reflecteren immers alleen statistische verbanden, die niet gelijk zijn te stellen aan juridische (redengevende) argumenten. Gewezen is daarnaast op de eis van *equality of arms* en op de noodzaak om ruimte te blijven bieden voor rechtsontwikkeling en interpretatie op basis van individuele gevallen. Aan dit soort eisen kan mogelijk worden voldaan als de inzet van algoritmen beperkt blijft tot zeer eenvoudige rechtsvragen en het functioneren van het algoritme voor eenieder inzichtelijk is, bijvoorbeeld omdat gebruikgemaakt wordt van een regelgebaseerd algoritme in plaats van zelflerende algoritmen, of als – zoals door sommige respondenten genoemd – voor sommige categorieën van zaken minder waarde wordt toegedicht aan motivering of rechtsontwikkeling.

6.6 Conclusie

In deze casestudy stond de (toekomstige) inzet van algoritmen in de rechtspraak centraal. De ontwikkelingen op dit terrein staan nog in de kinderschoenen. Gedeeltelijk is dit toe te schrijven aan het feit dat algoritmische systemen alleen goed kunnen functioneren in een gedigitaliseerde omgeving. Algoritmische systemen zijn voor hun werking immers afhankelijk van voor computers leesbare gegevens. Zolang die gegevens niet beschikbaar zijn en de procesvoering en dataverzameling niet zijn gedigitaliseerd, ontbreekt binnen de rechtspraak een vruchtbare bodem om algoritmen op grote schaal en op alle door ons beschreven wijzen te kunnen inzetten. Dit is alleen anders voor de inzet van algoritmen waarbij de rechterlijke macht gebruik kan maken van de gegevens die zij zelf creëert. Het gaat daarbij dus met name om interne processen. Waar de rechtspraak gebruikmaakt van externe gegevens afkomstig van partijen ligt dit ingewikkelder omdat de rechterlijke procedure in Nederland tot nu toe slechts in (zeer) beperkte mate is gerealiseerd. Illustratief in dit kader is dat het procesrecht in bijna alle domeinen van de rechtspraak het nog toestaat om handgeschreven stukken aan te leveren of processtukken per fax naar de rechterlijke instantie te sturen.

Ook de technische mogelijkheden die algoritmen op dit moment bieden, leveren beperkingen op voor het gebruik van algoritmen in de rechtspraak. Voor zover de rechtspraak in de nabije toekomst – de komende vijf tot tien jaar – tot de inzet van algoritmen zou beslissen, lijkt het aannemelijk dat iteratieve taken die in duidelijke regelstructuren te vatten zijn en dus door transparante regelgebaseerde algoritmen kunnen worden uitgevoerd, als eerst in aanmerking komen om te

worden geautomatiseerd. Dit geldt bijvoorbeeld voor de inzet op het niveau van de bedrijfsvoering, waar relatief eenvoudige procesmatige beslissingen worden genomen die samenhangen met doorlooptijden en aantallen procedures.

Voor de taken die door rechters worden uitgevoerd is dit anders. Veel van het rechterlijke werk, en in het bijzonder het inhoudelijk beslissen van een zaak, kan niet door een algoritme worden vervangen, eenvoudigweg omdat algoritmen daartoe technisch (nog) niet in staat zijn. Tegelijkertijd zijn er, ook met inachtneming van de kanttekeningen die hiervoor zijn gemaakt, in de wijze waarop de rechtspraak nu functioneert, rechterlijke procedures en handelingen aan te wijzen die zich in de toekomst mogelijk zouden kunnen lenen voor automatisering met behulp van algoritmen. Daarbij kan onder meer gedacht worden aan rechterlijke beslissingsprocessen die ook nu al voor een (groot) deel worden bepaald door vooraf vastgestelde richtlijnen. Zo maken strafrechters ook nu al gebruik van oriëntatiepunten die helpen bij het bepalen van de strafmaat. Het is voorstelbaar dat het berekenen van de strafmaat aan de hand van door een rechter geselecteerde variabelen in de toekomst aan een algoritme wordt overgelaten, mits dit algoritme voldoet aan de voorwaarden zoals ze in deze casestudy zijn besproken.

In deze casestudy is gewezen op een aantal kansen ten aanzien van de inzet van algoritmen in de rechtspraak. De bestudeerde algoritmische toepassingen dragen bijvoorbeeld de kans in zich dat de rechtseenheid en daarmee samenhangende rechtsgelijkheid erdoor worden vergroot. Dat is bijvoorbeeld het geval als de inzet van algoritmen in de bedrijfsvoering ertoe leidt dat procedures die inhoudelijk met elkaar samenhangen door dezelfde rechters worden behandeld, als rechters worden ondersteund door algoritmen bij het zoeken van bijvoorbeeld jurisprudentie en daardoor in gelijksoortige zaken vaker van dezelfde jurisprudentie gebruikmaken, of als algoritmen worden gebruikt om procedures op gelijkvormige wijze rechtstreeks te beslissen. De inzet van algoritmen biedt verder mogelijkheden om de publieke waarde van rechtsbescherming te realiseren. Ten eerste kan de inzet van algoritmen op het niveau van de bedrijfsvoering leiden tot een grotere transparantie over het functioneren van de rechtspraak. Deze transparantie leidt er niet alleen toe dat de rechtspraak op efficiëntere wijze kan worden georganiseerd en verantwoordelijk kan worden gehouden voor haar prestaties, maar een grotere transparantie kan ook betekenen dat rechtszoekenden een betere inschatting van hun rechtspositie kunnen maken. Voor zover algoritmen worden ingezet ter ondersteuning van rechters, of misschien zelfs als robotrechter dragen zij vooral de belofte in zich te leiden tot snellere en efficiëntere procedures.

Hoewel er daarmee zeker kansen zijn voor de inzet voor algoritmen in de rechterlijke macht, vooral als de daarvoor noodzakelijke digitale infrastructuur is gerealiseerd, is in het voorgaande ook een aantal risico's in beeld gebracht in verband met de publieke waarden die in dit onderzoek centraal staan (gegevensbescherming, non-discriminatie en rechtsbescherming). Uitgelegd is dat de mate

waarin die risico's zich kunnen verwezenlijken afhankelijk is van de precieze inzet van het algoritme (in de bedrijfsvoering, als beslissingsondersteuning of als robotrechter) en de aard van het gebruikte algoritme (regelgebaseerd of zelflerend). Hierna vatten wij de geïdentificeerde risico's per publieke waarde kort samen.

Gegevensbescherming

Bij het ontbreken van specifieke algoritmen waarvan het functioneren kan worden beoordeeld, zijn de risico's voor de waarde van gegevensbescherming die op dit moment het meest duidelijk zijn te identificeren vooral van institutionele en organisatorische aard. Het toezicht op de bescherming van persoonsgegevens binnen de rechtspraak is versplinterd. Daardoor bestaat er ten eerste een gevaar dat, wanneer tot de ontwikkeling en de inzet van algoritmen wordt besloten, niet altijd voldoende helder is welke instantie toezicht houdt op de verwerking van persoonsgegevens door algoritmen. Dat gevaar lijkt het grootst voor zover algoritmen toegepast gaan worden in de bedrijfsvoering. Ten tweede kan het gevolg van de versplintering zijn dat toezichtregimes uit elkaar gaan lopen doordat de verschillende toezichthouders het geldende wettelijk kader op verschillende wijzen gaan toepassen.

Non-discriminatie

Risico's voor de publieke waarde van non-discriminatie bestaan vooral wanneer algoritmen worden gebruikt bij de rechterlijke oordeelsvorming. Dit geldt zowel voor beslisondersteuningsalgoritmen als voor de zogenaamde robotrechters. In algemene zin bestaat het risico dat ongeoorloofde vooroordelen op structurele wijze een rol gaan spelen in de rechterlijke besluitvorming. Bij het ontwerp van de algoritmen die in de rechtspraak kunnen worden ingezet, zal de publieke waarde van non-discriminatie altijd in acht moeten worden genomen. In dat kader dient gebruik gemaakt te worden van representatieve, objectieve en neutrale gegevens. Daarnaast mogen ook in de besliscriteria van regelgebaseerde algoritmen geen ongeoorloofde vooroordelen doorklinken. Daarbij is van belang dat voldoende duidelijk is hoe algoritmen tot een besluit komen en dat gebruikers zich daarvan voldoende bewust zijn. Dat geldt zeker voor de inzet van zelflerende algoritmen. Bij de inzet van dat soort algoritmen bestaat namelijk het risico dat ook als bij het ontwerp neutrale uitgangspunten zijn gehanteerd, het algoritme eigenstandig verbanden legt die in juridische zin als indirecte discriminatie kwalificeren.

Rechtsbescherming

De inzet van algoritmen in de rechtspraak brengt ook voor de publieke waarde van rechtsbescherming een aantal risico's en voorwaarden met zich. Een algemeen risico is dat de onafhankelijkheid van de rechterlijke macht onder druk kan komen te staan. Dat risico doet zich vooral voor als aan de rechtspraak externe partijen betrokken zijn bij het ontwerp en de ontwikkeling van algoritmen voor de rechtspraak.

De waarde van rechtsbescherming brengt, evenals de waarde van gelijke behandeling, verder mee dat het altijd duidelijk moet zijn hoe een algoritme tot een uitkomst is gekomen. Dat geldt zeker als algoritmen rechters ondersteunen. De rechterlijke onafhankelijkheid en de daarmee verbonden professionaliteit eisen dat de rechter verantwoordelijkheid draagt en kan dragen voor alle beslissingen die hij neemt. Daaruit vloeit ook voort dat rechters een voldoende kritische houding moeten aannemen ten opzichte van de door hen gebruikte systemen. Als procespartijen met een algoritme worden geconfronteerd – direct omdat zij gebruikmaken van een robotrechter of indirect omdat een rechter wordt ondersteund door een algoritme – is het ook van belang dat zij weten hoe het algoritme functioneert. Dat stelt procespartijen in staat verweer te voeren over de wijze waarop het algoritme een procedure beïnvloedt.

De spanning met de publieke waarde van rechtsbescherming lijkt het grootst wanneer algoritmen worden gebruikt om procedures geautomatiseerd af te doen. Er kunnen dan problemen ontstaan met wezenskenmerken van een eerlijk proces. Zolang algoritmen niet in staat zijn om de juridisch relevante argumenten die tot een beslissing hebben geleid aan alle betrokkenen te presenteren, is de inzet van robotrechters onverenigbaar met de waarde van rechtsbescherming. Ook kan bij procedures die geautomatiseerd worden afgedaan de *equality of arms* onder druk komen te staan. Dat gevaar doet zich vooral voor wanneer *repeat players* in staat zijn de werking van het algoritme te doorgronden en daaruit voordeel te behalen.

Bestendigheid van het juridisch kader

Gezien het feit dat de toepassing van algoritmen in de rechtspraak nog in de kinderschoenen staat is het lastig om een volwaardige beoordeling te maken van de mate waarin het juridisch kader in staat is om de in dit onderzoek besproken publieke waarden te realiseren. Wel kan op een aantal aandachtspunten gewezen worden. Als uitgangspunt daarbij geldt dat naar onze inschatting het juridisch kader in algemene zin voldoende ruimte biedt om de ontwikkeling van algoritmen in de rechtspraak te faciliteren en te accommoderen op een wijze die recht doet aan de door ons besproken publieke waarden, mits daarbij de hierboven besproken aandachtspunten in acht worden genomen. Daarnaast is in het juridisch kader een aantal belemmeringen aan te wijzen als gevolg waarvan de kansen die de inzet van algoritmen in de rechtspraak zou kunnen bieden, niet volledig verwezenlijkt kunnen worden.

Daarbij gaat het onder meer om de vormvrijheid die wordt gehanteerd ten aanzien van het aanleveren van processtukken. Voor de ontwikkeling van sommige algoritmische systemen, dat geldt in het bijzonder voor de toepassing van algoritmen als beslisondersteuner en als robotrechter, is het nodig dat ook documenten die niet door de rechtspraak zijn geproduceerd in een voor een computer leesbare vorm beschikbaar zijn. In die gevallen zou de inzet van algoritmen

kunnen worden vereenvoudigd als procedures digitaal worden gevoerd. Voor het gebruik van algoritmen op het niveau van de bedrijfsvoering doen deze problemen zich niet voor. Op dat niveau kan worden volstaan met het gebruik van interne, door de rechtspraak geproduceerde data die ook nu al vaak digitaal beschikbaar is.

Daarnaast volgt ook uit de AVG een tweetal beperkingen. Zo is in art. 22 AVG een verbod op volledig geautomatiseerde besluitvorming neergelegd, voor zover die besluiten rechtsgevolgen of anderszins aanmerkelijke gevolgen hebben. Voor de inzet van robotrechters zal dan ook een wettelijke grondslag gecreëerd moeten worden, waarbij voorzien wordt in de nodige waarborgen. Een dergelijke (algemene) grondslag bestaat op dit moment in art. 40 UAVG voor geautomatiseerde besluitvorming die geen profilering inhoudt, mits dit noodzakelijk is om te voldoen aan een wettelijke verplichting of voor de vervulling van een taak van algemeen belang. Omdat het daarbij met name lijkt te gaan om de uitoefening van gebonden bevoegdheden waarbij weinig discretionaire ruimte bestaat, kan deze wettelijke bepaling waarschijnlijk niet voor alle vormen van algoritmische besluitvorming in de rechtspraak uitkomst bieden. Een laatste beperking die nog genoemd dient te worden volgt uit art. 9 AVG. Dit artikel bepaalt dat alleen in uitzonderlijke gevallen bijzondere categorieën persoonsgegevens zoals geslacht of etnische afkomst verwerkt kunnen worden. De inzet van algoritmen kan evenwel de paradoxale consequentie met zich brengen dat het nodig kan zijn om dergelijke persoonsgegevens te verwerken, juist om ongerechtvaardigd onderscheid op grond van dit soort kenmerken te voorkomen

Hoofdstuk 7. Casestudy Overheidsincasso bij verkeersboetes

Martje Boekema & Susanne Heeger

7.1 Introductie

In 2017 verscheen het onderzoeksrapport *Weten is nog geen doen: een realistisch perspectief op zelfredzaamheid* van de Wetenschappelijke Raad voor het Regeringsbeleid (WRR).[1] In dit onderzoeksrapport wijst de WRR erop, dat niet alle burgers in gelijke mate zelfredzaam zijn. De huidige samenleving doet een groot beroep op niet alleen de cognitieve vermogens van een individu, maar ook op diens niet-cognitieve vermogens (doenvermogen). De eisen aan burgers gaan regelmatig hun doenvermogen te boven. De WRR adviseert onder andere om bij de handhaving van overheidsbeleid te differentiëren en voor de groep 'niet-kunners' om actief, persoonlijk contact op te nemen. In de kabinetsreactie op dit rapport[2] is aangegeven, dat het kabinet de mogelijkheden zal verkennen om binnen massale besluitvormingsprocessen algoritmen in te zetten voor het signaleren van individuele zaken van burgers die door omstandigheden (tijdelijk) mogelijk niet zelfredzaam zijn, met als doel voor deze zaken een maatwerkaanpak in te zetten.

De onderzoeksopdracht voor onderhavig onderzoek betrof (onder meer) het signaleren van juridische kansen en risico's op drie publieke waarden: rechtsbescherming, non-discriminatie en de bescherming van persoonsgegevens. Voor deze casestudy heeft de onderzoeksopdracht geleid tot het volgende. De casestudy belicht de pilot 'Telefonisch Innen', die bij het Centraal Justititieel Incassobureau (CJIB) is uitgevoerd. Deze pilot illustreert op een aantal punten de juridische kansen en risico's die zich potentieel kunnen voordoen bij het middels een algoritme selecteren van burgers voor een maatwerkaanpak.[3] Daarnaast zullen, op basis van een literatuuranalyse, enkele samenhangende juridische kansen en risico's worden verkend die zich kunnen voordoen bij het uitvoeren van de beleidsaanbevelingen van het WRR-rapport.

De opbouw van deze casestudy is als volgt. Hierna worden in paragrafen 7.1.1 en 7.1.2 allereerst het WRR-rapport en de kabinetsreactie daarop geschetst. Vervolgens zal in paragraaf 7.1.3 verdere verantwoording worden gegeven voor de keuze van de casestudy en de methodologie. De tweede paragraaf (7.2) beschrijft de pilot Telefonisch Innen en het algoritme in deze pilot. De derde paragraaf (7.3) vertaalt de publieke waarde rechtsbescherming voor deze casestudy naar een juridisch kader, aanvullend op paragraaf 3.3 van dit onderzoek. Vervolgens worden juridische kansen en risico's beschreven die zich kunnen voordoen, zowel voor de pilot als voor een aantal

[1] WRR 2017.
[2] *Kamerstukken II* 2017/18, 34775 VI, nr. 88.
[3] De casestudy richt zich dus niet in algemene zin op het innen en incasseren van vorderingen door de overheid.

van de beleidsaanbevelingen uit het WRR-rapport. De vierde en vijfde paragraaf (7.4 en 7.5) zijn op dezelfde wijze opgebouwd voor de waarden non-discriminatie en de bescherming van persoonsgegevens.[4] Afsluitend volgt in paragraaf 7.6 een conclusie.

7.1.1 WRR-rapport 'Weten is nog geen doen' (2017)

De WRR heeft de afgelopen jaren een reeks rapporten uitgebracht waarin inzichten uit sociaalwetenschappelijk onderzoek worden omgezet in beleidsaanbevelingen. In eerdere rapporten richtte de WRR zich op de cognitieve beperkingen van individuen bij het maken van keuzes: de manier waarop het menselijk denken en de oordeelsvorming zijn begrensd. In het rapport 'Weten is nog geen doen' (2017) richt de WRR zich juist op niet-cognitieve persoonskenmerken. Het rapport biedt een overzicht van persoonskenmerken die ertoe leiden dat gemaakte keuzes niet altijd worden omgezet in gedrag.

De WRR verdeelt deze persoonskenmerken in vier factoren. Het gaat ten eerste om de mate waarin iemand een *avoidance temperament* of juist een *approach temperament* heeft (gaat men problemen uit de weg of worden deze juist aangepakt?).[5] Daarnaast onderscheiden de onderzoekers de capaciteit voor zelfcontrole van een individu; dit kenmerk typeren zij als een neurobiologisch verankerd vermogen. De derde factor is optimisme, gedefinieerd als de mate waarin een individu optimistisch denkt over zichzelf en over zijn omgeving.[6] Ten slotte is er de waargenomen controle: de overtuiging die een individu heeft over zichzelf, over de wereld en over de relatie tussen beide.[7] Deze vier persoonskenmerken zijn niet-cognitieve vermogens: het gaat niet om het denkvermogen van een individu maar om andere persoonskenmerken die zijn of haar gedrag beïnvloeden. De onderzoekers nemen deze persoonskenmerken samen onder de noemer 'doenvermogen'. Op basis van een literatuuronderzoek, gecombineerd met eigen kwantitatief onderzoek, concluderen de onderzoekers in het WRR-rapport dat het doenvermogen van een individu van invloed is op diens zelfredzaamheid. Onder zelfredzaamheid wordt in het rapport verstaan: de mate waarin een burger beschikt over vermogens om zijn doelen te bereiken en zich te redden in het leven. Burgers met verminderd doenvermogen zijn hier onvoldoende toe in staat.[8]

In het WRR-rapport wordt geconcludeerd dat burgers verschillen in de mate waarin zij beschikken over doenvermogen. Zowel laag- als hoogopgeleiden kunnen moeite hebben om hun zaken goed

[4] Anders dan in de andere casestudy's, wordt in deze casestudy rechtsbescherming als eerste besproken, daarna non-discriminatie en de bescherming van persoonsgegevens. De reden hiervoor is dat de latere paragrafen inhoudelijk voortbouwen op de paragraaf over rechtsbescherming. In deze paragraaf wordt beschreven op welke manier een burger kan opkomen tegen onwelgevallig overheidsoptreden. Bij de pilot Telfonisch Innen kan dan bijvoorbeeld worden gedacht aan een klacht bij de nationale ombudsman of bij de Autoriteit Persoonsgegevens (AP). In de latere paragrafen worden twee publieke waarden belicht waaraan in een dergelijke procedure wordt getoetst: het non-discriminatiebeginsel en de bescherming van persoonsgegevens.
[5] WRR 2017, p. 69.
[6] WRR 2017, p. 70.
[7] WRR 2017, p. 72.
[8] WRR 2017, p. 22.

te regelen. Ook is het zo dat burgers die normaal gesproken goed in staat zijn om zichzelf te redden, door stress en mentale belasting verminderd zelfredzaam zijn. De onderzoekers noemen baanverlies, echtscheiding en problematische schulden als mogelijke oorzaken hiervan. Juist als het leven flink tegenzit kan het doenvermogen van burgers volgens de onderzoekers onder druk komen te staan.[9]

De WRR ziet twee oplossingsrichtingen voor dit probleem. Als de overheid nieuw beleid en nieuwe regelgeving opstelt, zou dit gericht moeten zijn op het verminderen van verleidingen en stress voor burgers. Keuzevrijheid voor burgers zou bijvoorbeeld moeten wijken voor meer overheidssturing als burgers beslissingen moeten nemen die grote impact hebben op hun verdere leven. Als de overheid bestaand beleid uitvoert, pleiten de onderzoekers voor een persoonlijke en proportionele aanpak. De WRR beveelt in het rapport aan om vaker maatwerk toe te passen bij handhavend overheidsoptreden. Bij onregelmatigheden zou voorafgaand aan handhaving actief en persoonlijk contact met de burger moeten worden opgenomen. Op deze wijze kan in de uitvoering worden gedifferentieerd tussen burgers die niet *willen* handelen en burgers die dat niet *kunnen*.[10]

De WRR werkt deze oplossingsrichtingen op een aantal terreinen uit in beleidsaanbevelingen. Als het gaat om persoonlijke financiën concludeert het rapport dat wanneer een burger boetes niet voldoet, of niet reageert, een overheidsinstantie bij voorkeur eerst persoonlijk contact met de burger op zou moeten nemen. Dwanginvordering (bijvoorbeeld het inzetten van een deurwaarder) zou pas aan de orde moeten komen als andere mogelijkheden (bijvoorbeeld het treffen van een betalingsregeling) zijn gepasseerd.[11] Bij voorkeur worden contactmomenten ingebouwd voordat er een opeenstapeling van boetes kan ontstaan. De aanname van onderzoekers is dat door vroegtijdig contact op te nemen problematische schulden en bijkomende stress worden vermeden. Door bij te sturen op een moment dat mensen nog over voldoende doenvermogen beschikken, zijn burgers beter in staat om op eigen kracht uit de problemen te blijven.[12]

7.1.2 Kabinetsreactie op het WRR-rapport en probleemstelling

Het kabinet heeft in zijn reactie op het WRR-rapport ‘Weten is nog geen doen’ aangegeven op welke manier de overheid de inzichten uit het rapport wil gebruiken. De kabinetsreactie is opgedeeld in drie delen. De eerste twee delen beschrijven de implicaties van het WRR-rapport voor de inhoud van nieuw beleid en nieuwe regelgeving. In het derde deel van de kabinetsreactie (‘Uitvoering van beleid: verifiëren en differentiëren’) onderschrijft het kabinet het belang van vroegtijdig contact met de burger die problemen ondervindt bij de uitvoering van overheidsbeleid.[13]

[9] WRR 2017, p. 130.
[10] WRR 2017, p. 132.
[11] WRR 2017, p. 135.
[12] WRR 2017, p. 147.
[13] *Kamerstukken II* 2017/18, 34 775 VI, nr. 88, p. 7.

Het kabinet heeft in dit onderdeel van de reactie een toezegging gedaan aan de Tweede Kamer dat, als onderdeel van het voorliggend onderzoek, aandacht zal worden besteed aan de vraag 'welke kansen kunstmatige intelligentie kan bieden voor een tijdige signalering, vooral in massale besluitvormingsprocessen, van mensen die door omstandigheden (tijdelijk) mogelijk niet zelfredzaam zijn'.[14] Uit de context van de toezegging in het derde deel van de kabinetsreactie leiden wij af, dat de toezegging ziet op de inzet van algoritmen bij de uitvoering van bestaand beleid en bestaande regelgeving. De kabinetsreactie maakt melding van de term 'kunstmatige intelligentie'. Deze term wordt in de kabinetsreactie zelf niet gedefinieerd. Wij hebben in deze casestudy de definitie aangehouden uit par. 2.1.2 van dit onderzoek. Net als in die paragraaf gebruiken we de term 'zelflerende algoritmen' als synoniem voor kunstmatige intelligentie (AI).

De centrale vraag van onderhavig onderzoek, zoals geformuleerd en nader uitgewerkt in hoofdstuk 1, luidt:

Welke kansen en risico's doen zich voor bij algoritmische besluitvorming met betrekking tot de bescherming en realisering van publieke waarden en belangen, en zijn de bestaande juridische kaders voldoende bestendig om kansen te verwezenlijken en het intreden van geïdentificeerde risico's te voorkomen of de gevolgen daarvan te mitigeren?

Deze casestudy beperkt zich daarbij tot de drie algemene publieke waarden die in hoofdstuk 1 zijn geformuleerd: rechtsbescherming, non-discriminatie en de bescherming van persoonsgegevens.

Uit bovenstaande vraagstelling volgt dat potentiële kansen en risico's worden geschetst, waarbij de gekozen casestudy wordt gebruikt om deze kansen en risico's te illustreren. De pilot Telefonisch Innen zal daarbij niet worden geëvalueerd. We kunnen dus bijvoorbeeld geen uitspraken doen over of de doelen van de pilot zijn behaald. Daarom worden de beleidsafwegingen van de pilot ook alleen besproken voor zover deze vallen binnen de vraagstelling. Ook zullen we geen juridisch oordeel geven over de rechtmatigheid van de pilot zoals deze de afgelopen jaren is uitgevoerd door het CJIB.

De verkenning van kansen en risico's op de drie publieke waarden resulteert in deze casestudy, voor zover mogelijk, in een bespreking van de bestendigheid van het juridisch kader. Deze bespreking was niet altijd mogelijk, omdat een casestudy het slechts beperkt mogelijk maakt om generaliserende conclusies te trekken. Zie voor deze beperking van de gekozen onderzoekopzet verder par. 1.3.1.

[14] *Kamerstukken II* 2017/18, 34 775 VI, nr. 88.

7.1.3 Verantwoording keuze casestudy

Uit de onderzoekopdracht, zoals hierboven en in hoofdstuk 1 beschreven, volgen vier voorwaarden waarmee rekening moet worden gehouden bij de keuze voor de casestudy. Ten eerste moet de casestudy plaatsvinden in de uitvoeringsfase van overheidsbeleid. Het formuleren van nieuw beleid of nieuwe regelgeving valt hier dus niet onder. Ten tweede moet er sprake zijn van zelflerende algoritmen die burgers identificeren die in aanmerking komen voor een maatwerkaanpak. Ten derde moet de casestudy overheidshandelen inhouden dat grote aantallen burgers raakt ('massale besluitvormingsprocedures'). Ten slotte moet het algoritme erop gericht zijn om – sneller dan zonder een algoritme mogelijk zou zijn – burgers te selecteren die (tijdelijk) niet zelfredzaam zijn.[15] Hieronder worden de implicaties van deze vier criteria voor de gekozen casestudy besproken.

Vereist is ten eerste, dat de casestudy plaatsvindt in de uitvoeringsfase van overheidsbeleid. Gekozen is voor het pilotproject Telefonisch Innen bij het CJIB. Het CJIB is een uitvoeringsinstantie van het Rijk. De pilot betreft een veranderde aanpak in het innen van verkeersboetes. Daarmee vindt de pilot plaats in de uitvoeringsfase van overheidsbeleid.

Ten tweede is vereist, dat het algoritme dat binnen de pilot wordt gebruikt, kan worden getypeerd als zelflerend. De pilot bevat een algoritme dat als zodanig kan worden getypeerd, zie voor de beschrijving par. 7.2.2. De derde voorwaarde is, dat er bij de gekozen casestudy sprake is van massale besluitvormingsprocedures. In 2019 was het CJIB verantwoordelijk voor de afhandeling van ruim acht miljoen boetes op grond van de Wet administratiefrechtelijke handhaving verkeersvoorschriften (Wahv).[16] Bij het innen van deze vorderingen zet het CJIB diverse instrumenten in die zijn gericht op het ondersteunen van de debiteur in het voldoen aan diens betalingsverplichtingen. De pilot is een van deze instrumenten, waarbij een specifieke doelgroep proactief persoonlijk wordt benaderd. Binnen de pilot is er gemiddeld 2.000 keer per maand persoonlijk contact geweest met burgers met een openstaande Wahv-vordering.[17] Daarmee kan worden gesproken van massale besluitvormingsprocessen, waarbinnen algoritmen een rol kunnen spelen om zaken te selecteren waarvoor een maatwerkaanpak geschikt kan zijn.

De vierde voorwaarde die hierboven is beschreven, is dat het algoritme erop is gericht om – sneller dan zonder een algoritme mogelijk zou zijn – burgers te selecteren die (tijdelijk) niet zelfredzaam zijn. We nemen aan dat het gaat om onvoldoende zelfredzaamheid als gevolg van verminderd doenvermogen. Deze vierde eis is problematisch gebleken. De kabinetsreactie bevat geen nadere

[15] *Kamerstukken II* 2017/18, 34 775 VI, nr. 88, p. 9.
[16] 'Feiten & Cijfers 2019', cjb.nl.
[17] Burgers waardeerden dit persoonlijke contact doorgaans zeer (interview CJIB 9-10-2019). Tot op heden is er één klacht ontvangen bij het CJIB over het telefonisch benaderen van burgers (Ministerie van JenV, Uitvoeringstoets 2019).

invulling van de begrippen ‘zelfredzaamheid’ en ‘doenvermogen’ op dit punt. Het WRR-rapport definieert het begrip ‘zelfredzaamheid’ als de mate waarin individuen in staat om hun doelen te bereiken en zich te redden in het leven.[18] In het WRR-rapport is dit uitgewerkt op basis van een bestaande wetenschappelijke vragenlijst. Respondenten is gevraagd om zichzelf te scoren op 21 vaardigheden, zoals ‘toekomstige ontwikkelingen inschatten’ en ‘duidelijk formuleren wat ik wil bereiken’. Op basis van hun antwoorden hebben de respondenten een individuele score toegewezen gekregen voor zelfredzaamheid. Het begrip ‘doenvermogen’ is op eenzelfde manier gemeten. [19]

Deze definitie is naar onze mening onvoldoende afgebakend voor het verder invullen van het begrip ‘zelfredzaamheid’ voor de keuze van een concreet beleidsvoorbeeld als casestudy. De door de WRR in het rapport gegeven beleidsvoorbeelden boden wel handvatten voor nadere afbakening en daarmee voor de keuze van de casestudy. In het WRR-rapport wordt onder andere het kwadrantenmodel van het ministerie van Justitie en Veiligheid beschreven. Dit model geeft richtlijnen voor het innen van geldschulden door de overheid.[20]

Kern van dit kwadrantenmodel, zoals weergegeven in het WRR-rapport, is dat burgers in vier kwadranten worden verdeeld, op basis van het wel of niet willen en het wel of niet kunnen betalen van een vordering. Voor ‘niet-kunners’, die wel voldoende geld hebben maar die moeite hebben met de formulieren of met het plannen, zou de overheid hulp moeten bieden met de registratie en het betalen van de boete.[21] Burgers moeten zonodig actief en persoonlijk worden benaderd.[22]

Bovenstaande beleidsdoelen sluiten naar ons oordeel aan bij een van de doelen van de gekozen casestudy Telefonisch Innen. Het CJIB past in deze pilot een algoritme toe om zaken te selecteren die mogelijk in aanmerking komen voor een maatwerkaanpak, bestaande uit een telefonische betalingsherinnering. De zaken in de pilot betreffen niet-betaalde verkeersboetes. Het CJIB beoogt in de pilot om debiteuren te identificeren die wel voldoende geld hebben om de boete (op termijn) te betalen, maar die gedrag vertonen dat kan wijzen op verminderde zelfredzaamheid.[23] Niet elke burger is immers in staat om zijn of haar openstaande rekeningen te overzien en daarnaast te begrijpen wat voor elke rekening de consequenties zijn van niet-betalen. Ook is niet elke burger in staat om actie te ondernemen als betaling binnen de standaardtermijn niet mogelijk is.[24] De pilot is daarmee wat ons betreft een casestudy die aansluit bij de wens in de kabinetsreactie om middels

[18] WRR 2017, p. 22.
[19] WRR 2017, p. 74-78.
[20] ‘Innen en casseren’, cjib.nl. Bij de opzet van de pilot Telefonisch Innen heeft het CJIB willen aansluiten bij het hierboven beschreven kwadrantenmodel (WODC 2017; Live brengen analytics 2017, p. 12).
[21] WRR 2017, p. 146-148
[22] WRR 2017, p. 134-135.
[23] *Visie tenuitvoerlegging financiële sancties* 2015, p. 6; interview ministerie van JenV, 5-12-2019; Mulder, Klingenberg en Mifsud-Bonnici 2017, p. 6.
[24] Rapport Decisio & DSP 2019, p. 16-18; Hertogh, Winter & Schudde 2012, p. 57-59.

algoritmen burgers te selecteren die (tijdelijk) niet zelfredzaam zijn door verminderd doenvermogen.[25]

7.1.4 Methodologie

Ten behoeve van deze casestudy is literatuuronderzoek uitgevoerd, aangevuld met semigestructureerde interviews (zie voor de verdere beschrijving van de onderzoeksopzet hoofdstuk 1). De opzet en uitvoering van de pilot Telefonisch Innen en de inzet van het algoritme hierbinnen is nagegaan bij het CJIB middels twee interviews: één interview met de verantwoordelijke projectleider en één interview met een data-analist die het gebruikte algoritme mede heeft ontwikkeld. Naar aanleiding van deze twee interviews zijn interne stukken opgevraagd. De opzet en uitvoering van het project bij het CJIB vindt plaats in een pilot, en de opgevraagde stukken waren daardoor ten tijde van het onderzoek (deels) niet openbaar.[26]

De verkenning van kansen en risico's voor de geselecteerde publieke waarden, alsmede de verhouding tot het wettelijk kader, is beschreven met behulp van vier interviews met personen die werkzaam zijn bij relevante onderdelen van de overheid. Allereerst is gesproken met drie medewerkers van het ministerie van Sociale Zaken en Werkgelegenheid, die werken aan beleidsprojecten gericht op het voorkomen van problematische schulden bij burgers. Daarnaast is gesproken met drie medewerkers van de Vereniging Nederlandse Gemeenten (VNG), die kennis hebben van de inzet van ICT bij de dienstverlening van de overheid aan burgers. Met hen is ook gesproken over huidig en toekomstig beleid dat door gemeenten wordt uitgevoerd voor het voorkomen van het oplopen van schulden bij burgers. Verder is gesproken met een IT-jurist en een senior beleidsmedewerker van het ministerie van Justitie en Veiligheid (JenV), het ministerie waar het CJIB als uitvoeringsorganisatie onder valt.

Daarnaast heeft een feitencheck op de beschrijving van het pilotproject plaatsgevonden: feedback op de beschrijving van de pilot in de conceptcasestudy is gegeven door het CJIB en door de beleidsverantwoordelijke binnen het ministerie van JenV. De verkenning van kansen en risico's voor de geselecteerde waarden vanuit het perspectief van de burger heeft plaatsgevonden middels twee interviews. Ten eerste heeft een interview plaatsgevonden met een advocaat socialezekerheidsrecht, die ruime ervaring heeft met cliënten met schuldenproblematiek. Verder is gesproken met de directeur van het Instituut voor het Midden- en Kleinbedrijf (IMK). Dit interview is afgenomen omdat het IMK betrokken is bij de schuldenproblematiek van ondernemers. Voor ondernemers met schulden vervult het IMK een adviesfunctie.

[25] WRR 2017, p. 151.
[26] Aan ons is door het CJIB een aantal beleidsdocumenten verstrekt. De beslisboom die is gebruikt in de pilot Telefonisch Innen was ten tijde van dit onderzoek niet openbaar. De uitvoering van het project is op een aantal onderdelen wel in een aantal openbare stukken beschreven, zie onder andere Position paper CJIB 2018 en Rapport Decisio & DSP 2019.

7.2 Pilot Telefonisch Innen bij verkeersboetes

Hieronder zal allereerst in 7.2.1 de pilot Telefonisch Innen worden beschreven. Deze pilot omvat een beslisboom, die gebruikmaakt van een algoritme. Par. 7.2.2 bespreekt de werking van dit algoritme. Vervolgens verkent par. 7.2.3 een aantal toekomstige ontwikkelingen in de inzet van algoritmen bij de overheid. Afsluitend volgt een tussenconclusie.

7.2.1 De pilot

Het CJIB maakt onderdeel uit van het ministerie van JenV. Het CJIB is bij het grote publiek vooral bekend vanwege de inning van verkeersboetes op grond van de Wahv. Het CJIB int deze vorderingen namens de minister van JenV.[27] Het betreft een administratieve sanctie: een aan de Staat te betalen geldsom, wegens overtreding van (onder andere) de Wegenverkeerswet 1994.[28] Hoofdstuk VIII van de Wahv omschrijft precies hoe het CJIB bij incasso dient te handelen en wat de consequenties zijn van niet-betalen. De vordering dient binnen twee weken te zijn betaald nadat de Wahv-boetebeschikking juridisch onherroepelijk is geworden. Dit komt erop neer dat de boete moet worden betaald binnen acht weken nadat de beschikking door het CJIB is verzonden.[29] Als een burger een boete niet kan voldoen, kan de burger het CJIB verzoeken om het bedrag in termijnen te mogen betalen.[30] Als geen verzoek is gedaan en betaling na acht weken uitblijft, wordt de boete automatisch op grond van de Wahv verhoogd.[31] Er geldt dan een nieuwe betaaltermijn van vier weken.[32] Als opnieuw niet tijdig wordt betaald, verstuurt het CJIB een tweede aanmaning. Het boetebedrag wordt dan opnieuw op grond van de wet verhoogd.[33] Als de vordering ook na de tweede aanmaning niet tijdig[34] wordt betaald, zal de eerste incassostap worden gezet: het CJIB kan het boetebedrag afschrijven van de rekening van de debiteur.[35] Deze methode leidt niet in alle gevallen tot een volledige incasso, bijvoorbeeld als er onvoldoende geld op de bankrekening staat.[36] De volgende stap is dan het inschakelen van een deurwaarder.[37] Dit brengt voor burgers vaak aanzienlijke kosten met zich. Een geïnterviewde advocaat onderschrijft de ingrijpende gevolgen die het inschakelen van een deurwaarder kan hebben: ‘Een rekening van enkele tientallen euro’s kan uitgroeien tot vele honderden euro’s’.[38] De pilot Telefonisch Innen beoogt voor

[27] Art. 22 Wahv.
[28] Artt. 1 en 2 Wahv; Art. 63d Organisatiebesluit Ministerie van Justitie en Veiligheid.
[29] Art. 23 lid 1 Wahv jo. artt. 6:7 Awb jo. 6:8 Awb jo. art. 3:41 Awb. Het is mogelijk om tegen een boetebeschikking op grond van de Wahv binnen een termijn van 6 weken (art. 6:7 Awb) administratief beroep in te stellen bij de officier van justitie (hoofdstuk IV Wahv) en vervolgens beroep bij de kantonrechter (artt. 8:1 Awb jo 7:1 lid 1 onder a Awb). Een burger wordt pas gebeld door het CJIB nadat alle beroepsmogelijkheden onbenut zijn gelaten of juist zijn uitgeput. De vordering staat dan juridisch vast. Interview CJIB 26-6-2019.
[30] Art. 4:14 beleidsregel tenuitvoerlegging strafrechtelijke en administratiefrechtelijke beslissingen.
[31] Art. 23 lid 3 Wahv.
[32] Art. 24 lid 1 Wahv.
[33] Art. 25 lid 1 Wahv.
[34] Binnen de vervaltermijn van de aanmaning.
[35] Art. 27 lid 1 sub c Wahv jo art. 4:34 Beleidsregel tenuitvoerlegging strafrechtelijke en administratiefrechtelijke beslissingen. Zie voor een overzicht ‘Verkeersboete’, cjb.nl/verkeersboete-m.
[36] Schoneveld e.a. 2018, bijlage 2.
[37] Art. 26 Wahv, artt. 4:1 en 4:5 Beleidsregel tenuitvoerlegging strafrechtelijke en administratiefrechtelijke beslissingen.
[38] Interview met advocaat socialezekerheidsrecht 9-2-2019. Zie het Besluit tarieven ambtshandelingen gerechtsdeurwaarders en het Besluit buitengerechtelijke kosten. Voor een overzicht zie De Meulder & Yildirim 2018, p. 10-11.

de geselecteerde groep (tijdelijk) niet-zelfredzame burgers te voorkomen dat het nodig is om een gerechtsdeurwaarder in te schakelen met bijkomende kosten en spanningen.[39]

Uitgangspunt van de pilot blijft overigens wel dat de vordering door de burger wordt terugbetaald. Het CJIB is een overheidsinstantie en is daarom gebonden aan de beginselplicht tot handhaving, in dit geval handhaving van onder andere de Wegenverkeerswet. Indien een burger geen rechtsmiddelen instelt tegen een boete, komt de juridische rechtmatigheid hiervan vast te staan. De beginselplicht tot handhaving brengt dan de beginselplicht tot invordering met zich: openstaande vorderingen op burgers die voortvloeien uit een onherroepelijk geworden boete moeten daadwerkelijk worden geïnd.[40]

Het CJIB beschikt over historische data over de inning van Wahv-vorderingen. Van elke vordering zijn kenmerken bekend, bijvoorbeeld of de vordering uiteindelijk is betaald. Op basis van deze gegevens heeft het CJIB een beslisboom ontwikkeld die voor een onherroepelijk geworden vordering op een burger kan voorspellen wat de kans is dat, als de vordering wordt overgedragen aan de deurwaarder, deze de vordering binnen de daarvoor gestelde termijn van drie jaar kan verhalen op de burger.[41] In de pilot worden Wahv-vorderingen die juridisch onherroepelijk zijn geworden en die niet binnen de vervaltermijn van de tweede aanmaning zijn betaald getoetst aan deze beslisboom. Als de vordering 'hoog' scoort (dat wil zeggen: de kans op verhaalbaarheid van de vordering wordt als hoog ingeschat), betekent dit dat de betreffende burger in aanmerking komt voor telefonisch contact binnen de pilot.[42]

Aan bovenstaande keuzes in de beslisboom ligt de volgende beleidsaanname ten grondslag. Als de kans hoog is dat een deurwaarder een vordering kan verhalen, betekent dit dat de burger over de financiële middelen beschikt om de vordering (in termijnen) af te betalen. Bij de zaken die worden ingevoerd in het algoritme is de vordering niet betaald binnen twaalf weken nadat het CJIB de boetebeschikking aan de burger heeft verzonden. Ook heeft de betreffende burger geen regeling getroffen met het CJIB. Dit kan duiden op het niet-betalen van de boete als gevolg van verminderde zelfredzaamheid (men *wil* wel betalen, maar *kan* dit niet). Het CJIB noemt een slechte financiële administratie als mogelijke reden voor dit niet-kunnen.[43] Wij maken hieruit op, dat de

[39] Interview CJIB, 26-06-2019, interview ministerie JenV 5-12-2019. In dezelfde zin: Position paper CJIB 2018.
[40] De beginselplicht tot handhaving rust op het bevoegde bestuursorgaan, in het geval van de Wahv-boetes de minister van JenV, namens wie het CJIB de boetes int. Zie Michiels, Blomberg & Jurgens 2016, p. 288.
[41] De initiële dataset bestond uit WAHV-zaken die in de periode van 1-1-2013 tot 1-7-2013 zijn toebedeeld aan de deurwaarder. De afloop (heeft het inschakelen van de deurwaarder geleid tot uitstroom van de zaak, bijvoorbeeld doordat is betaald of een betalingsregeling is getroffen?) wordt gemeten drie jaar na toedeling aan de deurwaarder (WODC 2017, p. 8-10). De mogelijke periode waarin verhaal kan plaatsvinden was indertijd twee jaar nadat ten aanzien van de administratieve sanctie een onherroepelijke beslissing is genomen. Inmiddels is deze termijn per 1 januari 2018 gewijzigd in drie jaar (art. 25 lid 3 Wahv).
[42] Interview CJIB, 26-06-2019.
[43] Interview CJIB, 26-06-2019. *Visie tenuitvoerlegging financiële sancties* 2015.

pilot (mede) als doel heeft om burgers te traceren die niet zelfredzaam zijn als gevolg van verminderd doenvermogen.

Als de beslisboom een hoge betaalkans bij de deurwaarder voorspelt, beslist een medewerker van het CJIB of de burger ook daadwerkelijk wordt gebeld. Zo ja, dan neemt de medewerker telefonisch contact op met de betreffende burger.[44] In de gesprekken met burgers wordt ingezet op het vergroten van de betalingsbereidheid van de betrokkene dan wel op het doorbreken van diens betalingsonmacht (bijvoorbeeld door te wijzen op de mogelijkheid van betaling in termijnen).[45] In de juridische positie van de schuldenaar verandert niets: het uitgangspunt blijft dat de vordering wordt betaald.[46] Het CJIB geeft aan dat het inningstraject niet wordt opgeschort voor de groep die door het algoritme wordt geselecteerd.[47] De burgers die niet actief worden benaderd kunnen, door zelf te bellen met het CJIB, dezelfde behandeling bewerkstellingen als de groep die wel wordt gebeld. Beide groepen burgers kunnen een beroep doen op uitzonderingsregelingen, zoals een betalingsregeling of de 'noodstopprocedure geldelijke sancties'.[48] Wij zijn er daarom van uit gegaan dat het al dan niet opbellen van burgers in de pilot geen wijziging aanbrengt in de juridische positie van een burger.

7.2.2 Werking van het algoritme in de pilot

De in de pilot gehanteerde beslisboom gebruikt een algoritme dat is getraind met historische, gelabelde data.[49] De beslisboom omvat kenmerken van de betreffende Wahv-boete en kenmerken van eerdere Wahv-boetes van de burger, alsmede informatie over het betalingsgedrag van de burger bij Wahv-vorderingen in de afgelopen twee jaar. Belangrijke indicatoren zijn het bedrag van de boete in de huidige zaak, het openstaande bedrag van alle Wahv-boetes van de burger en of de burger in het afgelopen jaar een andere boete bij de deurwaarder heeft betaald.

Op basis van statistische analyse is door het CJIB een beslisboom opgesteld, waarbij voor een nieuwe vordering een beperkt aantal stappen wordt doorlopen. Bij elke nieuwe stap of 'vertakking' splitst de boom zich verder in tweeën. In de beslisboom komt een nieuwe vordering vijfmaal voor een dergelijke 'splitsing' te staan, waarbij iedere 'afslag' die wordt genomen bepaalt voor welke verdere 'splitsingen' de zaak komt te staan. Uit deze beslisboom resulteert de betaalkans van de

[44] Interview CJIB, 26-06-2019.
[45] *Visie tenuitvoerlegging financiële sancties* 2015.
[46] Zie bijvoorbeeld art. 4:17 van de Beleidsregel tenuitvoerlegging strafrechtelijke en administratiefrechtelijke beslissingen, voor de beperkingen ten aanzien van het treffen van een betalingsregeling. Zie voor enige ruimte art. 4:1 van de beleidsregel: deze maakt een persoonsgerichte aanpak mogelijk. Ook hier blijft betaling van de boete het uitgangspunt.
[47] Wel is het zo dat wanneer een burger een aanvraag voor een betalingsregeling doet, de zaak voor de beoordelingsperiode van deze aanvraag tijdelijk in de wacht wordt geplaatst (CJIB, schriftelijke aanvulling op 2-3-2020). Ook burgers die niet worden gebeld kunnen een betalingsregeling aanvragen; de website van het CJIB informeert burgers hierover 'Ik kan mijn boete niet in één keer betalen', cjb.nl.
[48] *Kamerstukken II* 2019/20, 24515 nr. 525.
[49] WODC 2017, p. 9.

vordering bij de deurwaarder. Bij een hoge betaalkans zal een CJIB-medewerker proberen de betreffende burger te bellen.[50]

Voor het trainen van het algoritme in de beslisboom is gebruikgemaakt van geanonimiseerde data.[51] Er is sprake van *supervised machine learning*: het algoritme waarop de beslisboom is gebaseerd is getraind met gelabelde voorbeelden (zie verder par. 2.1.2). Er is sprake van een gesloten dataset. De voorspellende kracht van de beslisboom wordt gemonitord en de trainingsdata voor het algoritme wordt, indien nodig, ververst.[52]

De uitkomsten van het algoritme zijn relatief stabiel. Dit komt onder andere doordat het algoritme is gebaseerd op een beperkt aantal voorspellende variabelen die een sterke statistische samenhang hebben met de voorspelde variabele. Bijvoorbeeld: kleinere bedragen worden sneller betaald dan grotere bedragen. Dit soort statistische correlaties zijn relatief stabiel, en daarmee is het algoritme dat ook.[53] De beslisboom in de pilot is relatief goed uitlegbaar in vergelijking met andere zelflerende algoritmen zoals *deep learning* of *unsupervised learning* (par. 2.1.2). De beslisboom heeft minder dan tien variabelen en er zijn in totaal vijf 'vertakkingen' in de boom. Ook is de bijdrage van elk van de variabelen aan de voorspelkracht van het model na te gaan in de beslisboom. De mate van uitlegbaarheid is een belangrijke factor geweest voor het CJIB bij de voorkeur voor een beslisboom boven een ander type algoritme.[54]

Het algoritme opereert niet autonoom, maar wordt beslissingsondersteunend ingezet. Het algoritme genereert een lijst met zaken waarin wordt geadviseerd om de betreffende burger te bellen.[55] De medewerker van het CJIB die dit advies krijgt, is niet verplicht om te bellen en wordt ook niet beoordeeld op het aantal gepleegde telefoontjes. Enkele voorbeelden waarin een CJIB-medewerker geen telefonisch contact opneemt, zijn situaties waarin de boete inmiddels toch is betaald, of waarin de burger inmiddels zelf al heeft gebeld.[56] De betreffende medewerker ziet deze en andere informatie die niet in het algoritme is verwerkt en kan vervolgens beslissen om al dan niet te bellen.[57]

[50] Ministerie van JenV, schriftelijke aanvulling op 30-1-2020.
[51] Belangrijk voor de vraag of een burger zal gaan betalen is verder of het gaat om een kentekenconstatering (constatering overtreding door bijvoorbeeld een flitspaal) of een staandehouding (politieagent of buitengewoon opsporingsambtenaar legt ter plaatse een boete op). De kans op betaling is groter bij een kentekenconstatering. In totaal gaat het om minder dan tien variabelen (WODC 2017, p. 13 en interview CJIB 9-9-2019).
[52] Interview CJIB 9-9-2019, CJIB, schriftelijke aanvulling op 30-1-2020. Het CJIB geeft aan, dat daarbij de trainingsdata zo wordt samengesteld, dat bias wordt voorkomen (bijvoorbeeld een zelfversterkend effect van het model). Daarnaast wordt gekeken naar de stabiliteit en generaliseerbaarheid van het nieuwe model.
[53] Interview CJIB 9-9-2019.
[54] Interview CJIB 9-9-2019.
[55] Interview CJIB, 26-06-2019; Interview Ministerie van JenV 5-12-2019; Ministerie van JenV, Uitvoeringstoets 2019, p. 11.
[56] Interview CJIB 9-9-2019.
[57] Interview CJIB 9-9-2019.

7.2.3 Blik op de toekomst

Door onderzoeksbureau Decisio is onderzoek gedaan naar het gebruik van algoritmen door overheden voor een persoonsgerichte inning. Hieruit kwam naar voren dat diverse overheidsinstanties (naast het CJIB ook o.a. DUO en UWV) algoritmen hebben ontwikkeld voor dit doel en deze in pilots testen. Het gaat dan om relatief eenvoudige statistische technieken[58] waarmee betaalprofielen worden ontwikkeld: een verzameling van kenmerken van een persoon waarmee debiteuren worden ingedeeld in groepen. Uit het onderzoek kwam ook naar voren, dat meer overheidsinstanties in de nabije toekomst dergelijke betaalprofielen willen ontwikkelen en testen.[59]

Als algoritmen in de toekomst op grotere schaal worden ingezet voor overheidsincasso, is te verwachten dat daarvoor de relatief minder complexe algoritmen zullen worden ingezet. Ook het algoritme dat is gebruikt in de pilot Telefonisch Innen is, zoals is beschreven, relatief eenvoudig. Het CJIB heeft er bewust niet voor gekozen complexere algoritmen in te zetten, onder andere omdat men de burger wil kunnen uitleggen waarom de overheid een bepaalde handeling verricht.[60] Ook andere overheden gaven in een onderzoek door het Centraal Bureau voor de Statistiek (CBS) aan dat ze bij het inzetten van algoritmen in beleid rekening houden met hoe goed het algoritme verklaard en uitgelegd kan worden en in welke mate het getoetst kan worden op nauwkeurigheid en juistheid.[61] De genoemde overwegingen spelen een nog grotere rol als het gaat om het nemen van bindende beslissingen jegens burgers.[62] Ook uit de door ons gehouden interviews komt naar voren dat overheden terughoudend zijn om in de toekomst complexere *machine learning*-algoritmen in te zetten voor het uitvoeren van overheidsbeleid.[63]

7.2.4 Tussenconclusie

In deze paragraaf is de pilot Telefonisch Innen geschetst, alsmede het algoritme dat in deze pilot een rol speelt. Ook is belicht, hoe overheden aankijken tegen de inzet van algoritmen in de nabije toekomst. Hieruit blijkt dat overheden terughoudend zijn om meer complexe vormen van *machine learning*-algoritmen in te zetten. Ook in de pilot Telefonisch Innen heeft het CJIB ervoor gekozen om een algoritme in te zetten dat relatief goed uitlegbaar is aan burgers. De volgende paragrafen belichten de casestudy vanuit de publieke waarden rechtsbescherming, non-discriminatie en de bescherming van persoonsgegevens.

[58] De onderzoekers beschrijven regressieanalyse en clusteranalyse (Rapport Decisio & DSP 2019, p. 10-15).
[59] Rapport Decisio & DSP 2019, p. 20-26.
[60] Het CJIB geeft aan, dat 'het relatief eenvoudige model goed presteert ten opzichte van complexere *machine learning* algoritmen (zoals een *random forest*), maar tegelijkertijd wel een stuk beter uitlegbaar is' (schriftelijke aanvulling op 30-1-2020).
[61] Doove & Otten 2018, p. 9.
[62] Interview VNG 29-8-2019.
[63] Interview VNG 29-8-2019; interview CJIB 9-9-2019.

7.3 Rechtsbescherming

Deze paragraaf is als volgt opgebouwd. Allereerst wordt de publieke waarde rechtsbescherming voor deze casestudy vertaald naar een juridisch kader. Hier zullen, in aanvulling op par. 3.3 van dit rapport, een aantal juridische aspecten worden uitgelicht die van belang zijn voor de casestudy. Daarna worden de huidige en toekomstige potentiële juridische kansen en risico's in kaart gebracht voor het pilotproject en voor de inzet van zelflerende algoritmen ter identificatie van niet-zelfredzame burgers. Afsluitend volgt een tussenconclusie.

7.3.1 Rechtsbescherming en algoritmen in overheidsincasso

Rechtsbescherming tegen de overheid

Er zijn verschillende procedures om geschillen met de overheid te beslechten. Tegen overheidsbeslissingen, inhoudende besluiten in de zin van art. 1:3 Awb, staat beroep open bij de bestuursrechter.[64] Echter, voordat beroep openstaat dient een voorprocedure (doorgaans bezwaar[65]) te worden doorlopen. Als een burger tegen een beslissing bezwaar maakt is het bestuursorgaan[66] dat dit besluit heeft genomen, gehouden om deze beslissing te heroverwegen.[67] Tegen overheidsoptreden dat *geen* besluit inhoudt, staat geen bezwaar of beroep open,[68] maar kan wel worden geprocedeerd bij de civiele rechter.[69] Dit is het geval in de pilot Telefonisch Innen: we zagen dat geen sprake is van een rechtshandeling, te weten een handeling die een wijziging beoogt in de rechtspositie van de burger. Dit is een van de vereisten die art. 1:3 Awb stelt voor een overheidshandeling om als besluit te worden aangemerkt.[70]

Naast een gang naar de rechter is het ook mogelijk in een buitengerechtelijke procedure op te komen tegen overheidsoptreden. Drie procedures zijn vooral relevant voor deze casestudy. Ten eerste kunnen burgers die een klacht hebben over vermeend discriminatoir handelen door de overheid deze voorleggen aan het College voor de Rechten van de Mens (CRM).[71] De uitspraken die het CRM doet zijn juridisch niet-bindend, wel kan het College beslissen om een zaak aan te brengen bij de civiele rechter, om zo een bindend oordeel te verkrijgen.[72] Het CRM heeft een breed mandaat en kan dus zich uitspreken over mensenrechten die zijn vastgelegd in de Nederlandse Grondwet, de AWGB of in verdragen en richtlijnen.[73]

[64] Art. 8:1 Awb.
[65] Art. 7:1 Awb
[66] Art. 1:1 Awb.
[67] Art. 7:11 Awb.
[68] Art. 8:1 Awb.
[69] HR 31 december 1915, ECLI:NL:HR:1915:AG1773 (*Noordwijkerhout/Guldemond*). De minister van JenV is een bestuursorgaan; het CJIB neemt besluiten namens (onder andere) de minister, zie par. 7.2.1.
[70] Bröring & De Graaf 2019, p. 159.
[71] Art. 10 Wet College voor de Rechten van de Mens.
[72] Art. 13 lid 1 Wet College voor de Rechten van de Mens.
[73] Artt. 1 en 3 Wet College voor de Rechten van de Mens; *Kamerstukken II* 2009/10, 32467, nr. 3, p. 5.

Verder is, als het gaat om de bescherming van persoonsgegevens, van belang dat een belanghebbende kan verzoeken aan de toezichthouder (in Nederland: de AP) om in actie te komen tegen onrechtmatige gegevensverwerking. De belanghebbende zal dan een klacht moeten indienen bij de AP. Tegen een reactie op een dergelijke klacht staat eerst bezwaar bij de AP en dan beroep bij de bestuursrechter open.[74]

Ten slotte kan een burger op grond van hoofdstuk 9 Awb zich wenden tot een lokale of de nationale ombudsman. De ombudsman behandelt klachten van burgers over overheidsoptreden.[75] Een klacht bij de ombudsman kan juist voor overheidsoptreden zoals in de pilot Telefonisch Innen een geschikt middel zijn. In de pilot is er immers sprake van een beleidsaanpak, waarbij tussen groepen wordt gedifferentieerd in de wijze van bejegening (telefonisch contact opnemen).[76] Het beoordelen van de kwaliteit van dienstverlening door de overheid aan burgers is een van de kerntaken van de ombudsman. In het vervolg van deze paragraaf zal daarom worden gefocust op klachtbehandeling door de ombudsman.

Klachtbehandeling door de ombudsman

Bij de klachtbehandeling door een ombudsman staat niet zozeer de juridische rechtmatigheid maar de behoorlijkheid van overheidsoptreden centraal.[77] De juridische basis voor klachtbehandeling is art. 5 Gw, dat burgers het recht geeft om bij het bevoegd gezag een schriftelijk verzoek in te dienen.

Een burger met een klacht over het optreden van een bestuursorgaan van de rijksoverheid kan terecht bij de nationale ombudsman.[78] Alvorens een klacht in te dienen bij de nationale ombudsman dient de burger eerst een klachtprocedure bij de overheidsinstantie zelf te doorlopen.[79] In dit geval dient een klachtprocedure te worden doorlopen bij de minister van JenV. Na het volgen van deze interne klachtprocedure kan de ombudsman worden verzocht om een onderzoek in te stellen naar de gedraging, aldus art. 9:18 Awb. Een verzoek kan zowel worden ingesteld door natuurlijke personen als door rechtspersonen.[80] De nationale ombudsman neemt ook telefonisch ingediende verzoeken in behandeling.[81] De procedure is daarmee relatief laagdrempelig voor burgers.

Art. 9:18 Awb omschrijft een gedraging als de wijze waarop een bestuursorgaan zich in een bepaalde aangelegenheid jegens de betreffende burger of jegens een ander heeft gedragen.

[74] Zie uitvoeriger Kranenborg & Verhey 2018.
[75] Marseille & Tolsma 2019, p. 417-443.
[76] Marseille & Tolsma 2019, p. 419.
[77] Art. 9:27 Awb.
[78] Art. 1a Wet Nationale ombudsman.
[79] Art. 9:20 lid 1 Awb.
[80] *Kamerstukken II* 2002/03, 28747, nr. 3, p. 26.
[81] Jaarverslag 2018 nationale ombudsman, p. 45.

Klachten over beleid, over de uitvoering van beleid of over het algemene optreden van een bestuursorgaan vallen onder dit bereik, voor zover deze betrekking hebben op de uitwerking hiervan op een specifieke burger die door dit beleid wordt geraakt. Het kan gaan om een rechtshandeling of om een feitelijke handeling (zonder juridische gevolgen), maar ook om een nalaten.[82] Als tegen het overheidsoptreden een bestuursrechtelijke procedure openstaat of openstond is het bestuursorgaan niet verplicht de klacht intern te behandelen.[83] Ook de ombudsman is in deze situaties niet verplicht om een onderzoek in te stellen.[84] De ombudsman behandelt in de praktijk vooral klachten over feitelijk overheidshandelen, waartegen geen bestuursrechtelijke procedure openstaat.[85]

In meer dan driekwart van de bij de nationale ombudsman ingediende verzoeken gaat de ombudsman niet over tot een inhoudelijk onderzoek. Het geschil wordt opgelost doordat de burger anderszins op weg wordt geholpen, bijvoorbeeld doordat wordt bemiddeld tussen de burger en de overheidsinstantie.[86] Indien het wel komt tot een onderzoek, beoordeelt de ombudsman of het betreffende bestuursorgaan behoorlijk heeft gehandeld.[87] Als leidraad voor de overheid en voor lokale ombudsmannen heeft de nationale ombudsman richtlijnen opgesteld.[88] Deze behoorlijkheidsnormen zijn afgeleid van de algemene beginselen van behoorlijk bestuur: algemene rechtsbeginselen die in de jurisprudentie van de civiele rechter en de diverse bestuursrechters zijn ontwikkeld.[89] In de literatuur bestaat discussie over de mate waarin de ombudsman (ook) een toetsing op juridische rechtmatigheid uitvoert of zou moeten uitvoeren. Indien een gedraging onrechtmatig is zal de ombudsman deze doorgaans ook als onbehoorlijk bestempelen. Daarmee omvat de behoorlijkheidstoetsing vaak mede een rechtmatigheids-toetsing.[90]

De nationale ombudsman kan uit ook eigen beweging onderzoek doen.[91] Zo'n onderzoek richt zich op mogelijke structurele problemen in de relatie tussen overheid en burger. Het gaat daarbij niet om individuele situaties. In 2019 publiceerde de nationale ombudsman negen van dergelijke onderzoeken, waaronder een onderzoek naar de knelpunten die burgers ervaren bij het invorderen van schulden door de overheid: 'Invorderen vanuit het burgerperspectief' (2019).[92]

[82] Langbroek & Rijpkema 2007, p. 39.
[83] Art. 9:8 lid 1 Awb.
[84] Art. 9:23 Awb.
[85] Marseille & Tolsma 2019, p. 417. Zie ook het hiervoor geschetste wettelijk kader.
[86] Marseille & Tolsma 2019, p. 439.
[87] Art. 9:27 Awb.
[88] Marseille & Tolsma 2019, p. 440.
[89] Deze beginselen zijn deels gecodificeerd in de Awb; Langbroek & Rijpkema 2007, p. 273.
[90] Marseille & Tolsma 2019, p. 440-442.
[91] Art. 9:26 Awb.
[92] Rapport nationale ombudsman 2019.

Het oordeel van de nationale ombudsman is niet dwingend voor de betreffende overheidsinstantie. De ombudsman beschikt niet over middelen waarmee naleving van een advies of een onderzoek kan worden afgedwongen. Door Langbroek en Rijpkema wordt het oordeel getypeerd als 'niet afdwingbaar, wel dwingend'. Het gezag van de uitspraken van de ombudsman volgt volgens hen onder andere uit de openbaarheid van diens adviezen en onderzoeken.[93]

In zijn Behoorlijkheidswijzer uit 2019 formuleert de nationale ombudsman onder andere de plicht van de overheid om de fundamentele rechten te respecteren, waaronder het recht om niet gediscrimineerd te worden. Verder dient de overheid zorg te dragen voor fatsoenlijke en hulpvaardige bejegening van burgers. Ook benadrukt hij de plicht om te streven naar maatwerk en individualisering in de besluitvorming. De ombudsman geeft aan dat ook bij overheidshandelen dat geen juridische gevolgen heeft, de overheid ernaar zou moeten streven om het handelen af te stemmen op de specifieke omstandigheden van een individuele burger.[94] Voor de pilot 'Telefonisch Innen' is met name het non-discriminatiebeginsel van belang, dat ook deel uitmaakt van de behoorlijkheidsnormen van de ombudsman. Dit beginsel zal in par. 7.4 worden besproken.

7.3.2 Kansen en risico's in relatie tot het juridisch kader

De pilot Telefonisch Innen beoogt maatwerk te leveren voor individuele burgers. Hiermee sluit de pilot aan bij de behoorlijkheidsnormen van de nationale ombudsman omtrent hulpvaardige bejegening en maatwerk. Ook sluit de pilot hiermee aan bij de aanbevelingen van de nationale ombudsman voor de invordering van schulden door de overheid. In het rapport *Invorderen vanuit het burgerperspectief* wordt gesteld, dat de overheid zich moet inspannen om verdere schulden voor burgers te voorkomen. Daarnaast dient de overheid duidelijk te communiceren en ernaar te streven om persoonlijk contact op te nemen waar nodig. Zo mogelijk dient de overheid maatwerk te leveren.[95]

In zijn jaarverslag over 2018 wijst de nationale ombudsman erop dat overheden steeds vaker gebruikmaken van profilering. Door de koppeling van grote hoeveelheden data en de toepassing van algoritmen hierop ontstaan profielen van burgers. De nationale ombudsman signaleert dat overheden ervan uit lijken te gaan dat profilering geen grote impact op individuen heeft. Het is daarom noodzakelijk om de risico's voor de burger bij profilering door de overheid nader in kaart te brengen, aldus de ombudsman.[96] Het doen van nader onderzoek hiernaar behoort tot de taken van de ombudsman. We zagen hiervoor immers dat de ombudsman onderzoek kan doen naar 'een gedraging van een bestuursorgaan'.

[93] Langbroek & Rijpkema 2007, p. 23.
[94] Behoorlijkheidswijzer 2019.
[95] Rapport nationale ombudsman 2019, p. 4.
[96] Jaarverslag 2018 nationale ombudsman, p. 27.

Een algemeen risico is echter dat de ombudsman, door een gebrek aan transparantie bij overheden, deze taak niet voldoende kan uitoefenen. Indien burgers niet weten waar en hoe algoritmen door de overheid worden ingezet, zullen zij hier geen klacht bij de ombudsman over indienen. Algoritmische transparantie is daarom een voorwaarde voor controle op de inzet van algoritmen door de overheid.[97] In een recent kabinetsstandpunt lijkt het kabinet te pleiten voor zoveel mogelijk transparantie naar burgers over het gebruik van algoritmen door de overheid, ook tijdens de pilotfase van projecten. Concrete toezeggingen worden echter op dit punt niet gedaan.[98]

Bovenstaand algemeen risico wordt in het hierboven besproken juridisch kader deels ondervangen doordat de ombudsman uit eigen beweging onderzoek kan doen (par. 7.3.1). Naast het opstarten van een onderzoek kan de ombudsman ook op meer informele wijze zicht blijven houden op de omgang met algoritmen door de overheid.[99] De vraag of bovenstaand risico voldoende wordt ondervangen, en daarmee het juridisch kader voldoende bestendig is, kan op basis van het individuele praktijkvoorbeeld dat in deze casestudy centraal staat niet worden beantwoord.

7.3.3 Tussenconclusie

We zagen dat een burger die een klacht heeft over een gedraging van een bestuursorgaan een klachtprocedure kan starten: eerst intern, vervolgens extern bij de ombudsman. De ombudsman kan ook uit eigen beweging een onderzoek starten, als er aanwijzingen zijn dat er structureel problemen bestaan bij de bejegening van burgers door de overheid. Het gericht benaderen van burgers, zoals dit plaatsvindt binnen de pilot Telefonisch Innen, sluit aan bij de aanbevelingen van de nationale ombudsman.

7.4 Non-discriminatie

Deze paragraaf is als volgt opgebouwd. Allereerst wordt de publieke waarde non-discriminatie voor deze casestudy vertaald naar een juridisch kader. Hier zal, in aanvulling op par. 3.2 van voorliggend onderzoek, een aantal juridische aspecten worden uitgelicht die van belang zijn voor de casestudy. Daarna worden de huidige en toekomstige juridische potentiële kansen en risico's in kaart gebracht voor het pilotproject en voor het inzetten van algoritmen voor het versterken van de zelfredzaamheid van burgers. Afsluitend volgt een tussenconclusie.

7.4.1 Non-discriminatie en algoritmen in overheidsincasso

Non-discriminatie hangt sterk samen met het gelijkheidsbeginsel als beginsel van de democratische rechtsstaat, dat vereist dat gelijke gevallen door de overheid gelijk worden

[97] Meuwese 2017, p. 168 en p. 174.
[98] Bijlage bij *Kamerstukken II* 2019/20, 26643, nr. 641, p. 15.
[99] Zie bijvoorbeeld Congresmagazine nationale ombudsman 2019, waar ook het project Telefonisch Innen wordt omschreven.

behandeld en ongelijke gevallen ongelijk, rekening houdend met de mate van verschil.[100] Art. 1 Gw bevat dit non-discriminatiebeginsel en benoemt een aantal gronden waarop discriminatie niet is toegestaan. Een andere belangrijke non-discriminatiebepaling voor publiekrechtelijk overheidshandelen is terug te vinden in art. 14 EVRM. Daarbij is relevant dat art. 14 EVRM enkel kan worden ingeroepen door een burger indien er tevens inbreuk wordt gemaakt op een van de andere grondrechten uit het verdrag.[101] Een ander voorbeeld van een non-discriminatiebepaling voor publiekrechtelijk overheidshandelen is terug te vinden in art. 1 van Protocol 12 bij het EVRM. Deze bepaling verbiedt overheidsinstanties ('enig openbaar gezag') om burgers te discrimineren op onder andere geslacht, ras en godsdienst, maar ook op vermogen of een andere status.[102]

Bij dienstverlenend overheidsoptreden, zoals in de pilot Telefonisch Innen, is het onderscheid tussen een formele en een materiële visie op gelijkheid van belang. Formele gelijkheid benadrukt de gelijke behandeling van gelijke gevallen. Het recht dient procedurele waarborgen te bieden en willekeur door de overheid tegen te gaan. Materiële gelijkheid ziet gelijkheid als doel op zich: het streven naar een samenleving waarin er sprake is van sociale gelijkheid. Deze visie benadrukt het belang van het ongelijk behandelen van ongelijke gevallen. Maatschappelijke ongelijkheden dienen in deze visie te worden rechtgezet.[103]

In de pilot Telefonisch Innen wordt gedifferentieerd tussen groepen burgers. Inherent hieraan bestaat het risico dat burgers menen dat zij worden gediscrimineerd. Het is op voorhand lastig aan te geven op welke non-discriminatiebepaling een burger zich zal beroepen indien deze meent, dat hij of zij als gevolg van de pilot Telefonisch Innen wordt gediscrimineerd. Dit zal afhangen van de omstandigheden van het geval. De betreffende burger kan een procedure aanhangig maken bij het College voor de Rechten van de Mens of bij de rechter. Zie voor de beschrijving van deze vormen van rechtsbescherming par. 7.3.

We zagen in de vorige paragraaf dat het handelen van het CJIB in de pilot Telefonisch Innen voorgelegd kan worden aan de nationale ombudsman. De behoorlijkheidsnorm is dan de toepasselijke materiële norm waarop het overheidshandelen wordt beoordeeld. Deze omvat mede een toetsing aan het non-discriminatiebeginsel.

7.4.2 Kansen en risico's in relatie tot het juridisch kader

Pilot Telefonisch Innen

De inzet van een telefonische betalingsherinnering door de overheid kan leiden tot meer materiële gelijkheid. Immers, als de overheid burgers die mogelijk minder zelfredzaam zijn identificeert en

[100] Bröring & De Graaf 2019, p. 45 e.v.
[101] Kortmann e.a. 2016, p. 415.
[102] Zie verder Vetzo, Gerards & Nehmelman 2018.
[103] Eleveld, *Sociaal Maandblad Arbeid* 2008, p. 313 e.v.

hen actief benadert, kan dit een tegenwicht bieden aan de ongelijke positie van burgers die de WRR in het rapport *Weten is nog geen doen* signaleert.

Tegelijkertijd bestaat bij elk overheidsbeleid dat differentieert tussen burgers het risico dat het non-discriminatiebeginsel als vereiste van behoorlijk overheidshandelen wordt geschonden (par. 7.3). Het discriminatieverbod is in het verleden met succes toegepast in procedures voor de ombudsman.[104] Een tweede potentieel risico voor non-discriminatie in de pilot is dat een statistische vertekening optreedt waardoor verschillen tussen groepen worden uitvergroot. Als algoritmen worden gebruikt om in beleid te differentiëren tussen burgers, bestaat het risico dat het algoritme zichzelf versterkt wanneer de data voor het hertrainen van het model niet zorgvuldig wordt gekozen. Door het algoritme te hertrainen op de uitkomsten van zaken die eerder door het algoritme zijn geselecteerd wordt mogelijk een statistische vertekening geïntroduceerd. Er kunnen dan zelfversterkende effecten optreden. Zie hierover verder par. 2.1.2. Het CJIB geeft aan, dat bij het trainen van het algoritme in de pilot Telefonisch Innen met dit risico rekening is gehouden.[105]

De inzet van zelflerende algoritmen voor het signaleren van niet zelfredzame burgers
Hierboven is al gesignaleerd, dat het bieden van maatwerk middels algoritmen kan leiden tot het in hogere mate verwezenlijken van het ideaal van materiële gelijkheid. Daarnaast kan de inzet van algoritmen bijdragen aan effectiever overheidsoptreden, als het algoritme juist die zaken kan selecteren, waarin een bepaalde interventie effect kan sorteren. Er zijn echter ook risico's. Uit het WRR-rapport volgen twee algemene risico's die spelen als overheden middels algoritmen burgers willen identificeren die onvoldoende zelfredzaam zijn.

Een eerste risico is dat in de trainingsdata van het algoritme statistische verbanden aanwezig zijn, waardoor bij de uitvoering van beleid op basis van dit algoritme strijd ontstaat met het non-discriminatiebeginsel. Uit het WRR-rapport blijkt bijvoorbeeld dat het doenvermogen, en daarmee de zelfredzaamheid, van burgers enigszins samenhangt met hun opleidingsniveau.[106] We leiden hieruit af dat, als een overheidsinstantie algoritmen inzet om burgers met (tijdelijk) verminderd doenvermogen te detecteren, het zou kunnen voorkomen dat daarbij (mede) wordt geselecteerd op opleidingsniveau. In een databestand kan opleidingsniveau statistisch samenhangen met migratieachtergrond.[107] Indien een algoritme op de mate van doenvermogen selecteert, zou het dus mede kunnen selecteren op (bijvoorbeeld) migratieachtergrond.

[104] Langbroek & Rijpkema 2004, p. 92.
[105] Interview CJIB, 26-06-2019.
[106] Het WRR-rapport geeft aan, dat er sprake was van een zwak, maar statistisch significant verband tussen opleidingsniveau en zelfredzaamheid, geoperationaliseerd in de Utrechtse Proactive Coping Competencies. Ook doenvermogen hing enigszins samen met opleidingsniveau (WRR 2017, p. 11).
[107] Zie 'De sociale staat van Nederland 2018: onderwijs', digitaal.scp.nl.

Als algoritmen worden gebruikt om te differentiëren tussen groepen in beleid kan dit bijvoorbeeld leiden tot ander beleid voor burgers met of zonder migratieachtergrond. Dergelijk overheidsbeleid zou in strijd kunnen zijn met het non-discriminatiebeginsel als vereiste van behoorlijk overheidshandelen. Ook zou een burger in een dergelijk geval een procedure kunnen starten bij het College voor de Rechten van de Mens of bij de rechter wegens schending van een non-discriminatiebepaling. Als het ongerechtvaardigde onderscheid leidt tot een besluit in de zin van de Awb, ligt een procedure bij de bestuursrechter voor de hand (zie 7.3.1). De inzet van algoritmen voor het signaleren van burgers met (tijdelijk) verminderd doenvermogen kan dus juridische risico's opleveren als het gaat om non-discriminatie. Overheden dienen zich dan ook bewust te zijn van de verbanden die aanwezig kunnen zijn in de trainingsdata van een algoritme.[108]

Een tweede, algemeen, risico is dat algoritmen die differentiëren op zelfredzaamheid of doenvermogen leiden tot het stigmatiseren van bepaalde groepen burgers.[109] Algoritmen kunnen voorspellingen doen over de *kans* dat een bepaald kenmerk zich voordoet bij een specifieke groep burgers. Echter, een dergelijke voorspelling geeft geen uitsluitsel over al het toekomstige gedrag van een individuele burger in die groep.[110]

In de kabinetsreactie wordt gesproken over algoritmen die burgers met (tijdelijk) verminderd doenvermogen kunnen identificeren. Indien dit soort algoritmen door de overheid zullen worden ingezet, is het dus van belang om de door het algoritme geselecteerde groep niet te stigmatiseren. Als iemand een etiket 'verminderd doenvermogen' of 'niet-kunner' krijgt, kan dit grote implicaties hebben voor het beeld dat anderen van deze persoon hebben. De WRR-onderzoekers signaleren dit risico ook. Zij geven aan dat wie eenmaal een stigma heeft, wellicht niet meer voor zelfredzaam wordt aangezien, of zich misschien voelt ontslagen van de plicht om zich in te zetten.[111]

7.4.3 Tussenconclusie

Het identificeren van burgers die (tijdelijk) niet zelfredzaam zijn, zoals wordt beoogd in de pilot Telefonisch Innen, zou kunnen leiden tot het meer verwezenlijken van het ideaal van materiële gelijkheid. Omdat in de pilot onderscheid wordt gemaakt tussen burgers (formele gelijkheid), speelt inherent het risico dat burgers hiertegen bezwaar zullen hebben. Een van de mogelijke vormen van rechtsbescherming is het hierover indienen van een klacht bij de ombudsman (par. 7.3). Deze beoordeelt dan het non-discriminatiebeginsel als onderdeel van zijn behoorlijkheidstoetsing.

Daarnaast zijn algemene kansen en risico's besproken die zich kunnen voordoen bij de inzet van algoritmen voor het signaleren van burgers met die niet zelfredzaam zijn. Als achterstandsgroepen

[108] Interview ministerie JenV 5-12-2019; interview IMK 8-7-2019; interview CJIB 9-9-2019.
[109] Vedder 2000, p. 165.
[110] Zie ook Meuwese 2017, p. 162.
[111] WRR 2017, p. 121.

met behulp van algoritmen kunnen worden gedetecteerd kan dit bijdragen aan effectiever overheidsbeleid. Dit is dus een kans. Een eerste risico heeft te maken met het feit dat door al aanwezige verbanden in de trainingsdata van een algoritme het algoritme (onbedoeld) een discriminatoir effect heeft. Daarnaast is van belang dat degenen die door een algoritme worden geselecteerd, niet worden gestigmatiseerd. Voorkomen moet worden dat mensen het etiket 'verminderd doenvermogen' of 'niet-kunner' krijgen.

7.5 Bescherming van persoonsgegevens

Deze paragraaf is als volgt opgebouwd. Allereerst wordt de publieke waarde bescherming van persoonsgegevens voor deze casestudy vertaald naar een juridisch kader. Hier zullen, in aanvulling op par. 3.3 van voorliggend onderzoek, een aantal juridische aspecten worden uitgelicht die van belang zijn voor de casestudy. Daarna worden de huidige en toekomstige juridische potentiële kansen en risico's in kaart gebracht voor het pilotproject en voor het inzetten van algoritmen voor het versterken van de zelfredzaamheid van burgers. Afsluitend volgt een tussenconclusie.

7.5.1 De bescherming van persoonsgegevens en algoritmen in overheidsincasso

Informationele privacy en het recht op bescherming van persoonsgegevens, zoals onder andere vastgelegd in art. 10 van de Grondwet en in art. 8 EVRM, zien in beginsel op de relatie tussen overheid en burger (par. 3.2). Bij het gebruik van persoonsgegevens door de overheid gelden dan ook verdergaande waarborgen dan bij het gebruik door een private partij. In de casestudy is met name de uitwerking van het grondrecht op bescherming van persoonsgegevens in de Algemene Verordening Gegevensbescherming (AVG) en de Nederlandse Uitvoeringswet AVG (UAVG) relevant. Hier worden immers de specifieke kaders aangegeven, waarbinnen de overheid bij de verwerking van persoonsgegevens dient te blijven.

Voor het trainen van het algoritme in de beslisboom worden geanonimiseerde gegevens gebruikt. Ook de invoer van de beslisboom bestaat uit niet of moeilijk tot een persoon herleidbare gegevens, zoals de hoogte van de Wahv-boete.[112] Identificerende gegevens, zoals een burgerservicenummer (BSN) of geboortedatum, worden niet als invoer voor het algoritme gebruikt. Indien data niet (meer) is te herleiden tot een persoon, is geen sprake van de verwerking van persoonsgegevens in de zin van art. 2 AVG; de AVG is dan niet van toepassing.[113]

Het benaderen van burgers binnen de pilot is alleen mogelijk indien persoonsgegevens worden verwerkt. Immers, de uitvoer van de beslisboom zal worden gekoppeld aan een individuele burger,

[112] CJIB, schriftelijke aanvullingen op de interviews 30-1-2020.
[113] Zie ook overweging 26 AVG. Zie hierover Voigt & Von dem Bussche 2017, p. 13.

die vervolgens al dan niet wordt opgebeld. Het verwerken van persoonsgegevens bij het gevolg geven aan de uitkomsten van de beslisboom valt daarmee onder het toepassingsbereik van de AVG.[114]

Voor een goed begrip van de pilot zijn twee privacyaspecten van belang: profilering en geautomatiseerde besluitvorming. De pilot Telefonisch Innen geeft een voorspelling van toekomstig gedrag van een individu op basis van historische data van soortgelijke gevallen. Het gaat daarmee om profilering in de zin van art. 4 aanhef en lid 4 AVG.[115] Profilering wordt vaak in één adem genoemd met het verbod op geautomatiseerde besluitvorming van art. 22 AVG. Dit verbod kan inderdaad een rol spelen bij profilering, maar waarschijnlijk niet in een situatie zoals in deze casestudy. Geautomatiseerde besluitvorming wordt in art. 22 AVG gedefinieerd als 'een uitsluitend op geautomatiseerde verwerking, waaronder profilering, gebaseerd besluit waaraan voor [de betrokkene] rechtsgevolgen zijn verbonden of dat hem anderszins in aanmerkelijke mate treft'. In deze casestudy is de vraag of er sprake is van een besluit dat een burger 'in aanmerkelijke mate treft'. Aan het al dan niet ontvangen van een telefoontje van een overheidsambtenaar zijn immers geen rechtsgevolgen verbonden voor de burger (par. 7.2.1).[116]

7.5.2 Kansen en risico's in relatie tot het juridisch kader

Pilot Telefonisch Innen

Het CJIB geeft aan zelf niet te beschikken over alle benodigde telefoonnummers van burgers. Daarom wordt een deel van de telefoonnummers ingekocht via een private partij. Hiervoor moeten persoonsgegevens waarover het CJIB beschikt worden verstrekt aan deze partij. Hierin schuilt echter ook een risico, omdat het deels gevoelige persoonsgegevens betreft. Het CJIB heeft een handhavende taak. Als het CJIB bij een private partij een telefoonnummer van een specifieke burger opvraagt, zou deze partij daaruit kunnen afleiden dat deze burger te maken heeft gehad met handhaving door de overheid.[117] De private partij zou dergelijke gevoelige informatie kunnen gebruiken voor andere doeleinden dan met de samenwerking wordt beoogd.

Zo zou een private partij informatie kunnen doorverkopen aan derden. Dit zou strijd kunnen opleveren met onder andere het principe van een rechtmatige gegevensverwerking.[118] Op basis van dit principe kan de verwerking van persoonsgegevens enkel plaatsvinden, indien er voor de specifieke verwerking een deugdelijke rechtsgrondslag aanwezig is.[119] De gedachte hierachter is

[114] Het CJIB heeft een publiekrechtelijke taak als rechtsgrondslag om persoonsgegevens te verwerken (art. 6 lid 1 onder e AVG), namelijk de ondersteuning van de overheidstaken van (onder andere) het Openbaar Ministerie. Deze taak vloeit onder meer voort uit art. 63d Organisatiebesluit Ministerie van Justitie en Veiligheid.
[115] Schermer, Hagenauw & Falot 2018, p. 82. Groep gegevensbescherming artikel 29 2018, p. 7.
[116] Zie Groep gegevensbescherming artikel 29 2018. De huidige casus sluit niet aan bij de voorbeelden die hier worden gegeven van 'aanmerkelijke mate'. Zie verder Van Breda, *Computerrecht* 2017, p. 223-229.
[117] Interview ministerie van JenV 5-12-2019.
[118] Art. 6 lid 1 AVG.
[119] Kranenborg & Verhey 2018, p. 141.

het beschermen van de autonomie van het individu (zie verder par. 3.1). Met dergelijke risico's moet de overheid rekening houden bij het samenwerken met private partijen.[120] Het CJIB geeft aan, dat hierover door het CJIB en de betreffende partij afspraken zijn gemaakt.[121] De evaluatie van deze afspraken valt buiten de reikwijdte van deze casestudy (par. 7.1.2).

Het actief en persoonlijk benaderen van burgers, zoals bepleit in het WRR-rapport, kan door de betreffende burger als onwenselijk worden ervaren.[122] Als een burger wordt gebeld door een overheidsinstantie, zoals in de pilot, bestaat het risico dat deze burger dit beschouwt als inbreuk op zijn of haar recht op privacy. Niet elke burger stelt het immers op prijs dat de overheid contact opneemt zonder dat hij of zij om hulp heeft gevraagd.[123] Het gaat dan niet alleen om de juridische rechtmatigheid van overheidshandelen, maar ook om de vraag, welke mate van inmenging in het privéleven van burgers wenselijk is. Een verregaande inmenging kan – ook als deze rechtmatig is – door een burger als negatief worden ervaren. Dit zou kunnen leiden tot een gegronde klacht bij de ombudsman (par. 7.3)

De inzet van zelflerende algoritmen voor het signaleren van niet-zelfredzame burgers
Het delen van meer persoonsgegevens van burgers maakt het mogelijk om persoonsgerichter te werken in de uitvoering van overheidsbeleid. Uit onderzoek van onderzoeksbureaus Decisio en DSP blijkt dat intensievere uitwisseling van gegevens tussen diverse publieke en private inningsinstanties (zoals gerechtsdeurwaarders en de belastingdienst) ervoor kan zorgen dat de overheid persoonsgerichter geldschulden kan invorderen. Deze bevindingen gelden ook voor het signaleren van burgers die wel willen betalen maar dat (om uiteenlopende redenen) niet doen. [124]

Deze kans om meer maatwerk toe te passen stuit echter op grenzen, door wettelijke beperkingen aan de uitwisseling tussen gegevens.[125] Eén van deze grenzen is de grondslag voor de uitwisseling van de persoonsgegevens. Indien de betrokkene geen toestemming heeft gegeven voor de uitwisseling van persoonsgegevens, kan de uitwisseling toch plaatsvinden op grond van een wettelijke verplichting of een taak van algemeen belang.[126] Lastig is daarbij, dat de omschrijving van publiekrechtelijke taken in wet- en regelgeving niet altijd voldoende basis biedt voor het uitwisselen van gegevens.[127]

[120] Ministerie van JenV, Uitvoeringstoets 2019, p. 19.
[121] CJIB, schriftelijke aanvullingen op de interviews van 30-1-2020.
[122] WRR 2017, p. 134.
[123] Interview ministerie SZW 15-7-2019, interview VNG 29-8-2019.
[124] Rapport Decisio & DSP 2019.
[125] Rapport Decisio & DSP 2019, p. 39.
[126] Met de term 'wettelijk' ziet de AVG niet op een wet in formele zin maar in materiële zin. Zie Voigt & Von dem Bussche 2017, p. 108.
[127] Interview VNG 29-8-2019. Zie hierover bijvoorbeeld ook de Memorie van Toelichting bij de aanpassing van de Wet gemeentelijke schuldhulpverlening: Kamerstukken *II* 2019/20, 35 316, nr. 3.

In de kabinetsreactie wordt de mogelijkheid besproken dat de overheid in massale besluitvormingsprocessen middels zelflerende algoritmen burgers identificeert die niet voldoende zelfredzaam zijn. Een belangrijk potentieel risico hierbij is dat dit te ver ingrijpt in het privéleven van burgers. Het gaat dan om de fundamentele vraag welke positie de overheid in Nederland moet willen innemen.[128] Hier speelt een dilemma tussen enerzijds een ingrijpende interventie in de privésfeer van burgers door het grootschalig delen en analyseren van (gevoelige) persoonsgegevens en anderzijds het niet tijdig bieden van de nodige ondersteuning aan kwetsbare burgers. Het kan in het belang van burgers zijn om bepaalde informatie juist wél mee te nemen, omdat dan in een individueel geval een betere uitkomst kan worden bewerkstelligd. In het huidige en toekomstige juridisch kader zullen beide belangen in evenwicht dienen te zijn.

7.5.3 Tussenconclusie

De pilot Telefonisch Innen bevestigt dat privacyregelgeving beperkingen stelt aan het middels algoritmen differentiëren tussen groepen burgers door de overheid. Meer algemeen kunnen overheden een spanningsveld ervaren tussen gegevensbescherming enerzijds en het niet tijdig bieden van de nodige ondersteuning aan minder redzame burgers anderzijds. Aanpassing van wet- en regelgeving (bijvoorbeeld het opnemen van een grondslag voor de gegevensverwerking) kan dit deels ondervangen.

Daarnaast bestaat bij het actief, persoonlijk benaderen van burgers het inherente risico dat een burger die ongevraagd wordt benaderd door de overheid dit beschouwt als inbreuk op zijn of haar privacy. Hier kan zich een dilemma voordoen: de keuze voor ‘paternalistisch’ optreden in het belang van de minder redzame burger versus terughoudend overheidsoptreden, uit respect voor de burger die geen inmenging in zijn privéleven wenst.

7.6 Conclusie

In het rapport ‘Weten is nog geen doen’ constateert de WRR dat niet alleen de cognitieve vermogens van burgers, maar ook hun niet-cognitieve vermogens (doenvermogen) begrensd zijn. Burgers verschillen in de mate waarin zij beschikken over doenvermogen. Zowel laag- als hoogopgeleiden kunnen moeite hebben om hun zaken goed te regelen. Ook is het zo dat burgers die normaal gesproken goed in staat zijn om zichzelf te redden door stress en mentale belasting over verminderd doenvermogen beschikken. Het doenvermogen van een individu is van invloed op diens zelfredzaamheid. Met zelfredzaamheid doelt de WRR op de mate waarin een burger beschikt over vermogens om zijn doelen te bereiken en zich te redden in het leven. Burgers met verminderd doenvermogen zijn hier onvoldoende toe in staat (par. 7.1.1).

[128] Interview VNG 29-8-2019.

In de kabinetsreactie op dit WRR-rapport is aangegeven dat zal worden onderzocht welke kansen algoritmen kunnen bieden voor een tijdige signalering van mensen die (tijdelijk) niet zelfredzaam zijn, met als doel maatwerk binnen massale besluitvormingsprocessen mogelijk te maken (par. 7.1.2). Deze kabinetsreactie is de aanleiding geweest voor deze casestudy.

De onderzoeksopdracht voor de casestudy's in onderhavig onderzoek vereist het signaleren van juridische kansen en risico's op drie publieke waarden, alsook het signaleren van de socio-technische ontwikkelingen in de casus en het analyseren van de verhouding van de kansen en risico's tot het juridisch kader (hoofdstuk 1). Daarbij beperkt deze casestudy zich tot de drie algemene publieke waarden die in hoofdstuk 1 zijn geformuleerd: rechtsbescherming, non-discriminatie en de bescherming van persoonsgegevens.

Als casestudy is gekozen voor de pilot Telefonisch Innen van het CJIB. Voor de verdere onderbouwing van de gekozen casestudy en de methodologie wordt verwezen naar par. 7.1.3-7.1.4. De pilot houdt op hoofdlijnen het volgende in (zie verder par. 7.2). Middels een algoritme selecteert het CJIB burgers met een onbetaalde verkeersboete. Het algoritme wordt ingezet met als doel om burgers te identificeren die over voldoende geld beschikken om het boetebedrag (op termijn) af te betalen, maar gedrag vertonen dat kan wijzen op verminderde zelfredzaamheid. Deze burgers worden vervolgens actief telefonisch benaderd door een medewerker van het CJIB, waarbij ze worden herinnerd aan het openstaande bedrag.

In voorgaande paragrafen is de pilot gebruikt om potentiële kansen en risico's op de drie publieke waarden te signaleren. Daarbij is tevens de verhouding van de kansen en risico's tot het juridisch kader beschreven. In de voorgaande paragrafen zijn daarnaast ook algemene kansen en risico's gesignaleerd die zich kunnen voordoen bij het middels algoritmen signaleren van niet zelfredzame burgers. Concluderend kwam hieruit het volgende naar voren.

Rechtsbescherming

Het algoritme in de pilot Telefonisch Innen is erop gericht om burgers te selecteren die in aanmerking komen voor een andere wijze van bejegening. In de juridische positie van de burger verandert niets. Bij dergelijk overheidsoptreden is een klachtprocedure bij de nationale ombudsman een voor de hand liggende vorm van rechtsbescherming. De ombudsman kan ook uit eigen beweging een onderzoek starten, als er aanwijzingen zijn dat er structureel problemen bestaan bij de bejegening van burgers door de overheid.

Het gericht benaderen van burgers, zoals dit plaatsvindt binnen de pilot Telefonisch Innen, sluit aan bij de aanbevelingen van de nationale ombudsman om, zo nodig, actief persoonlijk contact op

te nemen bij de invordering van geldschulden door de overheid. Het wijzen op de mogelijkheid van een betalingsregeling sluit aan bij de aanbeveling om maatwerk te leveren.

Ook het doen van nader onderzoek naar de risico's van het gebruik van algoritmen door de overheid valt binnen de functie van de ombudsman als bewaker van kwaliteit binnen overheidsorganisaties. Voor het succesvol vervullen van deze taak door de nationale ombudsman is algoritmische transparantie vanuit overheden van groot belang (par. 7.3).

Non-discriminatie

Het selecteren van burgers die niet zelfredzaam zijn voor een maatwerkaanpak, zoals in de pilot Telefonisch Innen, zou kunnen leiden tot het meer verwezenlijken van het ideaal van materiële gelijkheid. Omdat in de pilot onderscheid wordt gemaakt tussen burgers, speelt inherent het risico dat burgers hiertegen bezwaar zullen hebben. Eén van de mogelijkheden van rechtsbescherming is het indienen van een klacht bij de ombudsman. Deze beoordeelt dan het non-discriminatiebeginsel als onderdeel van zijn behoorlijkheidstoetsing.

Daarnaast is een aantal kansen en risico's benoemd voor het middels algoritmen signaleren van niet-zelfredzame burgers. Als achterstandsgroepen, zoals bijvoorbeeld burgers met verminderd doenvermogen, met behulp van algoritmen kunnen worden gedetecteerd kan dit bijdragen aan effectiever overheidsbeleid. Dit is dus een kans.

Een eerste risico is dat overheden zich onvoldoende bewust zijn van aanwezige verbanden in de trainingsdata van een algoritme. In databestanden kunnen factoren die wijzen op verminderd doenvermogen samenhangen met factoren als bijvoorbeeld opleidingsniveau of migratieachtergrond. Overheden dienen zich bewust te zijn van de risico's op gebied van non-discriminatie die dit met zich brengt. Risico's op gebied van discriminatie doen zich ook voor bij de toepassing van het algoritme in beleid. Degenen die door een algoritme worden geselecteerd moeten niet worden gestigmatiseerd. Voorkomen moet worden dat mensen het etiket 'verminderd doenvermogen' of 'niet-kunner' krijgen (par. 7.4).

Bescherming van persoonsgegevens

Uit de pilot Telefonisch Innen blijkt dat privacyregelgeving beperkingen stelt aan het middels algoritmen differentiëren tussen groepen burgers door de overheid. In zijn algemeenheid kunnen overheden bij het verwerken van persoonsgegevens een spanningsveld ervaren tussen privacy en het willen signaleren van achterstandsgroepen (par. 7.5). Als overheden in massale besluitvormingsprocessen met algoritmen niet-zelfredzame burgers willen identificeren om hen vervolgens actief, persoonlijk te benaderen is er een risico, dat de betreffende burgers dit zien als een inbreuk op hun privacy. De overheid ziet zich hier gesteld voor een dilemma tussen enerzijds

een ingrijpende interventie in de privésfeer van burgers door het grootschalig delen en analyseren van (gevoelige) persoonsgegevens en anderzijds het niet tijdig bieden van de nodige ondersteuning aan kwetsbare groepen. In het huidige en toekomstige juridisch kader zullen beide belangen in evenwicht dienen te zijn.

Uit bovenstaande blijkt dat op de drie besproken publieke waarden het juridisch kader grenzen stelt aan de inzet van algoritmen voor het identificeren van burgers die als gevolg van verminderd doenvermogen (tijdelijk) niet zelfredzaam zijn. Ook het ‘doenvermogen’ van de overheid is in dit opzicht begrensd.

Hoofdstuk 8. Kansen en risico's van algoritmische besluitvorming

Het gebruik van algoritmische besluitvorming kan leiden tot een grote verscheidenheid aan kansen en risico's voor de verwezenlijking van publieke waarden en belangen, zo blijkt ook uit de hierboven beschreven toepassingen in verschillende domeinen. Hoewel de kansen en risico's voor een groot deel domeinspecifiek zijn, is het desalniettemin mogelijk om een aantal rode draden te identificeren. Daarbij benadrukken we dat wij, voor de onderstaande inventarisatie van de kansen en risico's, leunen op de bevindingen van de casestudy's.[1] Wij zijn ons ervan bewust dat er buiten de domeinen van deze casestudy's andere kansen en risico's bestaan ten aanzien van andere publieke waarden en belangen. Bovendien zijn we bij de inventarisatie van kansen en risico's uitgegaan van de huidige inzet van algoritmen in de verschillende bestudeerde domeinen en van een inschatting van hoe die inzet zich in de komende vijf tot tien jaar zou kunnen ontwikkelen.

8.1 Kansen

De achtergrond van de inzet van algoritmen is veelal dat algoritmen in beginsel in staat zijn om (besluitvormings)processen sneller, beter of nauwkeuriger te doorlopen. Daarmee leveren zij, kort gezegd, efficiëntiewinst op. Regelgebaseerde algoritmen zijn met name goed inzetbaar in overzichtelijke processen met duidelijke kaders. Ze kunnen gegevens op nauwkeurige wijze verwerken zonder gehinderd te worden door menselijke eigenschappen als afleiding en verveling. Regelgebaseerde algoritmen kunnen besluitvormingsprocessen ook sneller doorlopen en kosten verlagen. Zelflerende algoritmen zijn juist goed inzetbaar in minder overzichtelijke processen. Op basis van aan hen gepresenteerde data ontwikkelen dergelijke algoritmen modellen en leggen zij verbanden. Kansen die ontstaan door de inzet van zelflerende algoritmen zullen daarom veelal samenhangen met de mogelijkheid om in grote hoeveelheden gegevens verbanden te kunnen ontdekken.

De efficiëntiewinst die met de inzet van algoritmen geboekt kan worden, heeft veelal (bedrijfs)economische waarde, maar kan ook een belangrijke bijdrage leveren aan het verwezenlijken van publieke waarden. Algoritmen zijn immers in staat om informatie nauwkeuriger te verwerken, dan mensen dat kunnen. De inzet van algoritmen voor 'makkelijke' of overzichtelijke processen kan bovendien als gevolg hebben dat er meer tijd overblijft voor de inzet van mensen voor ingewikkeldere besluitvormingsprocessen waarin maatwerk is vereist. Dergelijke kansen laten zich goed illustreren aan de hand van de casestudy naar de rechtspraak.[2] Het gebruik van algoritmen in de rechtsspraak kan ertoe leiden dat rechters en hun ondersteuning minder tijd kwijt zijn aan administratieve taken of het behandelen van routinezaken, waardoor er meer tijd overblijft

[1] Zie par. 1.3.1 over de selectie van casestudy's.
[2] Zie hoofdstuk 6.

voor het behandelen van ingewikkelder zaken. Door het bevorderen van efficiëntie kan de inzet van algoritmen zo een positieve bijdrage leveren aan de toegankelijkheid van de rechtspraak en de kwaliteit van de rechtspraak.

Naast kansen die samenhangen met efficiëntiewinst kan de inzet van algoritmen ook in algemene zin een bijdrage leveren aan het ondersteunen en verwezenlijken van bredere publieke waarden als menselijke autonomie en waardigheid. Algoritmen spelen bijvoorbeeld een cruciale rol in de technologie achter veel online platformen waarop mensen informatie delen of die het mogelijk maken om bezittingen als woningen of auto's te delen. Dergelijke platformen kunnen zo mogelijk maken dat verschillende groepen mensen toegang krijgen tot een veelheid aan waardevolle informatie en diensten, waardoor hun zelfontplooiingsmogelijkheden en keuzevrijheid worden vergroot.[3] Een ander voorbeeld zijn de algoritmen in zelfrijdende auto's die (in de toekomst) kunnen bijdragen aan het vergroten van de mobiliteit van mensen.[4]

De inzet van algoritmen kan tot slot kansen creëren voor de drie algemene publieke waarden die centraal staan in dit onderzoek: bescherming van persoonsgegevens, non-discriminatie, en rechtsbescherming. Deze kansen bespreken we hierna. Daarnaast zijn er in de casestudy's verschillende kansen geïdentificeerd ten aanzien van waarden zoals de vrijheid van meningsuiting en duurzaamheid. Voor bespreking van die kansen verwijzen wij naar de casestudy's en de daarin geformuleerde conclusies.

8.1.1 Bescherming van persoonsgegevens

Kansen voor de bescherming van persoonsgegevens die samenhangen met algoritmische besluitvorming hebben wij in het kader van dit onderzoek niet kunnen vaststellen. Algoritmen kunnen het in algemene zin mogelijk maken dat gegevens worden verwerkt die anders lastig of niet verwerkt zouden kunnen worden vanwege de bescherming van persoonsgegevens. Zo kunnen *privacy enhancing technologies* (PETs) worden ingezet om het gebruik van persoonsgegevens op een geautomatiseerde wijze te minimaliseren. Daarbij kan gedacht worden aan het geautomatiseerd pseudonimiseren van gegevens, bijvoorbeeld door het verwijderen van herleidbare gegevens.[5] De inzet van dergelijke technologieën komt echter niet zozeer de bescherming van persoonsgegevens ten goede. Wel kan die inzet ervoor zorgen dat andere kansen van algoritmische besluitvorming worden gerealiseerd met inachtneming van het belang van gegevensbescherming.

[3] Zie daarover Gerards 2019.
[4] Zie par. 5.6.2.
[5] Zie daarover De la Torre 2019; Hansen, Hoepman & Jensen 2015; Hernández Encinas e.a. 2015.

8.1.2 Non-discriminatie

De inzet van algoritmen kan op verschillende manieren kansen bieden voor het verwezenlijken van het recht op non-discriminatie. Algoritmen zijn in beginsel veelal beter dan mensen in staat om besluiten te nemen zonder aanzien des persoons. Waar mensen zich kunnen laten leiden door (mogelijke vooroordelen ten aanzien van) de persoon die zij tegenover zich hebben, zijn goed-geprogrammeerde en gevalideerde algoritmen in principe objectiever en kunnen zij dus een bijdrage leveren aan het verwezenlijken van het recht op non-discriminatie.

Algoritmen zijn bovendien in staat om veel informatie te verwerken en kunnen daarmee veel individuele kenmerken van personen meenemen in een besluitvormingsproces.[6] Op die manier kunnen besluitvormingsprocessen beter worden afgestemd op de betrokken personen, waardoor de inzet van algoritmen in beginsel kan bijdragen aan het realiseren van materiële gelijkheid. Ter illustratie daarvan kan gewezen worden op het gebruik van algoritmen voor overheidsincasso bij verkeersboetes met als doel om personen te identificeren die door omstandigheden hun boetes niet betalen terwijl ze daar financieel gezien waarschijnlijk wel toe in staat zouden zijn.[7] Een mogelijk effect daarvan is dat niet altijd een deurwaarder hoeft te worden ingezet, waardoor verdere financiële problemen voorkomen kunnen worden.[8]

Algoritmes kunnen daarnaast bijdragen aan gelijkheid in de vorm van consistentie van besluitvorming. Gebleken is bijvoorbeeld dat beslissings(ondersteunings)algoritmen binnen de rechtspraak kunnen bijdragen aan consistentere rechterlijke beslissingen, bijvoorbeeld doordat dergelijke algoritmen ervoor kunnen zorgen dat rechters steeds beschikking hebben over dezelfde informatie, zoals dossiers en jurisprudentie.[9]

Tot slot kunnen algoritmen worden ingezet om discriminatie te voorkomen of te detecteren. Algoritmen kunnen bijvoorbeeld in vacatureteksten leeftijdsdiscriminatie opsporen of onderzoeken of vrouwen in een bedrijf voor hetzelfde werk minder betaald krijgen.[10] In de casestudy naar contentmoderatie komt daarnaast naar voren dat algoritmen ook gebruikt kunnen worden om discriminerende uitingen van online platformen te verwijderen.[11]

[6] Lammerant, Blok & De Hert 2018, p. 8.
[7] Zie hoofdstuk 7.
[8] Zie par. 7.4.3.
[9] Zie par. 6.2.2.
[10] Lamprecht 2019; Khademi e.a. 2019. Zie ook SAPAI 2019, p. 42.
[11] Zie par. 4.4.2.

8.1.3 Rechtsbescherming

Uit de casestudy naar de rechtspraak volgt dat de inzet van algoritmen kansen biedt voor het bieden van effectieve rechtsbescherming.[12] Die kansen hangen daar vooral samen met het feit dat algoritmen een bijdrage kunnen leveren aan de efficiëntie van processen in de rechtspraak, wat ten goede kan komen aan de gerichte inzet van rechterlijke capaciteit. Daarnaast kan hier gewezen worden op de kans die de inzet van algoritmen kan opleveren voor het anonimiseren van rechtspraak. Dat kan een bijdrage leveren aan de openbaarheid van rechtspraak als daardoor meer rechtspraak gepubliceerd kan worden. Dergelijke openbaarheid van rechtspraak kan op haar beurt ten goede komen aan de waarde van rechtsbescherming.

8.2 Risico's

Algoritmische besluitvorming biedt niet enkel kansen, maar gaat ook gepaard met risico's. Daarbij valt te denken aan abstractere risico's zoals die van de-individualisatie en verlies van autonomie en menselijk contact. Ook kan de inzet van algoritmische besluitvorming ethische vraagstukken oproepen. De risico's voor publieke waarden en belangen die wij hieronder identificeren zijn echter concreter, in die zin dat ze rechtstreekser verband houden met de drie grondrechten die in dit onderzoek centraal staan.[13]

8.2.1 Bescherming van persoonsgegevens

Voor algoritmische besluitvorming zijn grote hoeveelheden gegevens nodig. Veelal zijn dit (ook) gegevens die direct of indirect te herleiden zijn tot personen en gaat het dus om persoonsgegevens. Risico's met betrekking tot de bescherming van persoonsgegevens ontstaan met name als algoritmen worden ingezet om op grote schaal persoonsgegevens te verzamelen of anderszins te verwerken. Als (regelgebaseerde) algoritmen worden ingezet om bestaande besluitvormingsprocessen te 'codificeren', dan hoeven daarvoor in beginsel niet meer persoonsgegevens te worden verwerkt. Echter, juist doordat de inzet van algoritmen het mogelijk maakt om grote hoeveelheden gegevens te betrekken in een besluitvormingsproces, worden daarvoor mogelijk meer persoonsgegevens gebruikt dan voor de inzet van algoritmen.

Persoonsgegevens worden lang niet altijd van personen zelf verkregen, maar kunnen ook worden verzameld door algoritmen die bijvoorbeeld het gedrag van internetters volgen en analyseren, of door sensoren die in de fysieke wereld gegevens verzamelen die vervolgens door algoritmen worden verwerkt. Algoritmen kunnen daarnaast op basis van bestaande data, waaronder gegevens over personen, nieuwe verbanden vinden tussen gegevens en op die manier meer te

[12] Zie par 6.5.2.
[13] Zie par. 1.2.2 voor onze definitie van publieke waarden en belangen.

weten te komen over personen.[14] Met name zelflerende algoritmen spelen een belangrijke rol in het verwerken van data en het vinden van nieuwe informatie in (*big*) datasets.[15]

Het ongelimiteerd verzamelen, combineren, of anderszins verwerken van persoonsgegevens in of met behulp van algoritmen kan op gespannen voet staan met het belang van bescherming van persoonsgegevens.[16] Als gevolgtrekkingen door algoritmen uit (*big*) datasets echter niet blijken te kloppen, dan kan dat bovendien grote gevolgen hebben voor de identiteit en reputatie van mensen. Daarnaast is het mogelijk dat algoritmen uit grote hoeveelheden 'normale' gegevens bijzondere categorieën van persoonsgegevens afleiden.[17] Zo kan uit gegevens over het aankoopgedrag van bepaalde personen informatie afgeleid worden over hun fysieke gesteldheid.[18] Naarmate meer gegevens (openbaar) toegankelijk zijn en verwerkt kunnen worden door algoritmen, neemt bovendien de herleidbaarheid van gegevens toe.[19] Aanvankelijk geanonimiseerde of gepseudonimiseerde gegevens worden dan opnieuw gekoppeld aan een persoon.[20]

De grote schaal van gegevensverwerking kan het ook aantrekkelijk maken voor kwaadwillenden om zich toegang tot de systemen te verschaffen, waardoor ook de informatieveiligheid kan worden geraakt.[21] Als dergelijke groepen ongeoorloofde toegang krijgen tot de systemen, zullen de gevolgen daarvan bovendien groot zijn vanwege de grote hoeveelheid gegevens die daarin is opgenomen.

8.2.2 Non-discriminatie

Het recht op non-discriminatie, en in bredere zin het recht op gelijke behandeling, vereist niet alleen dat vergelijkbare gevallen een gelijke behandeling krijgen, maar ook dat onvergelijkbare gevallen verschillend behandeld worden, in overeenstemming met hun verschillen. Als een algoritme bij de categorisering van gevallen onvoldoende rekening houdt met de relevante verschillen tussen bepaalde gevallen, worden die gevallen dus mogelijk ten onrechte gelijk behandeld. De inzet van algoritmen kan daardoor negatieve gevolgen hebben voor de waarde van non-discriminatie in situaties waarin het maken van een onderscheid op basis van non-discriminatieregels juist gewenst is.

[14] Zie in dat verband ook Groep gegevensbescherming artikel 29 2013, p. 47: '*More often than not, it is not the information collected in itself that is sensitive, but rather, the inferences that are drawn from it and the way in which those inferences are drawn, that could give cause for concern.*'
[15] ICO 2017, p. 10.
[16] Zwenne & Steenbruggen 2017, p. 89. Van Helden 2020, p. 97-98. Zie in dat kader ook Rb. Den Haag 5 februari 2020, ECLI:NL:RBDHA:2020:865 (*SyRI*).
[17] Zie over bijzondere categorieën van persoonsgegevens par. 3.1.
[18] Duhigg, *The New York Times* 16 februari 2012.
[19] Kulk & Van Loenen, *International Journal of Spatial Data Infrastructures Research* 2012, p. 196-206.
[20] Ohm, *UCLA Law Review* 2010, p. 1701-1777; Narayanan & Shmatikov 2008; Sweeney 2000.
[21] Zie in dat verband ook par. 5.3.2 over de zelfrijdende auto.

Deze risico's onstaan als er sprake is van fout-positieve en fout-negatieve resultaten. Als bijvoorbeeld een algoritme op grond van analyse van betaalgegevens laat zien dat er *in het algemeen* een bereidheid is bij een bepaalde groep mensen (bijvoorbeeld mannen) om een bepaalde prijs te betalen voor een bepaald product, dan is de kans groot dat er wel degelijk mannen zijn die *niet* bereid zijn om die prijs te betalen (overinclusiviteit), en ook dat er mensen zijn die niet tot de geselecteerde groep behoren (bijvoorbeeld vrouwen) die ook bereid zijn om die prijs te betalen (onderinclusiviteit). Onder- en overinclusiviteit doen zich bij alle vormen van categorisering onvermijdelijk voor. Zij zijn in het bijzonder problematisch als categorieën te grofmazig zijn of als er geen goede rechtvaardiging bestaat voor de in- en uitsluiting.[22]

Het discriminatieverbod kan verder in het gedrang komen als vooroordelen of onaanvaardbare stereotypen via de programmeur(s) van een algoritme of via anderen die betrokken zijn bij de ontwikkeling van het algoritme (bewust of onbewust) een weerslag hebben op het algoritme.[23] Als de werking van regelgebaseerde algoritmen redelijk inzichtelijk is, dan zal de oorsprong van de discriminatie op te sporen zijn en kan het algoritme daarop worden aangepast. Naarmate de werking van regelgebaseerde algoritmen complexer is, bijvoorbeeld door de veelheid aan gebruikte variabelen en toe te passen regels of door de interactie van verschillende algoritmen, dan zal het moeilijker zijn om erachter te komen of en op welke wijze het algoritme leidt tot discriminatie.[24]

Vergelijkbare problemen ten aanzien van de invloed van vooroordelen of onaanvaardbare stereotypen kunnen zich voordoen als zelflerende algoritmen worden ingezet in besluitvormingsprocessen. Dat is allereerst het geval als de data aan de hand waarvan het algoritme wordt getraind, gevalideerd, of getest niet voldoende representatief zijn voor de groep mensen waarover wordt beslist. Dat kan er dan toe leiden dat gevallen niet goed worden onderscheiden of zelfs helemaal niet worden herkend door het algoritme. Een voorbeeld hiervan kwamen we tegen in de casestudy naar de zelfrijdende auto. Als algoritmen in zelfrijdende auto's getraind zijn met niet-representatieve data dan kan dat ertoe leiden dat een zelfrijdende auto bijvoorbeeld niet in staat is om mensen met een bepaalde huidskleur te herkennen.[25]

Een vergelijkbaar risico van discriminatie bestaat als stigmatisering, stereotypering of vooroordelen onderdeel zijn van de data waarmee een algoritme wordt getrained. Dat kan leiden tot een voortzetting en, door de grote schaal waarop algoritmen kunnen beslissen, mogelijk ook versterking van patronen van achterstelling en maatschappelijke discriminatie. Een illustratie

[22] Zie uitgebreid Gerards 2002.
[23] Zie daarover ook de door Friedman & Nissenbaum besproken *pre-existing bias* (Friedman & Nissenbaum, *ACM Transactions on Information Systems* 1996, p. 330-347).
[24] Zie par. 2.2.
[25] Zie par. 5.4.2.

hiervan is te vinden in de casestudy naar de contentmoderatie-algoritmen die gebruikt kunnen worden om *hate speech* te detecteren. Als trainingsdata vooroordelen bevat ten aanzien van het taalgebruik van bepaalde sociale groepen, dan kunnen hun uitingen sneller aangemerkt worden als *hate speech*.[26] Dergelijke effecten kunnen zich nog sterker voordoen wanneer mensen handelen op basis van de bevooroordeelde uitkomsten van een algoritme, bijvoorbeeld in sollicitatieprocedures of bij het opsporen van strafbare feiten, en de daarbij verzamelde gegevens vervolgens weer worden gebruikt om het algoritme mee te trainen. Daardoor kan een *feedback loop* ontstaan waarin het discriminerende effect na verloop van tijd verder versterkt.[27]

Verder is het denkbaar dat zelflerende algoritmen bij het leggen van verbanden in de trainingsdata aspecten betrekken waarmee in wezen een discriminerend onderscheid wordt gemaakt. Het kan daarbij gaan om zowel directe als indirecte discriminatie. Die laatste vorm van discriminatie kan met name worden veroorzaakt door het gebruik van zogenaamde proxy-informatie in datasets, dat wil zeggen: informatie die correleert met het onderdeel zijn van een bepaalde (beschermde) groep.[28] Daarbij kan het bijvoorbeeld gaan om informatie met betrekking tot iemands lengte of postcode. Dergelijke informatie kan samenhangen met bijvoorbeeld geslacht, inkomen of etnische achtergrond, terwijl die kenmerken niet relevant zijn voor de uitkomst van het algoritme en het bovendien gaat om een verdachte grond die niet in besluitvorming mag worden betrokken. Daardoor kan het algoritme leiden tot een uitkomst die in strijd is met de gelijkebehandelingswetgeving.[29]

De hiervoor besproken risico's op discriminatie als gevolg van gebrekkige trainingsdata is tot op zekere hoogte nog groter bij *unsupervised learning*- en *deep learning* algoritmen, omdat deze algoritmen zelfstandig op zoek gaan naar mogelijk relevante verbanden, zonder dat mensen tussentijds kunnen controleren of discriminerende patronen de uitvoer van het algoritme bepalen en zonder dat zij het algoritme kunnen corrigeren door training aan de hand van gelabelde data.[30] De menselijke rol bij *supervised learning*-systemen is groter, met name vanwege de labeling die plaatsheeft in het kader van de training van het algoritme. Daardoor kunnen menselijke vooroordelen tegelijkertijd weer gemakkelijker het algoritme binnensluipen dan het geval is bij *unsupervised learning*-systemen.[31]

[26] Zie par. 4.5.
[27] Gebru 2020.
[28] Barocas & Selbst, *California Law Review* 2016, p. 691.
[29] European Union Agency for Fundamental Rights 2018, p. 5.
[30] Datatilsynet 2018, p. 11; Barocas & Selbst, *California Law Review* 2016, p. 680.
[31] Mitrou 2019, p. 42 e.v. Barocas & Selbst, *California Law Review* 2016, p. 692. Het effect dat menselijke vooroordelen hebben op algoritmen is potentieel nog groter als de algoritmen worden getraind aan de hand van eerdere menselijke beslissingen. Daardoor kunnen eventuele menselijke vooroordelen zich vertalen in het systeem, met als aanvullende complicerende factor dat ze moeilijker te identificeren en te corrigeren zijn. Zie hierover: Council of Europe 2018, p. 28.

8.2.3 Rechtsbescherming

Van belang bij het recht op rechtsbescherming is dat mensen die onderwerp zijn van algoritmische besluitvorming weten dat het betreffende besluit genomen is door of met behulp van algoritmen en dat zij in staat zijn om zich te verweren als dat besluit negatieve gevolgen heeft. Daarbij geldt dat algoritmen worden ingezet in een steeds verder digitaliserende samenleving, waarin processen aan elkaar gekoppeld worden en daarmee complexer kunnen worden. Dat maakt het lastiger om processen te ontwarren en dat kan tot gevolg hebben dat ook de uitkomsten van op zichzelf simpele algoritmen minder goed uitlegbaar worden.

Regelgebaseerde algoritmen werken aan de hand van vooraf bepaalde variabelen en regels die het in beginsel mogelijk maken om te achterhalen hoe een bepaalde uitkomst tot stand is gekomen. Voor zelflerende algoritmen geldt dat het een stuk ingewikkelder is om uit te leggen hoe het algoritme tot een bepaalde uitkomst is gekomen.[32] Als wordt gewerkt met zogenaamde dynamische modellen, ontwikkelt het model zich bovendien verder gedurende de toepassing van het algoritme. Zelflerende algoritmen baseren zich bovendien veelal niet op concepten of kenmerken die voor mensen duidelijk relevant zijn en die zij gemakkelijk als zodanig kunnen herkennen. Ook dat maakt het lastig en vaak zelfs onmogelijk om de werking van het systeem uit te leggen in termen die voor mensen te begrijpen zijn.[33]

Als een individu wil ageren tegen een beslissing, maakt de gebrekkige uitlegbaarheid van het door of aan de hand van een algoritme genomen besluit het moeilijk om argumenten te kunnen aanvoeren tegen dat besluit. De gebrekkige uitlegbaarheid is met name problematisch als er sprake is van beslissingen waarvan de gebruikte argumenten doorslaggevend zijn voor de inhoud van de beslissing (zoals rechterlijke beslissingen). De gebrekkige uitlegbaarheid zal in algemene zin minder problematisch zijn als een individu niet zozeer wil ageren tegen de argumentatieve onderbouwing van de beslissing maar tegen het effect van de beslissing. Dat kan bijvoorbeeld het geval zijn als zelfrijdende auto's belissingen nemen, zoals het wisselen van baan. Als zich schade voordoet als gevolg van een dergelijke beslissing, dan is het mogelijk om daartegen op te komen, zonder dat het nodig is om argumenten aan te voeren om de beslissing van de auto te weerleggen.[34] Uiteraard speelt dan wel de vraag wie er verantwoordelijk is voor de genomen beslissing. Daarmee raken we aan een andere risico voor rechtsbescherming.

[32] Datatilsynet 2018, p. 24.
[33] Datatilsynet 2018, p. 19; Mitrou 2019, p. 58; Eiband, Schneider & Buschek 2018. Alhoewel de begrippen 'begrijpelijkheid' *(interpretability)* en 'uitlegbaarheid' *(explainability)* veelal door elkaar worden gebruikt, wordt door sommige auteurs ook op de verschillen gewezen. Begrijpelijkheid wordt dan gezien als het uitleggen van de abstracte werking van een systeem op een voor mensen te bevatten wijze, terwijl de uitlegbaarheid zich juist richt op het menselijke begrip ten aanzien van de factoren die tot een specifieke uitkomst geleid hebben. Zie daarover: Došilović, Brčić & Hlupić 2018.
[34] Zie par. 5.5.

Voor effectieve rechtsbescherming is namelijk van belang dat de persoon wiens rechten geschonden zijn, weet tot wie hij zich moet richten.[35] In de praktijk kunnen bij de inzet van algoritmen veel verschillende partijen betrokken zijn, waardoor niet in alle gevallen duidelijk is wie verantwoordelijk is en wie dus kan worden aangesproken. Als bijvoorbeeld een besluit is genomen op basis van een algoritme, rijst de vraag of de ontwerper of trainer van het algoritme moet worden aangesproken, of (ook) degene die heeft besloten het algoritme in te zetten voor besluitvorming, of (ook) degene op basis van een algoritme uiteindelijk het concrete besluit heeft genomen.

Dergelijke vragen worden prangender naarmate processen in de samenleving verder digitaliseren en algoritmen in toenemende mate met elkaar interacteren. Algoritmen zijn dan steeds minder te zien als zelfstandig opererende software en meer als onderdelen van een groter systeem.[36] Dat creëert het zogenaamde *problem of many hands*: als gevolg van de complexiteit van de gebruikte technologieën en de hoeveelheid actoren die betrokken zijn, is het vaak zeer moeilijk of zelfs onmogelijk om iemand verantwoordelijk te houden.[37] In een complexe technologische, maatschappelijke en organisatorische context is niet altijd duidelijk waar de oorzaak van een fout ligt en wie daarvoor kan worden aangesproken. Ook hiervoor kan de casestudy naar zelfrijdende auto's als illustratie dienen. In zelfrijdende auto's werken verschillende algoritmen en sensoren samen. Door incorrecte sensordata of interpretatie van die data kan schade ontstaan. Schade kan echter ook het gevolg zijn van bijvoorbeeld foutieve informatie die de auto via communicatietechnologieën krijgt aangeleverd. Als in dergelijke situaties niet duidelijk is waar de oorzaak ligt van bijvoorbeeld een ongeluk en wie daarvoor kan worden aangesproken, dan is dat problematisch met het oog op rechtsbescherming.[38]

8.3 Organisatorische en maatschappelijk en context

De hierboven geïdentificeerde kansen en risico's hangen voor een deel samen met het type algoritme dat wordt gebruikt. Verschillende typen algoritmen hebben immers hun eigen mogelijkheden en onmogelijkheden.[39] De mate waarin kansen en risico's zich voordoen en de gevolgen die zij hebben, is echter ook voor een belangrijk deel afhankelijk van het domein en de organisatorische en maatschappelijke context waarin algoritmen worden ingezet.[40]

[35] Council of Europe 2018, p. 24.
[36] Zie daarover ook par. 2.2.
[37] Van de Poel e.a., *Science and Engineering Ethics* 2012, p. 50; Nissenbaum, *Science and Engineering Ethics* 1996, p. 25-42.
[38] Zie par. 5.5.2.
[39] Zie par. 2.1 over de verschillende typen algoritmen die we onderscheiden en wat de beperkingen zijn van deze typen algoritmen.
[40] Zie par. 2.2.

Zo is het doel waarvoor een organisatie een algoritme inzet van belang voor de risico's die dat met zich brengt. In algemene zin kan worden gesteld dat het inzetten van algoritmen met een beschrijvend of diagnosticerend doel in potentie minder grote gevolgen heeft voor personen dan het inzetten van voorspellende en voorschrijvende algoritmen. Daarbij is uiteraard ook van belang hoe zelfstandig het algoritme opereert in het besluitvormingsproces. In bijvoorbeeld (volledig) zelfrijdende auto's nemen algoritmen rechtstreeks, zonder menselijke tussenkomst, een beslissing. Algoritmen kunnen echter ook worden ingezet als onderdeel in een groter besluitvormingsproces waarin mensen de uiteindelijke beslissing nemen. Dit is bijvoorbeeld het geval als het algoritme informatie geeft of adviezen verstrekt die relevant zijn voor de te nemen beslissing. Daarbij kan gedacht worden aan de beslisondersteuningsalgoritmen die in de rechtspraak kunnen worden gebruikt of aan de algoritmen die online *hate speech* signaleren.

Ook de organisatorische context waarbinnen een algoritme wordt ingezet, is van invloed op (de mate) waarin risico's zich voordoen. Als mensen die werken met de uitvoer van algoritmen niet voldoende in staat zijn om risico's te herkennen, of zij onvoldoende ruimte hebben om die risico's te voorkomen, dan vergroot dat de kans dat die risico's zich daadwerkelijk verwezenlijken. Een voorbeeld daarvan kan gevonden worden in het domein van contentmoderatie door online platformen. In dat domein hebben de mensen die oordelen over online content slechts zeer beperkt de tijd om te besluiten over door algoritmen gesuggereerde te verwijderen content, wat de correctheid van genomen beslissingen niet ten goede komt.[41] Ook kan gedacht worden aan rechters die door algoritmen gegenereerde adviezen op waarde moeten schatten, terwijl zij niet altijd voldoende weten over de manier waarop het systeem tot het betreffende advies gekomen is.[42]

De mate waarin risico's zich verwezenlijken is ook afhankelijk van de mate waarin een organisatie invloed kan uitoefenen op de ontwikkeling en werking van een gebruikt algoritme. Algoritmen kunnen door een organisatie zelf worden ontwikkeld, maar ook door derden, of in het kader van een samenwerkingsverband. Gebruikte datasets kunnen uit de eigen organisatie of onderneming afkomstig zijn of extern worden ingekocht. Als data en/of algoritmen extern worden ontwikkeld dan kan het voor een organisatie lastiger zijn om bijvoorbeeld het risico op discriminatie te identificeren en te voorkomen. Voor de bescherming van persoonsgegevens is daarnaast van belang zijn hoe data worden opgeslagen en in hoeverre er wordt voortgebouwd op bestaande computercode en (cloud)platformen. Een organisatie kan dan ook afhankelijk worden van anderen voor de (correcte) werking van het algoritme. Uit de casestudy naar de rechtspraak volgt bijvoorbeeld dat de

[41] Zie par. 4.2.2.
[42] Zie par. 6.2.2.

onafhankelijkheid van de rechterlijke macht in het gedrang kan komen als gebruikgemaakt wordt van externe datasets.[43]

Tot slot kan voor wat betreft de risico's van algoritmische besluitvorming, een onderscheid gemaakt worden naar het type organisatie dat die algoritmen inzet. Voor overheidsorganisaties geldt dat zij burgers eenzijdig kunnen binden en burgers geen directe invloed kunnen uitoefenen op de inzet van algoritmen in besluitvormingsprocessen. Burgers worden dus per definitie geraakt door de risico's die samenhangen met de inzet van algoritmen. Dat is formeel gezien anders als actoren in de private sector algoritmen inzetten, omdat iedere burger dan keuzevrijheid heeft ten aanzien van het gebruikmaken van bepaalde dienstverlening. Hierbij geldt wel dat de daadwerkelijke keuzevrijheid beperkt kan zijn door omstandigheden als (markt)macht van een onderneming,[44] het belang van de dienstverlening voor de burger, en de wijze waarop de markt als geheel gebruikmaakt van (bepaalde) algoritmen in besluitvormingsprocessen.

8.4 Conclusie

In dit hoofdstuk hebben wij een overzicht gegeven van de kansen en risico's die kunnen ontstaan als gevolg van de inzet van algoritmische besluitvorming. Kansen hangen vooral samen met de grote hoeveelheden gegevens die door algoritmen snel en nauwkeurig verwerkt kunnen worden. De inzet van algoritmen leidt dan ook tot efficiëntiewinst. Daarnaast kunnen er kansen zijn voor de verwezenlijking van de publieke waarden en belangen die in dit onderzoek centraal staan, maar zoals hiervoor is gebleken, zijn die kansen relatief abstract. De risico's voor het verwezenlijken van deze publieke waarden en belangen zijn veel concreter.

Wel geldt dat bij de inzet van een algoritme zowel kansen als risico's kunnen bestaan. Om de kansen van algoritmische besluitvorming te realiseren zullen tegelijkertijd ook andere betrokken publieke waarden en belangen geborgd moeten worden. Dat vergt veelal een integrale (beleids)afweging in het licht van onder meer de publieke waarden en belangen die spelen in een bepaald domein.

[43] Zie par. 6.5.1.3.
[44] Zie in dat verband ook par 4.6.2.

Hoofdstuk 9. Bestendigheid van de juridische kaders

In dit hoofdstuk gaan wij in op de vraag in hoeverre de bestaande juridische kaders toereikend zijn om de in hoofdstuk 8 geïdentificeerde kansen en risico's voor publieke waarden die samenhangen met algoritmische besluitvorming te realiseren, respectievelijk te vermijden of te mitigeren. Daartoe brengen we in kaart in hoeverre bestaande wet- en regelgeving de onderzochte publieke waarden en belangen borgen en inventariseren we waar regels op dit moment mogelijk in de weg staan aan het verwezenlijken dan wel vermijden of mitigeren van de geïdentificeerde kansen en risico's.

9.1 Algemene en specifieke juridische kaders

Voor het bepalen van het toepasselijke juridisch kader moet steeds een onderscheid worden gemaakt tussen algemene juridische kaders en specifieke juridische kaders die voor een bepaald domein relevant zijn. De publieke waarden en belangen die centraal staan in dit onderzoek worden in ieder geval beschermd door algemene kaders zoals gelijkebehandelingswetgeving, de AVG, het algemene (civiele) aansprakelijkheidsrecht en de Awb. Deze kaders zijn uiteraard niet altijd van toepassing,[45] maar zij hebben wel een breed toepassingsbereik. Individuele domeinen worden daarnaast door specifieke regelgeving gereguleerd. Dit geldt bijvoorbeeld voor de rechtelijke macht (Wet RO), het verkeer (WVW) of het modereren van content (*Code of Conduct on Countering Illegal Hate Speech*). Op een concreet geval van algoritmische besluitvorming zal doorgaans dan ook steeds een andere combinatie van algemene en specifieke juridische kaders van toepassing zijn.

9.2 Bestendigheid specifieke juridische kaders

De bestendigheid van de in dit onderzoek betrokken specifieke juridische kaders is in de casestudy's onderzocht. Voor een uitvoerige analyse van de bestendigheid van die kaders wordt daarom naar de casestudy's terugverwezen.

In algemene zin volgt uit de casestudy's dat de daarin bestudeerde specifieke juridische kaders niet direct in de weg lijken te staan aan het realiseren van kansen ten aanzien van de in dit onderzoek betrokken publieke waarden en belangen. Wel kan het zo zijn dat de juridische kaders onvoldoende voedingsbodem bieden om te kunnen profiteren van de inzet van algoritmen – de kaders bevorderen de inzet van algoritmen dan niet. In de rechtspraak staat de inzet van algoritmen nog in de kinderschoenen en dat is deels toe te schrijven aan het beperkte digitale karakter van de huidige rechtspraak. Het specifieke juridische kader laat bijvoorbeeld ruimte voor

[45] Zo is de Awb alleen van toepassing op bestuursorganen die besluiten nemen.

analoge procesvoering, wat de digitalisering van rechtspraak, en daarmee de mogelijkheden voor het inzetten van algoritmen die de rechtsbescherming kunnen bevorderen, niet ten goede komt.[46]

Uit de casestudy's volgt daarnaast dat de bestudeerde specifieke juridische kaders veelal voldoende ruimte bieden om geïdentificeerde risico's van algoritmische besluitvorming te vermijden of te mitigeren. Daartoe is dan wel vaak vereist dat de bestaande ruim geformuleerde normen worden toegesneden op de inzet van algoritmen met inachtneming van de daarbij relevante waarden en belangen. Uit onze inventarisatie blijkt echter ook dat (gedeelten) van specifieke juridische kaders in sommige gevallen tekortschieten. Zo bieden de toepaselijke *Code of Conduct* en de Aanbeveling van de Europese Commissie in het domein van online contentmoderatie onvoldoende bescherming tegen de mogelijke discriminerende effecten van het gebruik van algoritmen door platformen.[47] En ten aanzien van zelfrijdende auto's geldt bijvoorbeeld dat de bewijspositie van slachtoffers in het productaansprakelijkheidsrecht onvoldoende recht lijkt te doen aan de complexiteit en ondoorzichtigheid van (de werking van) de algoritmen in zelfrijdende auto's.[48] Aanpassing of aanvulling van de daar toepasselijke juridische kaders lijkt aangewezen om risico's ten aanzien van de bescherming van publieke waarden en belangen te vermijden of te mitigeren.

9.3 Bestendigheid algemene juridische kaders

Naast de specifieke juridische kaders zijn er algemene juridische kaders, die het specifieke domein waarin algoritmen worden ingezet, overstijgen. Deze kaders zijn ook in de casestudy's aan de orde gekomen. Uit het geheel van de casestudy's onstaat het volgende beeld ten aanzien van de bestendigheid van deze algemene juridische kaders in het licht van de door ons geïdentificeerde kansen en risico's voor de algemene publieke waarden.

9.2.1 Bescherming van persoonsgegevens

Het belangrijkste instrument voor de bescherming van persoonsgegevens zijn de AVG en de Nederlandse invulling daarvan in de UAVG. Deze instrumenten reguleren de verwerking van persoonsgegevens in ruime zin; dat wil zeggen: alle handelingen die worden verricht met gegevens die direct of indirect te herleiden zijn tot personen.[49] Ook de verwerking van persoonsgegevens in of door algoritmen heeft te gelden als een verwerking van persoonsgegevens in de zin van de AVG.[50] Voor de toepasselijkheid van de AVG is evenmin relevant of persoonsgegevens worden verwerkt door de overheid of door een private partij. Wel

[46] Tegelijkertijd geldt dat het ontbreken van specifieke (digitale) vormvereisten goed te verklaren kan zijn in het licht van de toegankelijkheid van de rechtspraak.
[47] Zie par. 4.5.2.
[48] Zie par. 5.5.2.
[49] Artt. 1 jo. 4 AVG.
[50] Zie daarover ook overweging 15 AVG.

kunnen voor overheidsinstanties andere en vooral strengere regels gelden.[51] Door het ruime toepassingsbereik van de AVG dienen de gegevensbeschermingsrechtelijke regels te worden nageleefd in alle fasen van de levenscyclus van een algoritme waarin persoongegevens worden verwerkt. Zo heeft de AVG ook betrekking op het gebruik van persoonsgevens om zelflerende algoritmen te trainen, te valideren en te testen.

Art. 5 AVG bevat beginselen die handvatten bieden om de geïdenficeerde risico's met betrekking tot gegevensbescherming het hoofd te kunnen bieden. Het beginsel van dataminimalisatie vergt bijvoorbeeld dat niet méér gegevevens worden verzameld dan nodig is voor het doel van de verwerking en het beginsel van doelbinding vereist dat persoonsgegevens alleen worden verzameld voor *welbepaalde* doeleinden.[52] Deze beginselen bieden tezamen handvatten om ongelimiteerde verwerking van persoonsgegevens door of met behulp van algoritmen aan banden te leggen. Ook van belang in dit verband is het transparantiebeginsel, dat bepaalt dat toegankelijke en begrijpelijke informatie beschikbaar moet zijn met betrekking tot de verwerking. Dit beginsel geeft uitdrukking aan het belang van betrokkenen om te weten dat hun gegevens door middel van algoritmen worden verwerkt en hoe dat gebeurt.

Als algoritmen onjuiste vaststellingen doen over personen dan is voor effectieve bescherming van persoonsgegevens niet alleen van belang om grip te hebben op welke data worden verzameld, maar ook op de kwalificaties die daaraan met behulp van algoritmen worden gegeven en de mate waarin die kwalificaties corrigeerbaar zijn. Het juistheidsbeginsel en de daaraan gerelateerde rechten op rectificatie en gevenswissing bieden daarvoor reeds handvatten.

De AVG bevat naast de hierboven bedoelde beginselen een groot aantal nadere regels waaraan door verwerkingsverantwoordelijken en verwerkers van persoonsgegevens moet worden voldaan.[53] Het voert te ver om de normen in de AVG (en de UAVG) hier uitputtend te behandelen. Wel wijzen we op een aantal algemene verplichtingen die relevant zijn in het licht van algoritmische besluitvorming. Zo zijn verwerkingsverantwoordelijken en verwerkers van persoonsgegevens verplicht om technische en organisatorische maatregelen te treffen om te voldoen aan de AVG.[54] Bij de inrichting van werkprocessen en systemen moet ook, voorzover dat redelijkerwijze verlangd kan worden, rekening worden gehouden met de bescherming van persoonsgegevens (*privacy by design*).[55] Ook de verplichting voor verwerkingsverantwoordelijken om in bepaalde gevallen,

[51] Zie par. 3.1, voetnoot 11.
[52] Art. 5 lid 1 onder respectievelijk c en b AVG.
[53] De AVG maakt onderscheid tussen de 'gegevensverantwoordelijke' – degene die het doel van en de middelen voor de verwerking van vaststelt – en de 'verwerker' – degene die voor de gegevensverantwoordelijke de gegevens verwerkt (zie art. 4 leden 7 en 8 AVG). De regels waaraan een persoon of organisatie zich moet houden is afhankelijk van de vraag of men een verwerker of een gegevensverantwoordelijke is. Zie daarover ook hoofdstuk 4 van de AVG. Dit onderscheid blijft in dit onderzoek verder buiten beschouwing.
[54] Art. 24 AVG.
[55] Art. 25 AVG.

waaronder ook gevallen waarin persoonsgegevens geautomatiseerd worden verwerkt, een gegevensbeschermingseffectbeoordeling uit te voeren, draagt eraan bij dat vooraf goed wordt nagedacht over de risico's en over de wijze waarop die risico's voorkomen of gemitigeerd kunnen worden als algoritmen worden gebruikt in besluitvormingsprocessen waarin persoonsgegevens worden betrokken.[56] In bepaalde gevallen, zoals bij verwerking van persoonsgegevens door overheidsinstellingen, dient een functionaris voor gegevensbescherming te worden aangesteld. Deze moet binnen de organisatie informatie verstrekken en adviseren over de verplichtingen die voortvloeien uit de AVG en tevens toezien op de naleving van de AVG.[57]

Art. 22 AVG bevat een bijzondere bepaling ten aanzien van volledig geautomatiseerde individuele besluitvorming. Dit artikel behelst, behoudens een aantal uitzonderingen, een recht om niet onderworpen te worden aan besluiten met rechtsgevolgen of anderszins aanmerkelijke gevolgen, die gebaseerd zijn op volledig geautomatiseerde besluitvorming of op profilering. Daarnaast zijn er in de AVG specifieke regels opgenomen met betrekking tot de inzichtelijkheid van dergelijke besluitvorming.[58] Deze normen zijn niet alleen van belang voor de bescherming van persoonsgegevens, maar zij spelen ook een belangrijke rol in het bieden van rechtsbescherming. Zij worden daarom in dat kader verder besproken. [59]

Om in concrete individuele gevallen voldoende bescherming te kunnen bieden tegen de gegevensbeschermingsrechtelijke risico's van algoritmische besluitvorming is vereist dat aan bovengenoemde beginselen en nadere regels een algoritmespecifieke invulling wordt gegeven. Nu ook de AVG rechten toekent en verplichtingen veelal op hoofdlijnen vaststelt, kan er onduidelijkheid bestaan over de vraag wanneer deze rechten en verplichtingen gelden en wanneer daaraan is voldaan. Als rechtsontwikkeling uitblijft, worden risico's mogelijk onvoldoende voorkomen of gemitigeerd.

De bescherming van persoonsgegevens staat of valt bovendien bij de naleving van de regels in de AVG en het toezicht dat daarop wordt gehouden. Het uitgangspunt dat verwerkings-verantwoordelijkheden passende technische en organisatorische maatregelen moeten nemen om te waarborgen dat zij de AVG naleven, draagt idealiter bij aan de bescherming van persoonsgevens.[60] Ook hier geldt dat dan wel duidelijk moet zijn welke maatregelen genomen moeten worden als sprake is van algoritmische besluitvorming. Het transparantiebeginsel en de daarmee samenhangende rechten en verplichtingen zorgen ervoor dat voor betrokkenen zichtbaar wordt hoe hun persoonsgegevens worden verwerkt in algoritmische systemen. Dat draagt er ook

[56] Zie daarover Hoofdstuk IV, afdeling 3 AVG. Zie Goodman 2016 over de potentie van de gegevensbeschermingseffectbeoordeling in het kader van algoritmische besluitvorming.
[57] Zie daarover Hoofdstuk IV, afdeling 4 AVG.
[58] Art. 13 lid 2 onder f, art. 14 lid 2 onder g en art. 15 lid 1 onder h AVG.
[59] Zie par. 9.2.3.
[60] Zie daarover Hoofdstuk IV, afdeling 1 AVG.

aan bij dat betrokkenen inbreuken op hun rechten kunnen herkennen en aan de kaak kunnen stellen.[61] Met betrekking tot het toezicht dat wordt gehouden is tot slot nog noemenswaardig dat de Autoriteit Persoonsgevens (AP) heeft aangekondigd dat zij haar toezicht de komende jaren nadrukkelijk zal toespitsen op de inzet van algoritmen en kunstmatige intelligentie.[62]

9.2.2 Non-discriminatie

Het verbod op discriminatie is neergelegd in verschillende grondrechtelijke kenbronnen, waaronder de Grondwet, het EVRM en het Handvest. Daarbij geldt dat de relevante bepalingen voor een belangrijk deel enkel van toepassing zijn in verticale rechtsverhoudingen, dat wil zeggen: in verhoudingen tussen de burger en de overheid. Via het leerstuk van positieve verplichtingen en door de invulling van open (privaatrechtelijke) normen kunnen de bovengenoemde instrumenten echter ook doorwerken in verhoudingen tussen burgers. Een aantal meer specifieke instrumenten zoals de AWGB en de EU-wetgeving over gelijke behandeling geeft bovendien invulling aan het recht op non-discriminatie in horizontale verhoudingen. Deze instrumenten geven wel slechts regels die van toepassing zijn in specifieke rechtsrelaties of specifieke domeinen. Voor zover algoritmische discriminatie zich buiten deze domeinen voordoet, schieten deze instrumenten tekort, al kan een algemeen vangnet worden gevonden in de algemene gelijkheidsnormen zoals die hierboven zijn genoemd.

Als een slachtoffer aannemelijk maakt dat er sprake is van ongelijke behandeling (of juist gelijke behandeling) door middel van een algoritme, dan is het aan de organisatie die algoritmen inzet om aan te tonen dat er geen sprake is van discriminatie op verboden gronden of van ongerechtvaardige gelijke of ongelijke behandeling. Organisaties doen er daarom goed aan om zich in alle levensfasen van het algoritme bewust te zijn van de impact van de gemaakte keuzes op de verwezenlijking van de waarde van non-discriminatie en de mogelijke risico's te beperken.[63]

Ook de AVG komt voor de verwezenlijking van de waarde van non-discriminatie belangrijke betekenis toe. De AVG is in het bijzonder relevant vanwege het verbod op verwerking van bijzondere categorieën persoonsgegevens, zoals gegevens met betrekking tot afkomst, seksuele geaardheid of geslacht. Het zijn immers veelal deze gegevens waarvan de verwerking risico's op discriminatie meebrengt. Verwerkingsverantwoordelijken dienen ook maatregelen te treffen om ongerechtvaardigd onderscheid te voorkomen.[64] In geval van profiling dienen daartoe passende 'wiskundige en statistische procedures' te worden gebruikt.[65] Als een bepaalde verwerking een

[61] Zie voor het transparantiebeginsel art. 5 lid 1 onder a AVG. Zie voor rechten en verplichtingen met betrekking tot transparantie afdelingen 2 en 3 van Hoofdstuk 3 AVG.
[62] Autoriteit Persoonsgegevens 2019.
[63] Lammerant, Blok & De Hert 2018.
[64] Overweging 71 AVG. Zie ook Datatilsynet 2018, p. 16.
[65] Overweging 71 AVG.

hoog risico op discriminatie inhoudt, dient, zeker als bij die verwerking gebruikgemaakt wordt van algoritmen, een gegevensbeschermingseffectbeoordeling te worden uitgevoerd.[66]

De bescherming die in de AVG wordt geboden aan bijzondere categorieën van persoonsgegevens kan echter ook in de weg staan aan de kansen die algoritmen bieden om discriminatie te voorkomen. Het 'in beginsel'-verbod op verwerking van bijzondere categorieën persoonsgegevens maakt het bijvoorbeeld in veel gevallen onmogelijk om gegevens te verwerken met betrekking tot etnische afkomst, seksuele gerichtheid en andere gevoelige persoonskenmerken. Om discriminatie op grond van een van deze kenmerken te voorkomen is het soms juist nodig om dergelijke gegevens te verwerken.[67] De Minister voor Rechtsbescherming heeft in dat kader reeds opgeworpen dat het verwerken van bijzondere categorieën persoonsgegevens toegestaan zou moeten zijn als dat nodig is voor het voorkomen van onder meer discriminatie.[68]

9.2.3 Rechtsbescherming

Voor een goede rechtsbescherming bij algoritmische besluitvorming is van belang dat personen wier belangen geraakt worden toegang hebben tot de middelen die nodig zijn om hun belangen (in rechte) te verdedigen. In sommige gevallen kan een onjuist besluit van een algoritme, of een besluit dat met behulp daarvan is genomen, simpelweg worden teruggedraaid. In andere gevallen kan in het kader van rechtsbescherming een vergoeding van geleden schade aangewezen zijn. Worden algoritmen in civielrechtelijke rechtsverhoudingen ingezet, dan is onder meer art. 6:162 BW relevant, dat de grondslag biedt voor vergoeding van schade als gevolg van een onrechtmatige daad. Ook vormen van aansprakelijkheid voor personen en zaken (zoals opgenomen in afdeling 6.3.2 BW) en productaansprakelijkheid (afdeling 6.3.3 BW) kunnen in dit verband relevant zijn.[69] Gaat het om de inzet van algoritmen in bestuursrechtelijke sfeer, dan gelden hiervoor de procedures van de Awb. Voor zover schade is veroorzaakt door de inzet van een algoritme biedt de AVG de mogelijkheid om schade te vergoeden. Daarvoor is de regeling neergelegd in titel 8.4 van de Awb relevant.

Voor het waarborgen van het recht op rechtsbescherming is tevens van belang dat degene wiens belangen zijn geraakt vanwege een algoritmisch genomen besluit, weet tot wie hij zich moet wenden om zijn belangen te beschermen. In de AVG wordt de naleving van de beginselen van gegevensbescherming voor een groot deel neergelegd bij de verwerkingsverantwoordelijke.[70] De verwerkingsverantwoordelijke is een natuurlijke persoon of publieke of private rechtspersoon die,

[66] Art. 35 AVG.
[67] Lammerant, Blok & De Hert, *NTM/NJCM-bulletin* 2018, p. 10-11; Žliobaitė & Custers, *Artificial Intelligence and Law 2016*
[68] Brief van de Minister voor Rechtsbescherming van 8 oktober 2019, *Kamerstukken II* 2019/20, 26643 en 32761, nr. 641, p. 11.
[69] Deze aspecten komen met name uitdrukkelijk aan de orde in de casestudy naar de zelfrijdende auto.
[70] Artt. 24 en 82 AVG.

alleen of samen met anderen, het doel en de middelen voor de verwerking van persoonsgegevens vaststelt.[71] Tegelijkertijd kan ook de verwerker, dat wil zeggen de partij die ten behoeve van de verwerkingsverantwoordelijke persoonsgegevens verwerkt,[72] in bepaalde gevallen worden aangesproken. Dat is bijvoorbeeld het geval als hij de instructies van de verwerkingsverantwoordelijke niet nakomt of onvoldoende beveiligingsmaatregelen treft.[73] Daarnaast gelden op basis van de AVG verschillende informatieplichten die betrokkenen in staat stellen hun rechten te verwezenlijken.[74] In dit kader dienen betrokkenen onder meer op de hoogte gesteld te worden van de wijze van gegevensverwerking, de identiteit en contactgegevens van de verwerkingsverantwoordelijke, het doel van de verwerking en het type informatie dat verwerkt wordt.

Daarnaast is er ten aanzien van de rechtsbescherming in geval van algoritmische besluitvorming een rol weggelegd voor het reeds genoemde art. 22 AVG. [75] Dit artikel behelst een recht om niet te worden onderworpen aan een uitsluitend op geautomatiseerde verwerking gebaseerd besluit waaraan rechtsgevolgen zijn verbonden of dat een betrokkene anderszins in aanmerkelijke mate treft. Dat recht geldt ook ten aanzien van op basis van profilering genomen besluiten waaraan rechtsgevolgen zijn verbonden of die mensen anderszins op aanmerkelijke wijze raken. Daarbij gaat het om vormen van geautomatiseerde verwerking, waarbij aan de hand van persoonsgegevens bepaalde persoonlijke aspecten van een natuurlijke persoon worden geëvalueerd, met name met de bedoeling zijn beroepsprestaties, economische situatie, gezondheid, persoonlijke voorkeuren, interesses, betrouwbaarheid, gedrag, locatie of verplaatsingen te analyseren of te voorspellen. Art. 22 AVG heeft betrekking op besluiten waaraan voor de betrokkene rechtsgevolgen zijn verbonden of die hem anderszins in aanmerkelijke mate treffen. Hiervan kan sprake zijn als gebruikgemaakt wordt van algoritmen in een besluitvormingsproces waarin verder geen sprake is van menselijke tussenkomst. De Groep gegevensbescherming artikel 29 – de voorloper van het Europees Comité voor gegevensbescherming waarin nationale toezichthouders samenwerken – heeft gesteld dat sprake moet zijn van actieve menselijke tussenkomst om te voorkomen dat sprake is van een *uitsluitend* op geautomatiseerde verwerking gebaseerd besluit.[76] Wat precies als een relevante menselijke tussenkomst heeft te gelden, blijft echter onduidelijk en dat kan problematisch zijn gelet op het rechtsbeschermende karakter van art. 22 AVG.[77] Uit de casestudy naar contentmoderatie door online platformen blijkt dat de mensen die betrokken zijn bij het verwijderen van content die door

[71] Art. 4 onder 7 AVG.
[72] Art. 4 onder 8 AVG.
[73] Art. 82 AVG.
[74] Artt. 12-15 AVG.
[75] Schermer 2013, p. 145.
[76] Groep gegevensbescherming artikel 29 2017b. Zie over de verschillende interpretaties hiervan ook Mendoza & Bygrave 2017; Wachter, Mittelstadt & Floridi, *International Data Privacy Law* 2017, p. 76-99.
[77] Veale & Edwards, *Computer Law & Security Review* 2018, p. 398.

een algoritme als *hate speech* is geïdentificeerd, onder grote druk beslissingen moeten nemen.[78] Een nadere invulling van de vereisten in art. 22 AVG, met het oog op de kwaliteit van menselijke betrokkenheid bij algoritmische besluitvorming, zou kunnen bijdragen aan het mitigeren van risico's ten aanzien van de rechtsbescherming van individuen.

Tevens dient de verwerkingsverantwoordelijke nuttige informatie over de onderliggende logica van het gebruikte systeem te verstrekken om een behoorlijke en transparante verwerking te waarborgen.[79] In de praktijk kan het echter lastig of zelfs onmogelijk zijn om de werking van algoritmen uit te leggen op een voor mensen begrijpelijke wijze, zeker als het gaat om zelflerende algoritmen. Bovendien kan het vergroten van de uitlegbaarheid van een algoritme ten koste gaan van de nauwkeurigheid waarmee het opereert.[80] Dat roept een spanningsveld op. De AVG vergt echter niet noodzakelijkerwijs een ingewikkelde toelichting over de werking van het algoritme. Van belang is dat iemand die door algoritmische besluitvorming wordt geraakt, kan begrijpen wat de achterliggende gedachte is van het besluit en op grond van welke criteria dat besluit is genomen. Het gaat er daarbij in de kern om dat de betrokkene de redenen voor het geautomatiseerde besluit kan begrijpen.[81]

In gevallen waarin de overheid gebruikmaakt van algoritmen in een besluitvormingsproces, ook als het besluit niet *volledig* op geautomatiseerde verwerking is gebaseerd, is voor de verwezenlijking van rechtsbescherming tevens een rol weggelegd voor de Awb.[82] In veel gevallen zal de inzet van algoritmen door de overheid immers (direct of indirect) uitmonden in een besluit in de zin van de Awb. Door (een interpretatie van) de normen die in de Awb zijn neergelegd, kunnen publieke waarden en belangen voor een belangrijk deel gewaarborgd worden. Zowel het besluit als de voorbereiding daarvan dienen dan immers genomen te worden in overeenstemming met onder meer de algemene beginselen van behoorlijk bestuur. In de jurisprudentie is daarmee reeds een belangrijk begin gemaakt.[83] Zo heeft de ABRvS in het kader van het Programma aanpak stikstof (PAS) en de daarbij toegepaste algoritmegebaseerde AERIUS-calculator onder meer overwogen dat beoordelingen als gevolg van geautomatiseerde besluitvorming onvoldoende inzichtelijk en controleerbaar kunnen zijn vanwege een gebrek aan transparantie over de gemaakte keuzes en gebruikte gegevens en aannames.[84] Dit gebrek aan transparantie kan het

[78] Par. 4.2.2.
[79] Daarover art. 13 lid 2 onder f, art. 14 lid 2 onder g en art. 15 lid 1 onder h AVG.
[80] Adadi & Berrada, *IEEE Access* 2018, p. 52142 en p. 52145. Zie daarover ook par. 2.3.
[81] Groep gegevensbescherming artikel 29 2017b, p. 30.
[82] Het gaat daarbij onder meer over het zorgvuldigheidsbeginsel zoals neergelegd in afdeling 3.2 Awb en de motiveringsvereisten die volgen uit afdeling 3.7 Awb. Op grond van het zorgvuldigheidsbeginsel dat is neergelegd in afdeling 3.2 van de Awb dienen bestuursorganen in de voorbereiding van een besluit de nodige kennis over de relevante feiten en de af te wegen belangen te vergaren, en die afweging op een zorgvuldige wijze te maken. Volgens de motiveringsvereisten dienen besluiten te berusten op een deugdelijke motivering die in beginsel ook kenbaar moet zijn.
[83] Zie daarover ABRvS 17 mei 2017, ECLI:NL:RVS:2017:1259 (*Stichting Werkgroep Behoud de Peel/GS Noord Brabant)*; ABRvS 18 juli 2018, ECLI:NL:RVS:2018:2454 (*Blankenburg*); HR 17 augustus 2018, ECLI:NL:HR:2018:1316.
[84] ABRvS 17 mei 2017, ECLI:NL:RVS:2017:1259 (*Stichting Werkgroep Behoud de Peel/GS Noord Brabant*).

benutten van rechtsmiddelen volgens de ABRvS compliceren, aangezien het voor rechtszoekenden niet duidelijk is welke factoren tot een bepaalde beslissing hebben geleid, waardoor zij die beslissing niet gegrond ter discussie kunnen stellen.[85] Om deze mogelijke ondoorzichtigheid van geautomatiseerde beslissystemen te compenseren, rust volgens de ABRvS op bestuursorganen onder omstandigheden de verplichting 'om de gemaakte keuzes en de gebruikte gegevens en aannames volledig, tijdig en uit eigen beweging openbaar te maken op een passende wijze zodat deze keuzes, gegevens en aannames voor derden toegankelijk zijn'.[86] Zo zouden anderen in staat gesteld moeten worden om de gemaakte keuzes en aannames te beoordelen of te laten beoordelen en indien nodig te betwisten, en daarmee reële rechtsbescherming mogelijk te maken.[87] Ook de Hoge Raad heeft dit door de ABRvS ontwikkelde toetsingskader inmiddels in zijn rechtspraak toegepast.[88]

9.3 Conclusie

In dit hoofdstuk hebben we in kaart gebracht in hoeverre de toepasselijke juridische kaders in de weg staan aan de realisering van geïdentificeerde kansen en het voorkomen of mitigeren van geïdentificeerde risico's van algoritmische besluitvorming. We stellen vast dat de algemene juridisch kaders, zoals de AVG, de Awb en het aansprakelijkheidsrecht op hoofdlijnen voldoende in staat zijn om publieke waarden en belangen te borgen. Tegelijkertijd stellen we ook vast dat er ten aanzien van bepaalde algemene normen knelpunten kunnen bestaan ten aanzien van het voorkomen en mitigeren van risico's.

[85] ABRvS 17 mei 2017, ECLI:NL:RVS:2017:1259 (*Stichting Werkgroep Behoud de Peel/GS Noord Brabant*), r.o. 14.3.
[86] ABRvS 17 mei 2017, ECLI:NL:RVS:2017:1259 (*Stichting Werkgroep Behoud de Peel/GS Noord Brabant*), r.o. 14.4.
[87] ABRvS 17 mei 2017, ECLI:NL:RVS:2017:1259 (*Stichting Werkgroep Behoud de Peel/GS Noord Brabant*), r.o. 14.4. De ABRvS maakt in zijn uitspraak inzake de *Blankenburgtunnel* vervolgens een onderscheid tussen zogenaamde standaardinvoergegevens en 'maatwerk'invoergegevens. In die zaak waren standaardinvoergegevens gegevens afkomstig uit de kaart van de stikstofgevoelige habitattypen in Natura 2000-gebieden, uit de basiskaart met door het RIVM berekende achtergronddepositie van stikstof en van vaste rekenmodules. 'Maatwerk'invoergegevens daarentegen zijn gegevens die door de gebruiker zelf moeten worden ingevoerd. Bij een besluit moeten, aldus de ABRvS, de 'maatwerk'invoergegevens die dienen als grondslag van dat besluit uit eigen beweging op papier of op anderszins waarneembare wijze openbaar gemaakt worden. Als belanghebbenden verzoeken om (informatie over) standaardgegevens die niet in of met het besluit inzichtelijk zijn gemaakt, of om 'maatwerk'invoergegevens in de vorm zoals ingevoerd, dan dienen deze ook beschikbaar gesteld te worden. ABRvS 18 juli 2018, ECLI:NL:RVS:2018:2454 (*Blankenburg*).
[88] HR 17 augustus 2018, ECLI:NL:HR:2018:1316. Daarover ook: Rb. Amsterdam 4 juli 2019, ECLI:NL:RBAMS:2019:4799.

Hoofdstuk 10. Conclusie

In dit onderzoek stond de volgende vraag centraal:

Welke kansen en risico's doen zich voor bij algoritmische besluitvorming met betrekking tot de bescherming en realisering van publieke waarden en belangen, en zijn de bestaande juridische kaders voldoende bestendig om kansen te verwezenlijken en het intreden van geïdentificeerde risico's te voorkomen of de gevolgen daarvan te mitigeren?

Ter beantwoording van deze vraag hebben wij een algemeen literatuuronderzoek uitgevoerd en vier door het WODC gekozen casestudy's verricht op het terrein van: (1) contentmoderatie door online platformen, (2) zelfrijdende auto's, (3) de rechtspraak en (4) overheidsincasso bij verkeersboetes. Deze casestudy's geven een beeld van de manier waarop kansen en risico's voor de bescherming en realisering van publieke waarden en belangen in specifieke domeinen gestalte kunnen krijgen en hoe deze zich verhouden tot bestaande juridische kaders. Voor het definiëren van relevante publieke waarden en belangen hebben wij aansluiting gezocht bij de grondrechten. Drie algemene publieke waarden stonden in dit onderzoek centraal: de bescherming van persoonsgegevens, het recht op non-discriminatie en het recht op rechtsbescherming. In het onderzoek hebben we ons verder beperkt tot bestaande toepassingen van algoritmen, alsmede op de ontwikkelingen die in de komende vijf tot tien jaar zijn te verwachten.

10.1 Kansen en risico's

Algoritmen kunnen voor tal van besluitvormingsprocessen worden ingezet. Iedere toepassing in een specifiek domein gaat gepaard met eigen kansen en risico's. Algoritmen bieden in het algemeen kansen om processen sneller, beter of nauwkeuriger te te laten verlopen. Daarmee biedt het gebruik van algoritmen vaak (bedrijfs-)economische kansen en efficiëntievoordelen. Maar ook voor de verwezenlijking van andere publieke waarden en belangen kan de inzet van algoritmen kansen bieden. Zo kunnen algoritmen veel informatie verwerken en met veel factoren rekening houden, en daarmee in potentie beslissingen nemen of ondersteunen die beter recht doen aan iemands persoonlijke kenmerken. Daarnaast kan door de inzet van algoritmen in eenvoudige besluitvormingsprocessen, meer tijd overblijven voor menselijke beslisnemers om zich te buigen over complexere gevallen. Ook kunnen in bepaalde domeinen meer specifieke waarden en belangen gediend zijn met de inzet van algoritmen. Zo volgt uit de uitgevoerde casestudy's dat bijvoorbeeld ook de verkeersveiligheid en de vrijheid van meningsuiting bevorderd kunnen worden door de inzet van algoritmen.[89]

[89] Zie daarvoor respectievelijk de casestudy's naar zelfrijdende auto's (hoofdstuk 5) en contentmoderatie door online platformen (hoofdstuk 4).

Veelal levert de inzet van algoritmen in besluitvormingsprocessen enerzijds kansen op, maar brengt die anderzijds ook risico's met zich. Dat vergt dat steeds een afweging gemaakt wordt van de kansen en risico's die zich in een specifiek domein voordoen, met inachtneming van de waarden en belangen die in dat domein van belang zijn en de mogelijkheden om risico's te vermijden of te mitigeren. Ons onderzoek heeft zich niet gericht op hoe die afweging in concrete omstandigheden gemaakt moet worden. Het doel van dit onderzoek was om in kaart te brengen welke kansen en risico's zich in algemene zin kunnen voordoen bij algoritmische besluitvormingsprocessen en waar mogelijke knelpunten bestaan in de toepasselijke juridische kaders ten aanzien van het realiseren van die kansen en het voorkomen of mitigeren van die risico's.

Uit het onderzoek volgt dat risico's voor de bescherming van persoonsgegevens met name samenhangen met het feit dat voor de ontwikkeling en toepassing van algoritmen (grote hoeveelheden) persoonsgegevens nodig kunnen zijn. Hier speelt in het bijzonder een risico van verlies van controle van personen over hun gegevens. Daarnaast is het mogelijk om met (zelflerende) algoritmen verbanden te leggen tussen gegevens en zo meer te weten te komen over personen. Ook kunnen gegevens eenvoudiger worden herleid tot bepaalde personen, waardoor de grenzen tussen persoonsgegevens en niet-persoonsgegevens vervagen. Ook bestaat het risico dat algoritmen gevolgtrekkingen doen die niet kloppen, wat grote gevolgen kan hebben voor de identiteit en reputatie van een persoon.

De inzet van algoritmen brengt daarnaast diverse risico's mee ten aanzien van het recht op non-discriminatie. Die risico's kunnen bijvoorbeeld het gevolg zijn van het gebruik van niet-representatieve data waarmee zelflerende algoritmen worden getraind, getest en gevalideerd. Ook kan het zo zijn dat relevante persoonskenmerken niet of onvoldoende in besluitvormingsprocessen worden betrokken. Dat kan tot gevolg hebben dat groepen of individuen ten onrechte ongelijk of gelijk worden behandeld. Ook is het mogelijk dat bestaande stigma's, stereotyperingen of vooroordelen neerdalen in het algoritme, bijvoorbeeld als vooroordelen bij de ontwikkelaar van een algoritme of de organisatie die het algoritme inzet, bewust of onbewust worden doorvertaald in het algoritme, of als niet-representatieve of gekleurde data worden gebruikt om een algoritme te trainen. Als vervolgens naar de uitvoer van het algoritme wordt gehandeld en nieuw verzamelde gegevens worden gebruikt om het algoritme verder te trainen, dan kan bovendien een *feedback loop* ontstaan die het discriminerende effect versterkt.

Risico's voor het recht op rechtsbescherming kunnen ontstaan doordat de uitvoer van algoritmen niet altijd goed uitlegbaar is. Hoewel het bij regelgebaseerde systemen in beginsel veelal mogelijk is om uit te leggen hoe het systeem tot een bepaalde uitkomst is gekomen, wordt dat moeilijker als regelgebaseerde algoritmen aan elkaar worden gekoppeld en zij interacteren. De werking van

zelflerende algoritmen is inherent moeilijk uit te leggen. Zelflerende algoritmen brengen immers geen (juridisch) redengevende verbanden aan maar vinden correlaties in aangereikte data.

Ook de onduidelijkheid die door de inzet van algoritmen kan ontstaan over de verantwoordelijkheid voor het genomen besluit, kan problematisch zijn vanuit het oogpunt van rechtsbescherming. Voor effectieve rechtsbescherming is daarom van belang dat de persoon wiens rechten zijn geschonden, weet tot wie hij zich moet richten. Ook hier kunnen de steeds verder digitaliserende samenleving en daarmee samenhangende interactie van algoritmische systemen ervoor zorgen dat daarover in de toekomst grotere onduidelijkheid bestaat.

10.2 Bestendigheid juridisch kader

Het juridisch kader dat de inzet van algoritmen reguleert is, evenals de kansen en risico's, grotendeels domeinspecifiek. Bij zelfrijdende auto's speelt bijvoorbeeld het aansprakelijkheidsrecht een belangrijke rol, terwijl bij contentmoderatie ook rekening gehouden moet worden met de vrijheid van meningsuiting. Dat betekent dat bij de inzet van algoritmen niet alleen per domein of toepassing andere publieke waarden kunnen spelen (en tegen elkaar afgewogen moeten worden), maar dat daarbij ook steeds andere regelgeving geldt.

Uit het door ons uitgevoerde onderzoek volgt dat de algemene juridisch kaders, zoals de AVG, de Awb en het aansprakelijkheidsrecht op hoofdlijnen voldoende in staat zijn om publieke waarden en belangen te borgen. Deze kaders bestaan voor een belangrijk deel uit brede en open geformuleerde normen waarvan wij verwachten dat zij ontwikkelingen op het gebied van algoritmische besluitvorming kunnen absorberen. De breed en open geformuleerde normen maken het bovendien mogelijk om de juridische kaders ten aanzien van een zich snel ontwikkelende technologie geleidelijk en flexibel vorm te geven. Daarvoor is wel vereist dat de nodige rechtsontwikkeling plaatsheeft, door bijvoorbeeld rechters en toezichthouders, waarmee een op algoritmen toegesneden uitleg of interpretatie wordt gegeven aan deze algemene kaders.

Wij stellen echter wel vast dat bij bepaalde algemene normen knelpunten kunnen bestaan ten aanzien van het voorkomen en mitigeren van risico's. De vraag is bijvoorbeeld of de AVG voldoende waarborgen biedt in gevallen van geïndividualiseerde besluitvorming die niet volstrekt geautomatiseerd is, maar waarin de menselijke inbreng desalniettemin gering is. Daarnaast kunnen de strenge regels van de AVG met betrekking tot de verwerking van bijzondere categorieeën van persoonsgegevens in de weg staan aan het detecteren van discriminerende effecten van algoritmen, of zelfs verhinderen dat discrimininatie door algoritmen voorkomen wordt. Daarbij moet wel worden opgemerkt dat de Minister voor Rechtsbescherming in oktober 2019 heeft

aangegeven dat wordt overwogen om het verwerken van bijzondere categorieën persoonsgegevens toe te staan voor het voorkomen van onder meer discriminatie.

Uit de casestudy's is verder gebleken dat zich knelpunten kunnen voordoen ten aanzien van de specifieke juridische kaders die van toepassing zijn in een bepaald domein. Uit de casestudy naar het modereren van *hate speech* volgt bijvoorbeeld dat de toepasselijke zelfreguleringscodes en aanbevelingen onvoldoende ruimte bieden om in te kunnen spelen op het gevaar van discriminatie dat zich kan voordoen. En uit de casestudy naar de zelfrijdende auto blijkt dat de bewijsregels in het productaansprakelijkheidsrecht mogelijk aanpassing behoeven om rechtsbescherming van slachtoffers ten goede te komen. Nadere regelgeving kan hier aangewezen zijn.

10.3 Slot

Zoals gezegd zal ten aanzien van de eventuele regulering van algoritmische besluitvorming steeds een integrale (beleids)afweging gemaakt moeten worden, waarbij wordt geïdentificeerd welke van de relevante kansen en risico's zich voordoen en of het – in het licht van de behoeften, normen, waarden en belangen in een specifiek domein – mogelijk is om de risico's te vermijden of te mitigeren, terwijl de kansen wel kunnen worden gerealiseerd. De weging van kansen en risico's en het vinden van de balans daartussen is uiteindelijk een politiek proces.

Het juridisch kader dat daarvan het resultaat is, normeert en stuurt of en hoe algoritmen worden ingezet en heeft zo ook invloed op de mate waarin kansen en risico's worden gerealiseerd of vermeden. Regelgeving dient daarom zo te worden geformuleerd dat verantwoorde innovatie op het gebied van algoritmische besluitvorming mogelijk is. Voorkomen moet daarbij in het bijzonder worden dat normen te veel worden toegespitst op reeds bestaande technologieën en zij geen ruimte laten voor nieuwe ontwikkelingen. Dergelijke regelgeving biedt namelijk enkel rechtszekerheid en -bescherming zolang de specifieke technologie ook daadwerkelijk gereguleerd wordt door de opgestelde regels. Als de betreffende technologieën zich ontwikkelen, bestaat de mogelijkheid dat de betrokken (technologiespecifieke) regulering niet meer actueel is en zodoende geen houvast meer biedt voor de wijze waarop met nieuwe ontwikkelingen moet worden omgegaan. Bovendien kan het gebruik van technologiespecifieke regelgeving het zicht op de onderliggende uitgangspunten en beginselen ontnemen, wat uiteindelijk de rechtsontwikkeling niet ten goede komt. De publieke waarden die de wetgever probeerde te waarborgen, komen dan door de snelle ontwikkeling van de technologie weer op het spel te staan.

Het lijkt ons daarnaast, in het licht van de bevindingen van dit onderzoek, niet zinvol om het juridisch kader in te richten op algoritmische besluitvorming in algemene zin. De algemene kaders zoals die momenteel beschikbaar zijn, vertonen namelijk geen grote tekortkomingen of structurele

problemen. Integendeel: gebleken is dat ze een aanzienlijk absorberend vermogen hebben, wat betekent dat ze ook bij de ontwikkeling en inzet van nieuwe technologieën richting kunnen bieden. De casestudy's laten bovendien zien dat de kansen en risico's voor alle onderzochte publieke waarden en belangen sterk afhankelijk zijn van het domein en de organisatorische context waarin een algoritme wordt ingezet. Ook ten aanzien van de gevolgen die de inzet van algoritmen heeft voor mensen laten de toepassingen in de casestudy's zich niet eenvoudig vergelijken. De waardering van kansen en risico's van algoritmische besluitvorming is met name afhankelijk van de ernst van de gevolgen die dergelijke besluitvorming kan hebben en de mate waarin herstel mogelijk is. Zo is het onterecht bellen van een schuldenaar door een CJIB-medewerker van een andere orde dan een verkeersongeluk dat een zelfrijdende auto veroorzaakt door een verkeerde interpretatie in de sensoriek van de auto. Ook valt het modereerproces van online platformen, waarin algoritmen redengevend zijn maar niet zelf beslissen, niet te vergelijken met algoritmen die rechterlijke uitspraken automatisch kunnen anonimiseren. In alle gevallen worden de kansen en risico's die uit de inzet van algoritmen voortvloeien dus sterk gekleurd door het domein en de organisatorische context waarin de algoritmen worden ingezet. Nadere algemene regelstelling heeft voor het bestrijden van risico's voor de publieke waarden dan ook nauwelijks meerwaarde. De kansen en risico's voor publieke waarden en belangen moeten vooral in het licht van de dynamiek in een domein en de verhoudingen tussen betrokken partijen in kaart worden gebracht en gewogen. Het voorgaande pleit er volgens ons dan ook voor om, daar waar knelpunten worden ervaren, deze knelpunten zo veel mogelijk domeinspecifiek aan te pakken. Alleen op die manier kan voldoende recht worden gedaan aan de specifieke kansen en risico's voor de specifieke publieke waarden en belangen die op het spel staan.

Bijlage 1. Begeleidingscommissie

Voorzitter

prof. mr. A.R. Lodder, hoogleraar internet governance and regulation, Vrije Universiteit

Leden

D. Frijters MA, BSc, MT-lid & programma- en teammanager ECP | Platform voor de InformatieSamenleving

mr. E.C. van Ginkel, projectbegeleider, WODC

L. Kool MSc, MA, coördinator Digitale Samenleving, Rathenau Instituut

prof. mr. V. Mak, hoogleraar privaatrecht, Tilburg University

mr. S.W. Mul, raadadviseur wetgevingsbeleid, Ministerie van Justitie en Veiligheid

Bijlage 2. Klankbordgroep

Casestudy Contentmoderatie door online platformen

- Pieter van Koetsveld (Ministerie van Onderwijs, Cultuur en Wetenschap)
- Amber Mechelse (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties)
- Bastiaan Winkel-Boer (Ministerie van Justitie en Veiligheid)

Casestudy Zelfrijdende auto's

- Edwin Nas ((*tot 1 april 2020*) Ministerie van Infrastructuur en Waterstaat)

Casestudy De rechtspraak

- Twee beleidsmedewerkers van het Ministerie van Justitie en Veiligheid

Casestudy Overheidsincasso bij verkeersboetes

- Remco Boersma (Ministerie van Justitie en Veiligheid)
- Emiel Reck (Ministerie van Justitie en Veiligheid)
- Marije Zijlstra (Centraal Justitieel Incassobureau)

Bijlage 3. Lijst van deelnemers Expertmeeting 27 november 2019[1]

- Floris Bex (Tilburg University, Universiteit Utrecht)
- Jan Broersen (Universiteit Utrecht)
- Jan De Bruyne (KU Leuven)
- José van Dijck (Universiteit Utrecht)
- Hub Dohmen (Dohmen advocaten)
- Anna van Duin (Universiteit van Amsterdam)
- Mireille van Eechoud (Universiteit van Amsterdam)
- Serge Gijrath (Universiteit Leiden)
- Aviva de Groot (Tilburg University)
- Ran Haase (Vereniging Nederlandse Gemeenten)
- Martin van Hemert (LexIQ)
- Vincent van Os (Universiteit Utrecht)
- Manuella van der Put (Rechtbank Oost Brabant, Tilburg University)
- Laurens Naudts (KU Leuven)
- Jeroen Naves (Pels Rijcken)
- Merel Noorman (Tilburg University)
- Rachel Rietveld (ArbeidsmarktResearch UvA BV / Universiteit van Amsterdam)
- Esther Talal (Universiteit Utrecht)
- Tjerk Timan (TNO)
- Kees de Vey Mestdagh (Foundation for Law & ICT, Software Borg Instituut)
- Maranke Wieringa (Universiteit Utrecht)

[1] Alle op deze lijst genoemde personen hebben toestemming gegeven voor het vermelden van hun naam en organisatie in dit overzicht.

Bijlage 4. Geïnterviewde personen[2]

Respondenten die zijn geïnterviewd voor hoofdstukken 1-3 en 8-10[3]

- Jeroen Goudsmit, Global AML & Sanctions Officer bij Rabobank, Docent Compliance & Integriteitsmanagement bij de Vrije Universiteit Amsterdam
- Lynda Hardman, Manager Onderzoek & Strategie Centrum Wiskunde & Informatica, hoogleraar Multimedia Discourse Interaction Universiteit Utrecht, Directeur Amsterdam Data Science
- Sophie Horsman**,** stagiaire TNO
- Lotte Houwing, beleidsadviseur en onderzoeker Bits of Freedom
- Just Stam, hoofd van het Beleidsteam privacy van het ministerie van Justitie en Veiligheid
- Marc Steen, senior onderzoeker TNO

Respondenten die zijn geïnterviewd voor Casestudy Contentmoderatie door online platformen (Hoofdstuk 4)[4]

- Sjarrel de Charon, voormalig medewerker Arvato Bertelsmann; auteur van 'De Achterkant van Facebook'
- Edo Haveman, Head of Public Policy Facebook Netherlands
- Hans Klok, coördinerend beleidsmedewerker Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
- Pieter van Koetsveld, beleidsadviseur Ministerie van Onderwijs, Cultuur en Wetenschap
- Tarlach McGonagle, bijzonder hoogleraar Mediarecht & Informatiesamenleving Universiteit Leiden; senior onderzoeker en docent Instituut voor Informatierecht
- Amber Mechels, beleidsadviseur Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
- Maarten Sap, onderzoeker University of Washington
- Bastiaan Winkel-Boer, beleidsadviseur Ministerie van Justitie en Veiligheid

Respondenten die zijn geïnterviewd voor Casestudy Zelfrijdende auto's (Hoofdstuk 5)[5]

- Jos Elfring, Assistant Professor Mechanical Engineering TU Eindhoven
- Frans van Waes, Innovation & New Business Developer Vialis
- Edwin Nas, (*tot 1 april 2020*) Lead for Connected and Automated Driving Ministerie van Infrastructuur en Waterstaat
- Simeon Calvert, Research fellow Transport & Planning TU Delft

[2] Alle op deze lijst genoemde personen hebben toestemming gegeven voor het vermelden van hun naam en organisatie in dit overzicht.
[3] Respondenten Horsman en Steen zijn gezamenlijk geïnterviewd.
[4] Respondenten Klok, Van Koetsveld en Mechels zijn gezamenlijk geïnterviewd.
[5] Respondenten Hottentot en De Jong zijn gezamenlijk geïnterviewd.

- Chris Hottentot, Adviseur Public Affairs ANWB
- Ronald de Jong, Teammanager Mobiliteit & Verkeersveiligheid ANWB

Respondenten die zijn geïnterviewd voor de Casestudy De rechtspraak (Hoofdstuk 6)[6]

- Bart Jan van Ettekoven, voorzitter Afdeling bestuursrechtspraak van de Raad van State
- Martin van Hemert, CEO *LexIQ*
- Manuella van der Put, senior rechter Rechtbank Oost-Brabant en promovenda Universiteit van Tilburg
- Dory Reiling, gepensioneerd senior rechter, internationaal Rechtspraak en IT expert
- Graham Ross, international adviseur op het gebied van online dispute resolution en oprichter van *The Resolver*
- Bart Schellekens, Adviseur / onderzoeker Recht en ICT, Raad voor de rechtspraak
- Jos Smits, strategisch adviseur IV, Raad voor de rechtspraak
- Nico Tuijn, senior raadsheer Hof den Bosch
- Twee beleidsadviseurs, Ministerie van Justitie en Veiligheid (*op hun uitdrukkelijk verzoek wordt hun naam niet opgenomen*)

Respondenten die zijn geïnterviewd voor de Casestudy Overheidsincasso bij verkeersboetes (Hoofdstuk 7)[7]

- Joke de Boer, data scientist, Centraal Justitieel Incassobureau
- Roxane Daniels, teamleider, Vereniging van Nederlandse Gemeenten
- Han Dieperink, algemeen directeur, Instituut voor Midden- en Kleinbedrijf
- Caroline de Groot, advocaat sociaal zekerheidsrecht, Advocatenkantoor De Binnenstad
- Suzanne Hartholt, IT-jurist, Ministerie van Justitie en Veiligheid
- Juliette van der Jagt-Jobsen, senior juridisch adviseur, Vereniging van Nederlandse Gemeenten
- Marc Minnee, projectmanager, Ministerie van Sociale Zaken en Werkgelegenheid
- Nora Otto, senior beleidsmedewerker, Vereniging van Nederlandse Gemeenten
- Emiel Reck, senior beleidsmedewerker, Ministerie van Justitie en Veiligheid
- Pieter Roos, senior beleidsmedewerker, Ministerie van Sociale Zaken en Werkgelegenheid
- Mieke van der Vegt, senior beleidsmedewerker, Ministerie van Sociale Zaken en Werkgelegenheid
- Marije Zijlstra, senior adviseur, Centraal Justitieel Incassobureau

[6] Respondenten Schellekens en Smits zijn gezamenlijk geïnterviewd. Ook de twee beleidsadviseurs van het Ministerie van Justitie en Veiligheid zijn gezamenlijk geïnterviewd.
[7] Respondenten Daniels, Minnee, Otto en Van der Jagt-Jobsen zijn gezamenlijk geïnterviewd. Tevens zijn respondenten Roos en Van der Vegt gezamenlijk geïnterviewd.

Bijlage 5. Vragenlijst semi-gestructureerde interviews

Structuur

De interviews zijn opgezet volgens de structuur van de levenscyclus van een algoritme. Daarin hebben we in ons onderzoek drie fases onderscheiden:

(1) De probleemanalyse. In deze fase wordt onder andere het probleem dat het algoritme beoogt te adresseren, onderzocht en geanalyseerd. Het te bereiken doel van het algoritme wordt dan ook bepaald.
(2) De ontwikkelfase. In deze fase wordt het algoritme ontworpen, geschreven en getest. In geval van machine learning wordt benodigde data verzameld, bewerkt en aangeboden. Het algoritme wordt in deze fase ook getest. Daarvoor wordt veelal testdata gebruikt als invoer.
(3) De gebruiksfase. In deze fase wordt het algoritme in gebruik genomen door de toepasser. Dit is ook de fase waarin een betrokkene de gevolgen kan ondervinden van het gebruik van het algoritme.

De vragen met betrekking tot de probleemanalyse zijn het meest geschikt om te stellen aan de toepasser die het algoritme gebruikt, of de ontwikkelaars die betrokken zijn geweest in deze fase. De vragen met betrekking tot de ontwikkelingsfase zijn het meest geschikt om aan een ontwikkelaar van algoritmen te stellen. De vragen over de gebruiksfase zijn het meest van toepassing op de personen die daadwerkelijk met het algoritme hebben gewerkt of de personen die onderworpen zijn aan algoritmische besluitvorming of aan verenigingen die de belangen van deze personen vertegenwoordigen.

Het ligt voor de hand dat in het domein van de casestudy meerdere complexe algoritmen aan het werk zijn en het misschien beter is om te spreken van een systeem van algoritmen. Om de leesbaarheid van de vragen te bevorderen spreken we echter steeds van ‘een algoritme’ of ‘het algoritme’.

Verklarende woordenlijst:

- **Ontwikkelaar:** de persoon of groep van personen die het algoritme programmeert.
- **Toepasser**: de organisatie die het algoritme gebruikt.
- **Betrokkene**: de persoon die de gevolgen ondervindt van het door het algoritme genomen besluit.
- **Besluit**: het maken van een keuze die de belangen van een rechtssubject raakt. Dit omvat dus besluiten in meer traditionele juridische zin, maar ook besluiten met feitelijke gevolgen.

Vraag	Focusgebied onderzoek
Algemene introductie	
Kunt u zichzelf kort introduceren? Op welke manier bent u betrokken bij (de ontwikkeling / gebruik / regulering van) het algoritme?	Context
Kunt u iets vertellen over het gebruik van algoritmen in [domein casestudy]?	Context
Kunt u een korte introductie geven van het betreffende algoritme en de werking daarvan?	Context
Hoe zal, naar uw verwachting, het gebruik van het algoritme zich in de komende vijf tot tien jaar kunnen ontwikkelen?	Ontwikkelingen
Probleemanalyse	
Kunt u vertellen welk probleem wordt geadresseerd met het algoritme? Wat is het doel van het algoritme?	Context
In hoeverre is het doel van het algoritme door de toepasser omschreven en gedocumenteerd?	Context
In hoeverre wordt gemonitord of het algoritme de problemen daadwerkelijk oplost?	Context
Welke doelstellingen zal, naar uw verwachting, de toepasser met het algoritme de komende vijf tot tien jaar mogelijk willen en kunnen realiseren?	Context Ontwikkelingen
Ontwikkeling van het algoritme	
Kunt u iets vertellen over de wijze waarop het algoritme is ontwikkeld?	Context
Kunt u het type algoritme kwalificeren? Is er sprake van een regelgebaseerd algoritme, machine learning of deep learning?	Context Voorspelbaarheid Uitlegbaarheid

Als sprake is van machine learning, deep learning of reinforcement learning: hoe wordt het algoritme getraind? Wat voor data worden daarvoor gebruikt?	Voorspelbaarheid Uitlegbaarheid
In hoeverre is diversiteit een overweging bij het samenstellen van het team dat een algoritme ontwikkelt?	Publieke waarden: non-discriminatie Risico's
Is de broncode van het algoritme toegankelijk? Zo ja, onder welke voorwaarden, en voor voor wie?	Context Voorspelbaarheid Uitlegbaarheid Kansen/Risico's
Wat zijn de overwegingen voor het wel/niet openbaar maken van de broncode?	Context
In hoeverre is het ontwikkelingsproces van het algoritme gedocumenteerd?	Context Voorspelbaarheid Uitlegbaarheid
In hoeverre is documentatie over de ontwikkeling van het algoritme openbaar?	Context Voorspelbaarheid Uitlegbaarheid
Wat zijn de overwegingen voor het wel/niet openbaar maken van de documentatie?	Context Voorspelbaarheid Uitlegbaarheid Kansen/Risico's
In hoeverre wordt bij het ontwerpen van het algoritme de begrijpelijkheid van de uiteindelijke werking daarvan in acht genomen? En voor wie moet het algoritme begrijpelijk zijn?	Uitlegbaarheid
In hoeverre zijn er in het algoritme zelf mechanismen ingebouwd die de uitkomst van het algoritme kunnen helpen verklaren in termen die voor de toepasser of de betrokkene begrijpelijk zijn?	Uitlegbaarheid
In hoeverre wordt er feedback gevraagd door de toepasser binnen de eigen organisatie, aan betrokkenen, of anderen bij de ontwikkeling en het gebruik van het algoritme? Met name als het gaat om kansen en risico's van het gebruik van het algoritme voor publieke waarden?	Publieke waarden Kansen en risico's

Zijn er criteria geformuleerd waarlangs de ontwikkeling en het gebruik van het algoritme kan worden getoetst?	Publieke waarden Kansen en risico's
In hoeverre vinden er effectbeoordelingen plaats? En zo ja; hoe? Wanneer vinden die plaats en worden die gedocumenteerd? Op welke wijze wordt met de uitkomsten daarvan rekening gehouden?	Publieke waarden Kansen en risico's
In hoeverre wordt getest of de verwachte uitvoer van het algoritme overeenkomt met de daadwerkelijke uitvoer van het algoritme? En hoe wordt dat getest?	Publieke waarden Kansen en risico's
Als het algoritme leidt tot onwenselijke uitkomsten of besluiten, in hoeverre is er een procedure of methode om het algoritme op basis daarvan te verbeteren (en zo de onwenselijke uitkomsten of besluiten in de toekomst te voorkomen)?	Publieke waarden Kansen en risico's
Zijn er in het [domein casestudy] standaarden of codes voor het documentatie, de uitlegging, en het toezicht op het algoritme? En in hoeverre worden die nageleefd? Waarom wel/niet?	Uitlegbaarheid
Gebruik van het algoritme	
Kunt u iets vertellen over de manier waarop het algoritme werkt?	Context
Hoe voorspelbaar zijn de besluiten van het algoritme volgens u? Het gaat daarbij om de voorspelbaarheid van uitkomsten in een concreet geval.	Voorspelbaarheid
Hoe uitlegbaar zijn de besluiten van het algoritme volgens u? Daarmee wordt bedoeld in hoeverre de uitkomsten van het algoritme en de totstandkoming daarvan begrepen en verklaard kunnen worden in een concreet geval.	Uitlegbaarheid
Hoe zelfstandig is het algoritme volgens u? Hiermee wordt gedoeld op de mate waarin mensen aan de uiteindelijke besluitvorming te pas komen.	Zelfstandig
In hoeverre is het mogelijk om genomen besluiten ongedaan te maken?	Context Zelfstandigheid Rechtsbescherming
Als een mens het besluit neemt, in hoeverre is de uitvoer van een algoritme leidend bij het nemen van het besluit?	Zelfstandigheid

Als het algoritme een advies geeft, in hoeverre is het mogelijk om daarvan af te wijken?	Zelfstandigheid
Waar ligt de verantwoordelijkheid voor het gegeven advies of het genomen besluit?	Zelfstandigheid
Is er iemand werkzaam bij de toepasser (die het algoritme heeft gemaakt of die het algoritme gebruikt) die de uitvoer van het algoritme kan verklaren? Indien er niemand werkzaam is bij de toepasser (die het algoritme heeft gemaakt of die het algoritme gebruikt) die de uitvoer van het algoritme zou kunnen verklaren, kan de uitvoer dan op een andere manier worden verklaard?	Uitlegbaarheid
Wordt gedocumenteerd op welke wijze het algoritme tot een bepaalde uitkomst komt? Zo ja, op welke manier wordt dit gedaan?	Uitlegbaarheid
In hoeverre worden er aan het algemene publiek inlichtingen verstrekt over de werking van het algoritme?	Voorspelbaarheid
In hoeverre wordt aan betrokkenen kenbaar gemaakt dat zij zijn onderworpen aan (deels) algoritmische besluitvorming? Wordt de werking van het algoritme aan hen uitgelegd en zo ja; wordt getoetst of zij de uitleg hebben begrepen?	Uitlegbaarheid Publieke waarden: rechtsbescherming
In hoeverre kunnen mensen die zijn onderworpen aan de besluitvorming ageren tegen het algoritmisch genomen besluit? En worden zij daarover geïnformeerd?	Rechtsbescherming
In hoeverre wordt getoetst of de verwachte uitvoer overeenkomt met de daadwerkelijke output van het algoritme?	Context
In hoeverre vinden er effectbeoordelingen plaats? En zo ja; hoe? Wanneer vinden die plaats en worden die gedocumenteerd? Op welke wijze wordt met de uitkomsten daarvan rekening gehouden?	Kansen en risico's
Wordt er door de toepasser toezicht gehouden op de ontwikkeling en het gebruik van het algoritme? Zo ja, hoe ziet dat toezicht eruit? In hoeverre wordt er verslag gelegd van dat toezicht?	Publieke waarden Kansen en risico's

Als het algoritme leidt tot onwenselijke uitkomsten of besluiten, in hoeverre is er een procedure of methode om die in de toekomst te voorkomen en het algoritme op basis daarvan te verbeteren?	Context
Is er nagedacht over mogelijke problemen die kunnen ontstaan, en vastgesteld wanneer die problemen leiden tot omschakeling naar een plan B (bijv. abort, ingrijpen door een mens, ander soort KI systeem)?	Risico's
Welke **kansen** doen zich reeds voor of kunnen zich in de toekomst voordoen door het gebruik van het algoritme ten aanzien van de verwezenlijking van publieke waarden **in het algemeen**? Ten aanzien van het toekomstperspectief: welke factoren en ontwikkelingen kunnen daarbij van belang zijn?	Publieke waarden Kansen en risico's
Welke **risico's** doen zich reeds voor of kunnen zich in de toekomst voordoen door het gebruik van het algoritme ten aanzien van de verwezenlijking van publieke waarden in het algemeen? Ten aanzien van het toekomstperspectief: welke factoren en ontwikkelingen kunnen daarbij van belang zijn?	Publieke waarden Kansen en risico's
Welke **kansen** doen zich reeds voor of kunnen zich in de toekomst voordoen ten gevolge van het gebruik van algoritmen binnen **[domein casestudy]** ten aanzien van de verwezenlijking van het recht op **non-discriminatie/gegevensbescherming/rechtsbescherming[casusspecifieke waarde]?** Ten aanzien van het toekomstperspectief: welke factoren en ontwikkelingen kunnen daarbij van belang zijn?	Publieke waarden, Kansen en risico's
Welke **risico's** doen zich reeds voor of kunnen zich in de toekomst voordoen ten gevolge van het gebruik van algoritmen binnen **[domein casestudy]** ten aanzien van de verwezenlijking van het recht op **non-discriminatie/gegevensbescherming/rechtsbescherming/[casusspecifieke waarde]?** Ten aanzien van het toekomstperspectief: welke factoren en ontwikkelingen kunnen daarbij van belang zijn?	Publieke waarden Kansen en risico's
In hoeverre worden gebruikers van het algoritme op de hoogte gesteld van de mogelijke risico's van het gebruik?	Risico's
Wat zijn eventuele technologische obstakels voor het optimaal benutten van de kansen die zich door de toepassing van dit algoritme kunnen voordoen?	Ontwikkelingen

Zijn er richtlijnen (zowel intern, als extern (bijv. in samenwerkingsverband, codes, etc)) voor het gebruik van het algoritme? In hoeverre is het gebruik van het algoritme geclausuleerd? Wanneer mag het wel/niet gebruikt worden?	Kansen en risico's
In hoeverre is de toepassing van het algoritme gereguleerd?	Wettelijk kader
Kunt u een korte beschrijving geven van het juridisch kader dat het gebruik van het betreffende algoritme reguleert?	Wettelijk kader
In hoeverre loopt u aan tegen juridische knelpunten bij het verwezenlijken van kansen of mitigeren van risico's? Bijv. Intellectuele eigendomsrechten, AVG, wet gelijke behandeling, wet RO?	Kansen en risico's Wettelijk kader
Hoe zou het juridisch kader eventueel moeten worden aangepast om de geïdentificeerde risico's te verkleinen en de kansen te benutten (zowel nu, als in de toekomst)?	Wettelijk kader

Bibliografie

AARvS 2018

AARvS, *Ongevraagd advies over de effecten van de digitalisering voor de rechtsstatelijke verhoudingen, Kamerstukken II* 2017/18, 26643, nr. 557.

Van den Acker, *VR* 2015, p. 366-372

B. van den Acker, 'Visies op de autonome auto', *VR* 2015/108, afl. 10, p. 366-372.

Adadi & Berrada, *IEEE Access* 2018, p. 52138-52160

A. Adadi & M. Berrada, 'Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)', *IEEE Access* (6) 2018, p. 52138-52160.

Agenda Digitale Overheid (NL Digibeter) 2019

Agenda Digitale Overheid (NL Digibeter), *Kansen en uitdagingen in de digitale samenleving*, 2019, vng.nl/sites/default/files/nieuws_attachments/2018/7283-bzk_brede_agenda_digitale_overheid-infographicv4_002-bt2.pdf.

AI HLEG 2019

AI HLEG, *Ethische Richtsnoeren voor Betrouwbare KI*, 8 april 2019, doi:10.2759/61918.

Alaba e.a., *Journal of Network and Computer Applications* 2017, p. 10-28

F.A. Alaba e.a., 'Internet of Things security: A survey', *Journal of Network and Computer Applications* (88) 2017, afl. C, p. 10-28.

Aletras e.a., *PeerJ Computer Science* 2016

N. Aletras e.a., 'Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective', *PeerJ Computer Science* 2016, 2:e93doi.org/10.7717/peerj-cs.93.

Angelopoulos e.a. 2015

C.J. Angelopoulos e.a., *Study of fundamental rights limitations for online enforcement through self-regulation*, Amsterdam: IViR 2015.

Angwin e.a., propublica.org 23 mei 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.org 23 mei 2016.

Ashley 2017

K.D. Ashley, *Artificial Intelligence and Legal Analytics. New Tools for Law Practice in the Digital Age,* Cambridge: Cambridge University Press 2017.

Ashton, *RFID Journal* 22 juni 2009

K. Ashton, 'That "Internet of Things" Thing', *RFID Journal* 22 juni 2009 rfidjournal.com/articles/pdf?4986.

Autoriteit Persoonsgegevens 2019

Autoriteit Persoonsgegevens, *Focus AP 2020-2023*, 11 november 2019, autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/focus_ap_202-2023_groot.pdf.

Autoriteit Persoonsgegevens 2020

Autoriteit Persoonsgegevens, *Connected car? Bescherm uw privacy!*, 2 maart 2020, autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/handleiding_privacybescherming_connected_vehicles_def.pdf.

AVG Commissie Bestuursrechtelijke Colleges 2019

AVG Commissie Bestuursrechtelijke Colleges, *Advies Gegevensverwerking (AVG) en gerechtelijke taken*, 2019, raadvanstate.nl/publish/pages/113611/advies_van_de_avg-commissie.pdf.

Baarda 2012

B. Baarda, *Basisboek Methoden en Technieken*, Groningen: Noordhoff 2012.

Babuta & Oswald 2019

A. Babuta & M. Oswald, *Data Analytics and Algorithmic Bias in Policing* (Briefing Paper Royal United Services Institute for Defence and Security Studies), 2019.

Baehrens e.a., *Journal of Machine Learning Research* 2010, p. 1803-1831

D. Baehrens e.a., 'How to Explain Individual Classification Decisions', *Journal of Machine Learning Research* 2010, afl. 11, p. 1803-1831.

Bakker 2017

F. Bakker, 'Speech Dag van de Rechtspraak 2017', 28 september 2017, rechtspraak.nl/SiteCollectionDocuments/speech-frits-bakker-dvdr-2107.pdf.

Bakker, *NJBlog* 17 augustus 2017

F. Bakker, 'Wettelijke regeling publicatie rechterlijke uitspraken', *NJBlog* 17 augustus 2017.

Bansak e.a., *Science* 2018, p. 325-329

K. Bansak e.a., 'Improving refugee integration through data-driven algorithmic assignment', *Science* (359) 2018, afl. 6373, p. 325-329.

Barendrecht e.a. 2016

M. Barendrecht e.a., *ODR and the courts: The promise of 100% access to justice?* (HiiL Trend Report IV), 2016.

Barocas, Hardt & Narayanan 2015

S. Barocas, M. Hardt & A. Narayanan, *Fairness and machine learning. Limitations and Opportunities*, fairmlbook.org (online tekstboek, laatst bijgewerkt op 6 december 2015).

Barocas & Selbst, *California Law Review* 2016, p. 671-732

S. Barocas & A.D. Selbst, 'Big Data's Disparate Impact', *California Law Review* (104) 2016, afl. 3, p. 671-732.

Bauw, *AA* 2018, p. 890-893

E. Bauw, 'Geschillen als handelswaar: Over cliffhangers en e-Court-soap', *AA* 2018, afl. 11, p. 890-893.

Behoorlijkheidswijzer 2019

Behoorlijkheidswijzer, Den Haag: Nationale ombudsman 2019.

Besson, *Human Rights Law Review* 2008, p. 647-682

S. Besson, 'Gender Discrimination under EU and ECHR Law: *Never Shall the Twain Meet*?', *Human Rights Law Review* (8) 2008, afl. 4, p. 647–682.

Bex & Prakken, *AA* 2020, p. 255-259

F.J. Bex & H. Prakken, 'De juridische voorspelindustrie: onzinnige hype of nuttige ontwikkeling?', *AA* 2020, p. 255-259

Bijlsma, Meynen & Bex, *NJB* 2019, **p. 3313-3319**

J. Bijlsma, G. Meynen & F.J. Bex, 'Artificiële intelligentie en risicotaxatie. Drie kernvragen voor strafrechtjuristen', *NJB* 2019/2778, afl. 44, p. 3313-3319.

Binns e.a. 2017

R. Binns (e.a.), 'Like Trainer, Like Bot? Inheritance of Bias in Algorithmic Content Moderation', in: G. Ciampaglia, A. Mashhadi & T. Yasseri (red.), *Social Informatics* (SocInfo 2017, Lecture Notes in Computer Science, vol 10540), Cham (Zwitserland): Springer 2017, p. 405-415.

Boeglin, *Yale Journal of Law and Technology* 2015, p. 171-203

J. Boeglin, 'The Costs of Self-Driving Cars: Reconciling Freedom and Privacy with Tort Liability in Autonomous Vehicle Regulation', *Yale Journal of Law and Technology* (17) 2015, afl. 1, p.171-203.

Borking 2010

J.J.F.M. Borking, *Privacyrecht is Code. Over het gebruik van privacy-enhancing technologies* (diss. Leiden), Leiden: Kluwer 2010, hdl:1887/15660.

Brandão 2019

M. Brandão, *Age and gender bias in pedestrian detection algorithms* (Workshop on Fairness Accountability Transparency and Ethics in Computer Vision (FATE CV) at CVPR 2019), 2019, arXiv:1906.10490.

Branting, *Artificial Intelligence and Law* 2017, p. 5-27

L.K. Branting, 'Data-centric and logic-based models for automated legal problem solving', *Artificial Intelligence and Law* (25) 2017, afl. 1, p. 5-27.

Brechtel, Gindele & Dillmann 2014

S. Brechtel, T. Gindele & R. Dillmann, *Probabilistic Decision-Making under Uncertainty for Autonomous Driving using Continuous POMDPs* (IEEE International Conference on Intelligent Transportation Systems), oktober 2014.

Van Breda, *Computerrecht* 2017, p. 223-229

B.C. van Breda, 'Profilering in de AVG: nieuwe regels, voldoende bescherming?', *Computerrecht* 2017/154, afl. 4, p. 223-229.

Brown, *Law and Philosophy* 2017, p. 419-468

A. Brown, 'What is hate speech? Part 1: The Myth of Hate', *Law and Philosophy* (36) 2017, afl. 4, p. 419-468.

Bröring & De Graaf 2019

H.E. Bröring & K.J. de Graaf (red.), *Bestuursrecht 1. Systeem; bevoegdheid; bevoegdheidsuitoefening; handhaving*, Den Haag: Boom juridisch 2019.

De Bruin, *European Journal of Risk Regulation* 2016, p. 485-501

R.W. de Bruin, 'Autonomous Intelligent Cars on the European Intersection of Liability and Privacy: Regulatory Challenges and the Road Ahead', *European Journal of Risk Regulation* (7) 2016, afl. 3, p. 485-501.

Buijze 2013

A.W.G.J. Buijze, *The principle of transparency in EU law* (diss. Utrecht), 's-Hertogenbosch: Uitgeverij BOXPress 2013.

Burkov 2019

A. Burkov, *The Hundred-Page Machine Learning Book*, Burkov 2019.

Calvert e.a., *Theoretical Issues in Ergonomics Science* 2019, p. 1-29

S.C. Calvert e.a., 'A human centric framework for the analysis of automated driving systems based on meaningful human control', *Theoretical Issues in Ergonomics Science* 2019, p. 1-29, doi.org/10.1080/1463922X.2019.1697390.

Calvert e.a., *IEEE Intelligent Transportation Systems Magazine* 2020

S.C. Calvert e.a., 'Gaps in the control of automated vehicles on roads', *IEEE Intelligent Transportation Systems Magazine* 2020, doi.org/10.1109/MITS.2019.2926278.

Carneiro e.a., *Artificial Intelligence Review* 2014, p. 211-240

D. Carneiro e.a., 'Online dispute resolution: an artificial intelligence perspective' *Artificial Intelligence Review* (41) 2014, afl. 2, p. 211-240.

CFTC & SEC 2010

CFTC & SEC, *Findings regarding the market events of May 6, 2010*. *Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues*, 30 september 2010.

Cheng, Danescu-Niculescu-Mizil & Leskovec 2015

J. Cheng, C. Danescu-Niculescu-Mizil & J. Leskovec, *Antisocial Behavior in Online Discussion Communities*, 2015, arXiv:1504.00680.

Chesney & Citron, *Foreign Affairs* 2019, p. 147
R. Chesney & D. Citron, 'Deepfakes and the new disinformation war: The coming age of post-truth geopolitics', *Foreign Affairs* (98) 2019, afl. 1, p. 147.

Chouldechova 2017
A. Chouldechova, 'Fair prediction with disparate impact: A study of bias in recidivism prediction instruments', 28 februari 2017, arXiv:1703.00056v1.

Citron & Norton, *Boston University Law Review* 2011, p. 1435-1484
D.K. Citron & H. Norton, 'Intermediaries and Hate Speech: Fostering Digital Citizenship for Our Information Age', *Boston University Law Review* (91) 2011, afl. 4, p. 1435-1484.

De Cock Buning & De Bruin, *Connection Science* 2017, p. 189-199
M. de Cock Buning & R.W. de Bruin, 'Autonomous intelligent cars: proof that the EPSRC Principles are future-proof', *Connection Science* (29) 2017, afl. 3, p. 189-199.

Cohen-Almagor, *Policy & Internet* 2011, p. 1-26
R. Cohen-Almagor, 'Fighting Hate and Bigotry on the Internet', *Policy & Internet* (3) 2011, afl. 3, p. 1-26.

Coldewey, techcrunch.nl 10 september 2019
D. Coldewey, 'Hatebase catalogues the world's hate speech in real time so you don't have to', techcrunch.nl 10 september 2019.

Congresmagazine nationale ombudsman 2019
Nationale ombudsman, *Wie doet er mee? Burger en overheid in 2030* (*Congresmagazine* december 2019), Den Haag: Nationale ombudsman 2019.

Conger, *The New York Times* 9 juli 2019
K. Conger, 'Twitter Backs Off Broad Limits on "Dehumanizing" Speech', *The New York Times* 9 juli 2019, nytimes.com/2019/07/09/technology/twitter-ban-speech-dehumanizing.html.

Cormen e.a. 2009
T.H. Cormen e.a., *Introduction to Algorithms*, Cambridge: MIT Press 2009.

Council of Europe 2018

Council of Europe, *Algorithms and Human Rights. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications* (report prepared by the Committee of Experts on Internet Intermediaries MSI-NET), DGI(2017)12, maart 2018.

Cowan, *Journal of Social Issues* 2002, p. 247-263

G. Cowan e.a., 'Hate Speech and Constitutional Protection: Priming Values of Equality and Freedom', *Journal of Social Issues* (58) 2002, afl. 2, p. 247-263.

Dang, *Reuters* 21 februari 2019

S. Dang, 'AT&T pulls ads from YouTube over videos exploiting children', *Reuters*.com 21 februari 2019.

Datatilsynet 2018

Artificial intelligence and privacy (Rapport van de Noorse gegevensbeschermingsautoriteit Datatilsynet van januari 2018), Oslo (Noorwegen): Datatilsynet 2018.

Davidson e.a. 2017

T. Davidson e.a., *Automated Hate Speech Detection and the Problem of Offensive Language* (Proceedings of the Eleventh International AAAI Conference on Web and Social Media, ICWSM), 2017, p. 1-4.

Davidson, Bhattacharya & Weber 2019

T. Davidson, D. Bhattacharya & I. Weber, *Racial Bias in Hate Speech and Abusive Language Detection Datasets* (Proceedings of the Third Workshop on Abusive Language Online), 2019, p. 25-35.

Dickson, *TechTalks* 28 augustus 2017

B. Dickson, 'What is machine learning?', *TechTalks*.com 28 augustus 2017.

Dieterich, Mendoza & Brennan 2016

W. Dieterich, C. Mendoza & T. Brennan, *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity* (technisch onderzoeksrapport van 8 juli 2016 van het Northpointe Inc. Research Department).

Dolhansky e.a. 2019

B. Dolhansky e.a., *The Deepfake Detection Challenge (DFDC) Preview Dataset*, 2019, arXiv:1910.0885.

Doove & Otten 2018

S. Doove & D. Otten, *Verkennend onderzoek naar het gebruik van algoritmen binnen overheidsorganisaties*, Den Haag: CBS 2018.

Došilović, Brčić & Hlupić 2018

F.K. Došilović, M. Brčić & N. Hlupić, *Explainable Artificial Intelligence: A Survey* (41st International Convention on Information and Communication Technology, Electronics and Microelectronics), 2018, doi.org/10.23919/MIPRO.2018.8400040.

Dressel & Farid, *Science Advances* 2018, p. 1-5

J. Dressel & H. Farid, 'The accuracy, fairness, and limits of predicting recidivism', *Science Advances* (4) 2018, afl. 1, p. 1-5.

Du & Yu, *China Justice Observer* 16 december 2018

G. Du & M. Yu, 'China Establishes Three Internet Courts to Try Internet-Related Cases Online', chinajusticeobserver.com 16 december 2018.

Duhigg, *The New York Times* 16 februari 2012

C. Duhigg, 'How Companies Learn Your Secrets', *The New York Times* 16 februari 2012, nytimes.com/2012/02/19/magazine/shopping-habits.html.

Van Duin, *Sdu Blog* 5 februari 2018

A. van Duin, 'E-court en 'robotrechtspraak': efficiëntie ten koste van rechtsbescherming', *Sdu Blog* 5 februari 2018, sdu.nl/blog.

Dunk, *Advocatenblad* 28 augustus 2019

S. Dunk, ''Verschil vonnissen ontslagrecht bij advocaten al bekend'', *Advocatenblad* 28 augustus 2019, advocatenblad.nl

Van Eck 2013

B.M.A. van Eck, 'Is er meer tussen mens en machine?', 14 juni 2013, marliesvaneck.wordpress.com.

Van Eck 2018

B.M.A. van Eck, *Geautomatiseerde Ketenbesluiten & Rechtsbescherming. Een onderzoek naar de praktijk van geautomatiseerde ketenbesluiten over een financieel belang in relatie tot rechtsbescherming* (diss. Tilburg), Tilburg: Universiteit Tilburg 2018.

Eiband, Schneider & Buschek 2018
M. Eiband, H. Schneider & D. Buschek, 'Normative vs Pragmatic: Two Perspectives on the Design of Explanations in Intelligent Systems', *IUI Workshops* 2018.

Van den Eijnden 2011
P.M. van den Eijnden, *Onafhankelijkheid van de rechter in constitutioneel perspectief* (Staat en recht, deel 3), Deventer: Wolters Kluwer 2011.

Eleveld, *Sociaal Maandblad Arbeid* 2008, p. 313-317
A. Eleveld, 'Een uitkering voor zwangere zelfstandigen: formele of materiële gelijkheid', *Sociaal Maandblad Arbeid* 2008/63, afl. 7/8, p. 313-317.

Elfring e.a., *Sensors* 2016, p. 1668-1695
J. Elfring e.a., 'Effective World Modeling: Multisensor Data Fusion Methodology for Automated Driving', *Sensors* (16) 2016, afl. 10, p. 1668-1694.

Van Emmerik, Loof & Schuurmans 2014
M.L. van Emmerik, J.P. Loof & Y.E. Schuurmans, *Systeemwaarborgen voor de kernwaarden van de rechtspraak* (Rechtspraak Research Memoranda 2014, nummer 2, jaargang 10), Leiden: Raad voor de rechtspraak 2014.

Van Emmerik & Schuurmans, ***NJB* 2016, p. 795-799**
M.L. van Emmerik & Y.E. Schuurmans, 'Meer transparantie bij rechterlijke zaakstoedeling dringend gewenst', *NJB* 2016/593, afl. 12, p. 795-799.

Engelhard, *AA* 2017, p. 230-236
E.F.D. Engelhard, 'Wetgever, pas op! De (vrijwel) autonome auto komt eraan', *AA* 2017, afl. 3, p. 230-236.

Engelhard & De Bruin 2018
E.F.D. Engelhard & R.W. de Bruin, *Liability for Damage Caused by Autonomous Vehicles*, Den Haag: Eleven International Publishing 2018.

ERTRAC Working Group "Connectivity and Automated Driving" 2017
ERTRAC Working Group "Connectivity and Automated Driving", *Automated Driving Roadmap*, 29 mei 2017.

ERTRAC Working Group "Connectivity and Automated Driving" 2019
ERTRAC Working Group "Connectivity and Automated Driving", *Connected Automated Driving Roadmap*, 8 maart 2019.

Van Est e.a. 2019
R. van Est e.a., *Waardevol digitaliseren. Hoe lokale bestuurders vanuit publiek perspectief mee kunnen doen aan het 'technologiespel',* Den Haag: Rathenau Instituut 2019.

Van Est & Gerritsen 2017
R. van Est & J.B.A. Gerritsen, *Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality* (Expert report written for the Committee on Culture, Science, Education and Media of the Parliamentary Assembly of the Council of Europe (PACE)), Den Haag: Rathenau Instituut 2017.

Ethik-Kommission Automatisiertes und Vernetztes Fahren 2017
Ethik-Kommission Automatisiertes und Vernetztes Fahren, *Bericht* (Bundesministerium für Verkehr und digitale Infrastruktur), juni 2017.

Van Ettekoven & Prins 2018
B.-J. van Ettekoven & C. Prins, 'Data analysis, artificial intelligence and the judiciary system', in: V. Mak, E. Tjong Tjin Tai & A. Berlee (red.), *Research Handbook in Data Science and Law*, Cheltenham (VK): Edward Elgar 2018, p. 425-447.

European Network of Councils for the Judiciary 2014
European Network of Councils for the Judiciary, *Minimum Judicial Standards IV. Allocation of Cases* (ENCJ Report 2013-2014, adopted in Rome, 13 June 2014).

European Union Agency for Fundamental Rights & Council of Europe 2018
European Union Agency for Fundamental Rights & Council of Europe, *Handbook on European non-discrimination law*, Luxembourg: Publications Office of the European Union 2018.

European Union Agency for Fundamental Rights 2018
European Union Agency for Fundamental Rights, '#BigData: Discrimination in data-supported decision making', European Union Agency for Fundamental Rights, 2018, doi:10.2811/128407

Europese Commissie 2011
Europese Commissie, *Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system*, 23 maart 2011, COM(2011) 144 final.

Europese Commissie 2016

Europese Commissie, *Saving Lives: Boosting Car Safety in the EU*, 12 december 2016, COM(2016) 787 final.

Europese Commissie 2018

Europese Commissie, *On the road to automated mobility: An EU strategy for mobility of the future*, 17 mei 2018, COM(2018) 283 final.

European Data Protection Board 2020

European Data Protection Board, *Guidelines 1/2020 on processing personal data in the context of connected vehicles and mobility related applications* (version 1.0*)*, 28 januari 2020.

European Union Agency for Cybersecurity 2019

European Union Agency for Cybersecurity, *ENISA Good Practices for Security of Smart Cars*, ENISA november 2019.

Evas 2018

T. Evas, *A common EU approach to liability rules and insurance for connected and autonomous vehicles* (European Parliamentary Research Service), 2018.

Expert Group on Liability and New Technologies - New Technologies Formation 2019

Expert Group on Liability and New Technologies - New Technologies Formation, *Liability for Artificial Intelligence and other emerging digital technologies*, European Union, 2019, doi:10.2838/573689.

Fagnant & Kockelman, *Transportation Research Part A* 2015, p. 167-181

D.J. Fagnant & K. Kockelman, 'Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations', *Transportation Research Part A* (77) 2015, p. 167-181.

Fehrenbacher, FORTUNE.com 16 oktober 2015

K. Fehrenbacher, 'How Tesla is ushering in the age of the learning car', FORTUNE.com 16 oktober 2015.

Fenton & Neil 2018

N.E. Fenton & M. Neil, 'Criminally Incompetent Academic Misinterpretation of Criminal Data - and how the Media Pushed the Fake News', *ResearchGate* januari 2018, p. 1-5, doi.org/10.13140/RG.2.2.32052.55680.

Friedman & Nissenbaum, *ACM Transactions on Information Systems* 1996, p. 330-347
B. Friedman & H. Nissenbaum, 'Bias in Computer Systems', *ACM Transactions on Information Systems* (14) 1996, afl. 3, p. 330–347.

Frissen, Van Eck en Drouen 2019
V. Frissen, M. Van Eck en T. Drouen, *Onderzoek toezicht op het gebruik van algoritmen door de overheid*, Hooghiemstra en partners 25 november 2019, bijlage bij *Kamerstukken II* 2019-20, 26643, nr. 657.

Fulda, *Artificial Intelligence and Law* 2012, p. 321-333
J.S. Fulda, 'Implications of a logical paradox for computer-dispensed justice reconsidered: some key differences between minds and machines', *Artificial Intelligence and Law* (20) 2012, afl. 3, p. 321-333.

Gaeta, *Diritto Mercato Tecnologia* 2017, p. 1-20
M.C. Gaeta, 'The issue of data protection in the Internet of Things with particular regard to self-driving cars', *Diritto Mercato Tecnologia* 2017, p. 1-20.

Gagnon 2010
Y.-C. Gagnon, *The Case Study as Research Method: A Practical Handbook*, Quebec:
Presses de l'Université du Québec 2010.

Gal, *Michigan Technology Law Review* 2018, p. 59-104
M.S. Gal, 'Algorithmic Challenges to Autonomous Choice', *Michigan Technology Law Review* (25) 2018, afl. 1, p. 59–104.

Gawron e.a., *Environmental Science & Technology* 2018, p. 3249-3256
J.H. Gawron e.a., 'Life Cycle Assessment of Connected and Automated Vehicles: Sensing and Computing Subsystem and Vehicle Level Effects', *Environmental Science & Technology* (52) 2018, afl. 5, p. 3249-3256.

Gebru 2020
T. Gebru, 'Race and Gender', in: M.D. Dubber, F. Pasquale & S. Das, *Oxford Handbook of Ethics of AI*, Oxford: Oxford University Press 2020.

Gerards 2002
J.H. Gerards, *Rechterlijke toetsing aan het gelijkheidsbeginsel* (diss. Maastricht), Den Haag: Sdu 2002.

Gerards, *NJCM-Bulletin* 2004, p. 176-198

J.H. Gerards, 'Gelijke behandeling en het EVRM. Artikel 14 EVRM: van krachteloze waarborg naar 'norm met tanden'?', *NJCM-Bulletin* 2004/29, afl. 2, p. 176-198.

Gerards 2018

J.H. Gerards, 'Chapter 30. Prohibition of Discrimination Art. 14. Art. 1, Prot. 12', in: P. van Dijk e.a. (red.), *Theory and Practice of the European Convention on Human Rights*, Antwerpen: Intersentia 2018, p. 997-1028.

Gerards 2019

J.H. Gerards, 'Grondrechten in de platformeconomie', in: J.H. Gerards & A.C. van Schaick, *Digitalisering, vermogensrecht, de platformeconomie en grondrechten. Preadvies Vereniging voor Burgerlijk Recht*, Zutphen: Uitgeverij Paris 2019, p. 95-197.

Gerritsen, Kattenberg & Vermeulen 2018

S. Gerritsen, M. Kattenberg, W. Vermeulen, *Regionale plaatsing vergunninghouders en kans op werk* (CPB Policy Brief 2018/07)

Gibbs, *The Guardian* 7 april 2017

S. Gibbs, 'Google to display fact-checking labels to show if news is true or false', *The Guardian* 7 april 2017.

Gillespie 2016

T. Gillespie, 'Algorithm', in: B. Peters (red.), *Digital Keywords. A Vocabulary of Information Society and Culture*, Princeton: Princeton University Press 2016.

Gillespie 2018

T. Gillespie, *Custodians of the Internet. Platforms, content moderation, and the hidden decisions that shape social media*, New Haven: Yale University Press 2018.

Glancy, *Santa Clara Law Review* 2012, p. 1171-1239

D.J. Glancy, 'Privacy in Autonomous Vehicles', *Santa Clara Law Review* (52) 2012, afl. 4, p. 1171-1239.

Goodfellow, Bengio & Courville 2016

I. Goodfellow, Y. Bengio & A. Courville, *Deep Learning*, Cambridge: MIT University Press 2016.

Griffioen 2011
H. Griffioen, *Privacy en vormen van 'intelligente' mobiliteit* (WRR Webpublicaties), Amsterdam: Amsterdam University Press 2011.

Grimmelikhuijsen, *Rechtstreeks* 2018, p. 13-35
S.G. Grimmelikhuijsen, 'Van gegeven naar verdiend gezag. Hoe kan transparantere rechtspraak (blijvend) bijdragen aan legitimiteit?', *Rechtstreeks* 2018, afl. 2, p. 13-35.

Groep gegevensbescherming artikel 29 2007
Groep gegevensbescherming artikel 29, *Advies 4/2007 over het begrip persoonsgegeven* (01248/07/NL, WP 136), 20 juni 2007.

Groep gegevensbescherming artikel 29 2012
Groep gegevensbescherming artikel 29, *Opinion 3/2012 on developments in biometric technologies* (00720/12/EN, WP193), 27 april 2012.

Groep gegevensbescherming artikel 29 2013
Groep gegevensbescherming artikel 29*, Opinion 03/2013 on purpose limitation* (00569/13/EN, WP 203), 2 april 2013

Groep gegevensbescherming artikel 29 2017a
Groep gegevensbescherming artikel 29, *Richtsnoeren voor gegevensbeschermingseffectbeoordelingen en bepaling of een verwerking "waarschijnlijk een hoog risico inhoudt" in de zin van Verordening 2016/679* (17/NL, WP 248 rev.01), 4 april 2017 (laatst herzien op 4 oktober 2017).

Groep gegevensbescherming artikel 29 2017b
Groep gegevensbescherming artikel 29, *Richtsnoeren inzake geautomatiseerde individuele besluitvorming en profilering voor de toepassing van Verordening (EU) 2016/679* (17/NL, WP251rev.01), 3 oktober 2017 (bijgewerkt 6 februari 2018).

Gröndahl e.a. 2018
T. Gröndahl e.a., *All You Need is "Love": Evading Hate Speech Detection* (Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security), 2018, arXiv:1808.09115.

Grosan & Abraham 2011

C. Grosan & A. Abraham, *Intelligent Systems: A Modern Approach*, Berlijn/Heidelberg: Springer 2011.

Gryffroy, Fiten & Surinx, *Nieuw Juridisch Weekblad* 2019, p. 542-556

P. Gryffroy, B. Fiten & D. Surinx, 'Zelfrijdende wagens anno 2019. Waar staan we juridisch en maatschappelijk?', *Nieuw Juridisch Weekblad* 2019/18, afl. 406, p. 542-556.

Hacker, *International Data Privacy Law* 2017, p. 266-286

P. Hacker, 'Personal data, exploitative contracts, and algorithmic fairness: autonomous vehicles meet the internet of things', *International Data Privacy Law* (7) 2017, afl. 4, p. 266-286.

Hansen, Hoepman & Jensen 2015

M. Hansen, J.-H. Hoepman & M. Jensen, *Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies. Methodology, Pilot Assessment, and Continuity Plan* (Rapport van het European Union Agency for Network and Information Security), ENISA 2015.

Harding, *CBS NEWS* 11 mei 2018

J. Harding, 'Cisco pulls ads from YouTube for "brand safety"', *CBS NEWS*.com 11 mei 2018.

Hardman & Beauxis-Aussalet 2018

L. Hardman & E. Beauxis-Aussalet, 'When Computers Decide: Understanding Uncertainty in Data-driven Systems' (presentatie Informatics Europe 2018), homepages.cwi.nl/~lynda/talks/2018/Lynda_Hardman_Gulbenkian_181127.pdf.

Hecker e.a., *Journal of Communications* 2011, p. 115-124

T. Hecker e.a., 'Model Car Testbed for Development of V2X Applications', *Journal of Communications* (6) 2011, afl. 1, p. 115-124.

Van Helden 2020

J. van Helden, 'Privacyrechtelijke aspecten', in: N. van Duuren & V. de Pous (red.), *Multidisciplinaire aspecten van artificial intelligence*, Amsterdam: deLex 2020.

Herández Encinas e.a. 2015

L. Hernández Encinas e.a., *Online privacy tools for the general public. Towards a methodology for the evaluation of PETs for internet & mobile users* (Rapport van het European Union Agency for Network and Information Security), ENISA 2015.

Hertogh, Winter & Schudde 2012

M.L.M. Hertogh, H.B. Winter & L.T. Schudde, *Een factuurtje uit Leeuwarden: De invloed van legitimiteit op het nalevings-en betalingsgedrag van CJIB-debiteuren*, Groningen: Rijksuniversiteit Groningen, Vakgroep Bestuursrecht & Bestuurskunde 2012.

High Level Group on the Competitiveness and Sustainable Growth of the Automotive Industry in the European Union 2017

High Level Group on the Competitiveness and Sustainable Growth of the Automotive Industry in the European Union, *GEAR 2030* (final report), oktober 2017.

Hildebrandt & Gutwirth 2008

M. Hildebrandt & S. Gutwirth (red.), *Profiling the European Citizen*, Dordrecht: Springer 2008.

HLEG on Fake News and Online Disinformation 2018

High Level Expert Group on Fake News and Disinformation, *A multi-dimensional approach to disinformation. Report of the independent High level Group on fake news and online disinformation*, Luxemburg: Publications Office of the European Union 2018, doi:10.2759/739290.

Hobert e.a., *IEEE Communications Magazine* 2015, p. 64-70

L. Hobert e.a., 'Enhancements of V2X Communication in Support of Cooperative Autonomous Driving', *IEEE Communications Magazine* (53) 2015, afl. 12, p. 64-70.

Holstein, Dodig-Crnkovic & Pelliccione 2018

T. Holstein, G. Dodig-Crnkovic & P. Pelliccione, *Ethical and Social Aspects of Self-Driving Cars*, 5 februari 2018, arXiv:1802.04103v1.

Holtmaat & Rodrigues 2015

H.M.T. Holtmaat & P.R. Rodrigues, 'Discriminatie bezien vanuit juridisch perspectief. Naar een holistische benadering?', in: M. Davidović & A. Terlouw (red.), *Diversiteit en discriminatie. Onderzoek naar processen van in- en uitsluiting*, Amsterdam: Amsterdam University Press 2015, p. 125-148.

Huang e.a., *Scientific Reports* 2018, p. 1-8

C. Huang e.a., 'Machine learning predicts individual cancer patient response to therapeutic drugs with high accuracy', *Scientific Reports* (8) 2018/16444, p. 1-8.

ICO 2017

Information Commissioner's Office, *Big data, artificial intelligence, machine learning and data protection*, 2017, version 2.2, ico.org.uk.

Ingelse 2010

P. Ingelse, 'Rechter: tussen persoon en instituut', in: K.M. van Hassel & M.P. Nieuwe Weme (red.), *Willems' wegen: opstellen aangeboden aan prof. mr. J.H.M. Willems*, Deventer: Kluwer 2010.

***Resolution on Data Protection in Automated and Connected Vehicles* 2017**

Resolution on Data Protection in Automated and Connected Vehicles, 39th International Conference of Data Protection and Privacy Commissioners, september 2017.

International Working Group on Data Protection in Telecommunications 2018

International Working Group on Data Protection in Telecommunications, *Connected Vehicles*, Working paper, 63rd meeting, Budapest, 9-10 April 2018.

Jaarverslag 2018 nationale ombudsman

Jaarverslag 2018: iedereen moet mee kunnen doen, Den Haag: Nationale ombudsman 2019.

Jadaan, Zeater & Abukhalil, *Procedia Engineering* 2017, p. 641-648

K. Jadaan, S. Zeater & Y. Abukhalil, 'Connected Vehicles: An Innovative Transport Technology', *Procedia Engineering* (187) 2017, p. 641-648.

Jak & Bastiaans, *NJB* 2018, p. 3018-3025

N. Jak & S. Bastiaans, 'De betekenis van de AVG voor geautomatiseerde besluitvorming door de overheid', *NJB* 2018/2102, afl. 40, p. 3018-3025.

Van der Jagt 2013

F. van der Jagt, 'Het recht op bescherming van persoonsgegevens', in: J.H. Gerards (red.), *Grondrechten. De nationale, Europese en internationale dimensie*, Nijmegen : Ars Aequi Libri 2013 p. 164-183.

Jeltes, cursor.tue.nl 13 september 2018

T. Jeltes, 'Proef met zelfrijdende auto op campus', cursor.tue.nl 13 september 2018.

Jurgens, Chandrasekharan & Hemphill 2019

D. Jurgens, E. Chandrasekharan & L. Hemphill, *A Just and Comprehensive Strategy for Using NLP to Address Online Abuse* (Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics), 2019, p. 3658–3666.

Kamarinou, Millard & Singh 2016

D. Kamarinou, C. Millard & J. Singh, 'Machine Learning with Personal Data: Profiling, Decisions and the EU General Data Protection Regulation' (Queen Mary School of Law Legal Studies Research Paper No. 247/2016).

Katsh & Rifkin 2001

M.E. Katsh & J. Rifkin, *Online Dispute Resolution: Resolving Conflicts in Cyberspace*, San Francisco: Jossey-Bass 2001.

Katz, Bommarito & Blackman, *PloS ONE* 2017

D.M. Katz, M.J. Bommarito & J. Blackman, 'A general approach for predicting the behavior of the Supreme Court of the United States', *PloS ONE* (12) 2017, afl. 4.

Kauffman & Knowlton 2018

B.K.T. Kauffman & N.A. Knowlton, *Redefining case management* (Rapport Institute for the Advancement of the American Legal System), Denver: IAALS, april 2018.

Kaye 2018

D. Kaye, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression* (A/HRC/38/35, United Nations), 6 april 2018.

Keane, *Netherlands Quarterly of Human Rights* 2007, p. 641-663

D. Keane, 'Attacking Hate Speech under Article 17 of the European Convention on Human Rights', *Netherlands Quarterly of Human Rights* (25) 2007, afl. 4, p. 641-663.

Khademi e.a. 2019

A. Khademi e.a., 'Fairness in Algorithmic Decision Making: An Excursion Through the Lens of Causality', 27 maart 2019, arXiv:1903.11719.

Kirilenko e.a., *The Journal of Finance* 2017, p. 967-998

A. Kirilenko e.a., 'The Flash Crash: High-Frequency Trading in an Electronic Market', *The Journal of Finance* (72) 2017, afl. 3, p. 967-998.

Kleinberg, Mullainathan & Raghavan 2016

J. Kleinberg, S. Mullainathan & M. Raghavan, 'Inherent Trade-Offs in the Fair Determination of Risk Scores', 17 november 2016, arXiv:1609.05807.

Kleinberg e.a., *Journal of Legal Analysis* 2018, p. 113-174

J. Kleinberg e.a., 'Discrimination in the Age of Algorithms', *Journal of Legal Analysis* (10) 2018, p. 113-174.

Klitou 2012

D. Klitou, *Privacy-Invading Technologies: Safeguarding Privacy, Liberty & Security in the 21st Century* (diss. Leiden), Leiden: Universiteit Leiden 2012.

Kok, automobielmanagement.nl 4 september 2019

R. Kok, 'CBR ontwikkelt rijbewijs voor zelfrijdende auto's', automobielmanagement.nl 4 september 2019.

Kool e.a. 2017

L. Kool e.a., *Opwaarderen – Borgen van publieke waarden in de digitale samenleving.* Den Haag, Rathenau Instituut 2017

Koops e.a., *University of Pennsylvania Journal of International Law* 2017, p. 483-575

B.-J. Koops e.a., 'A Typology of Privacy', *University of Pennsylvania Journal of International Law* (38) 2017, afl. 2, p. 483-575.

Kortmann e.a. 2016

C.A.J.M. Kortmann e.a., *Constitutioneel Recht*, Deventer: Kluwer 2016.

Kourou e.a., *Computational and Structural Biotechnology Journal* 2015, p. 8-17

K. Kourou e.a., 'Machine learning applications in cancer prognosis and prediction', *Computational and Structural Biotechnology Journal* (13) 2015, p. 8-17.

KPMG 2019

KPMG, *Autonomous Vehicles Readiness Index. Assessing countries' preparedness for autonomous vehicles*, KPMG International 2019, assets.kpmg/content/dam/kpmg/xx/pdf/2019/02/2019-autonomous-vehicles-readiness-index.pdf.

Kramer 2016

X.E. Kramer, 'Access to Justice and Technology: Transforming the Face of Cross-Border Civil Litigation and Adjudication in the EU', in: K. Benyekhlef e.a. (red.), *eAccess to Justice*, Ottawa: University of Ottawa Press 2016, p. 351-375.

Kranenborg & Verhey 2018

H.R. Kranenborg & L.F.M. Verhey, *De Algemene Verordening Gegevensbescherming in Europees en Nederlands perspectief* (Mastermonografieën staats- en bestuursrecht), Deventer: Kluwer 2018.

Kreulen, *Trouw* 26 augustus 2019

E. Kreulen, 'Juridische databank laat zien: zoveel rechters, zoveel uitspraken', *Trouw* 26 augustus 2019, trouw.nl/nieuws/juridische-databank-laat-zien-zoveel-rechters-zoveel-uitspraken~bff1c576/.

Kulk 2019

S. Kulk, *Internet Intermediaries and Copyright Law: EU and US Perspectives*, Alphen aan den Rijn: Kluwer Law International 2019.

Kulk & Van Loenen, *International Journal of Spatial Data Infrastructures Research* 2012, p. 196-206

S. Kulk & B. van Loenen, 'Brave New Open Data World?', *International Journal of Spatial Data Infrastructures Research* (7) 2012, afl. 1, p. 196-206.

Lammerant, Blok & De Hert, *NTM/NJCM-bulletin* 2018

H. Lammerant, P.H. Blok & P. De Hert, 'Big data besluitvormingsprocessen en sluipwegen van discriminatie', *NTM/NJCM-bulletin* (43) 2018, afl. 1, p. 3-24.

Lamprecht, *Emerce* 5 september 2019

A. Lamprecht, 'Zijn algoritmes de oplossing om discriminatie in vacatureteksten te voorkomen?', *Emerce*.nl 5 september 2019.

Laney 2001

D. Laney, '3D Data Management: Controlling Data Volume, Velocity and Variety', *META Delta* 6 februari 2001, blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Langbroek en Rijpkema 2004
P.M. Langbroek & P.P. Rijpkema (red.), *Ombudsprudentie: over de behoorlijkheidsnorm en zijn toepassing*, Den Haag: Bju 2004.

Langbroek & Rijpkema 2007
P.M. Langbroek & P.P. Rijpkema, 'Ombudsprudentie in ontwikkeling', in: A. Brenninkmeijer (red.), *Werken aan behoorlijkheid. De Nationale ombudsman in zijn context*, Den Haag: Bju 2007, p. 269-297.

Larus e.a. 2018
J. Larus e.a., *When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making* (Technical Report Informatics Europe & EUACM), januari 2018.

Lastdrager, tno.nl 27 februari 2019
E. Lastdrager, 'Duurzame mobiliteit: economische groei, maatschappelijke impact', tno.nl 27 februari 2019.

Lavrijssen & Weitering, *VR* 2019, p. 167-171
N. Lavrijssen & M. Weitering, 'De zelfrijdende auto en het overmachtsverweer van art. 185 WVW', *VR* 2019/64, afl. 5, p. 167-171.

LeCun, Bengio & Hinton, *Nature* 2015, p. 436-444
Y. LeCun, Y. Bengio & G. Hinton, 'Deep learning', *Nature* (521) 2015, afl. 7553, p. 436-444.

Lester & Pachamanova, *UCLA Entertainment Law Review* 2017, p. 51-73
T. Lester & D. Pachamanova, 'The Dilemma of False Positives: Making Content ID Algorithms more Conducive to Fostering Innovative Fair Use in Music Creation', *UCLA Entertainment Law Review* (24) 2017, afl. 1, p. 51-73.

Von der Leyen 2019
U. von der Leyen, *A Union that strives for more. My agenda for Europe. Political guidelines for the the next European Commission 2019-2024*, 9 oktober 2019, ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf.

Li & Lyu 2019
Y. Li & S. Lyu, 'Exposing DeepFake Videos By Detecting Face Warping Artifacts', 22 mei 2019, arXiv:1811.00656.

Lim & Taeihagh, *Energies* 2018, p. 1062-1085

H.S.M. Lim & A. Taeihagh, 'Autonomous Vehicles for Smart and Sustainable Cities: An In-Depth Exploration of Privacy and Cybersecurity Implications', *Energies* (11) 2018, afl. 5, p. 1062-1085.

Lin 2015

P. Lin, 'Why Ethics Matters for Autonomous Cars', in: M. Maurer e.a. (red.), *Autonomes Fahren. Technische, rechtliche und gesellschaftliche Aspekte*, Berlijn/Heidelberg: Springer 2015, p. 69-85.

Lindenbergh, in: *T&C BW* 2019

S.D. Lindenbergh, commentaar op art. 6:162 BW, in: H.B. Krans, C.J.J.M. Stolker & W.L. Valk (red.), *Tekst & Commentaar Burgerlijk Wetboek*, Deventer: Wolters Kluwer 2019 (online, bijgewerkt 27 januari 2020).

Liu 2016

H.-Y. Liu, 'Structural Discrimination and Autonomous Vehicles: Immunity Devices, Trump Cards and Crash Optimisation', in: J. Seibt, M. Nørskov & S. Schack Andersen (red.), *What Social Robots Can and Should Do*, Amsterdam: IOS Press 2016, p. 164-173.

Live brengen analytics 2017

CJIB, Live brengen analytics – Risicoprofielen in het gerechtsdeurwaardertraject bij Mulderboetes, 28 februari 2017 (document niet openbaar).

Lodder, *Information & Communications Technology Law* 2006, p. 143-155

A.R. Lodder, 'The Third Party and Beyond. An Analysis of the Different Parties, in particular The Fifth, Involved in Online Dispute Resolution', *Information & Communications Technology Law* (15) 2006, afl. 2, p. 143-155.

Lodder e.a. 2014

A.R. Lodder, N.S. van der Meulen, T.H.A. Wisman, L. Meij, & C.M.M. Zwinkels, *Big Data, Big Consequences? Een verkenning naar privacy en big data gebruik binnen de opsporing, vervolging en rechtspraak*. Amsterdam: WODC - Vrije Universiteit 2014.

López & López 2017

C.A. López & R.M. López, 'Hate Speech in the Online Setting', in S. Assimakopoulos, F.H. Baider & S. Millar (red.), *Online Hate Speech in the European Union. A Discourse-Analytic Perspective*, Cham (Zwitserland): Springer 2017, p. 10-12.

MacAvaney e.a., *PloS ONE* 2019

S. MacAvaney e.a., 'Hate speech detection: Challenges and solutions', *PloS ONE* (14) 2019, afl. 8.

Makridakis, *Futures* 2017, p. 46-60

S. Makridakis, 'The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms', *Futures* (90) 2017, p. 46-60.

Mantelero, *Computer Law and Security Review* 2014, p. 643-660

A. Mantelero, 'The future of consumer data protection in the E.U. Re-thinking the "notice and consent" paradigm in the new era of predictive analytics', *Computer Law and Security Review* (30) 2014, afl. 6, p. 643-660.

Marseille & Tolsma 2019

A.T. Marseille & H.D. Tolsma (red.), *Bestuursrecht 2. Rechtsbescherming tegen de overheid*, Den Haag: Boom juridisch 2019.

Marshall, WIRED.com 31 maart 2018

A. Marshall, 'The Uber Crash Won't Be the Last Shocking Self-Driving Death', WIRED.com 31 maart 2018.

Martens 2014

M.H. Martens, *ITS en human factors: op de grens tussen mens en techniek* (oratie Twente), Enschede: Universiteit Twente 2014.

Massaro, *William and Mary Law Review* 1991, p. 211-266

T.M. Massaro, 'Equality and Freedom of Expression: The Hate Speech Dilemma', *William and Mary Law Review* (32) 1991, afl. 2, p. 211-266.

Matsakis, WIRED.com 22 maart 2018

L. Matsakis, 'A Window Into How YouTube Trains AI To Moderate Videos', WIRED.com 22 maart 2018.

McGonagle 2013

T. McGonagle, *The Council of Europe against online hate speech: Conundrums and challenges* (Expert Paper for the Council of Europe Conference of Ministers responsible for Media and Information Society, 'Freedom of Expression and Democracy in the Digital Age: Opportunities, Rights, Responsibilities', MCM(2013)005), Belgrado 2013.

McGonagle e.a. 2018

T. McGonagle e.a., *Inventarisatie methodes om "nepnieuws" tegen te gaan*, Amsterdam: IviR 2018.

Meuwese 2017

A.C.M. Meuwese, 'Grip op normstelling in het datatijdperk', in: W.J.M. Voermans, R.J.B. Schutgens & A.C.M. Meuwese, *Algemene regels in het bestuursrecht*, Den Haag: Boom juridisch 2017, p. 145-179.

De Meij e.a., *Mediaforum* 2006, p. 121-142

J.M. de Meij e.a., 'Toegang tot rechtelijke uitspraken: rapport van de VMC-studiecommissie Openbaarheid van rechtspraak', *Mediaforum* (18) 2006, afl. 4, p. 121-142.

Mendelts 2002

P. Mendelts, *Interpretatie van grondrechten. Grondrechtenclaims en verschuivingen in de reikwijdte van grondrechten* (diss. Utrecht), Deventer: Tjeenk Willink 2002.

Mendoza & Bygrave 2017

I. Mendoza & L.A. Bygrave, 'The Right Not to be Subject to Automated Decisions Based on Profiling', in: T.-E. Synodinou e.a. (red.), *EU Internet Law*, Berlijn/Heidelberg: Springer International Publishing 2017, p. 77-98.

De Meulder & Yildirim 2018

A. de Meulder & M. Yildirim, *Onderzoek invordering schulden door rijksoverheidsorganisaties*, Atos Consulting 2018, bijlage bij *Kamerstukken II* 2017/18, 24515, nr. 446.

Meyers West, *New Media & Society* 2018, p. 4366-4383

S. Meyers West, 'Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms', *New Media & Society* (20) 2018, afl. 11, p. 4366-4383.

Michiels, Blomberg & Jurgens 2016

F.C.M.A. Michiels, A.B Blomberg & G.T.J.M. Jurgens, *Handhavingsrecht* (Handboeken staats- en bestuursrecht), Deventer: Kluwer 2016.

Milakis, Van Arem & Van Wee, *Journal of Intelligent Transportation Systems* 2017, p. 324-348

D. Milakis, B. van Arem & B. van Wee, 'Policy and society related implications of automated driving: A review of literature and directions for future research', *Journal of Intelligent Transportation Systems* (21) 2017, afl. 4, p. 324-348.

Ministerie van JenV, Uitvoeringstoets 2019

Ministerie van JenV, *Telefonisch Innen, Uitvoeringstoets*, 26 april 2019 (document niet openbaar).

Mishra e.a. 2018

P. Mishra e.a., *Author Profiling for Abuse Detection* (Proceedings of the 27th International Conference on Computational Linguistics), Association for Computational Linguistics 2018, p. 1088–1098.

Mitrou 2019

L. Mitrou, *Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR) 'Artificial Intelligence-Proof'?*, SSRN 3 juni 2019, doi.org/10.2139/ssrn.3386914.

Mommers, Zwenne & Schermer, *NJB* 2010, p. 2072-2078

L. Mommers, G.-J. Zwenne & B. Schermer, 'Het best bewaarde geheim van de raadkamer. Over de ontoegankelijkheid van de rechtspraak', *NJB* 2010/1692, afl. 32, p. 2072-2078.

Mondal, Silva & Benevenuto 2017

M. Mondal, L.A. Silva & F. Benevenuto, *A Measurement Study of Hate Speech in Social Media* (Proceedings of the 28th ACM Conference on Hypertext and Social Media), 2017, p. 85-94, doi.org/10.1145/3078714.3078723.

Mozur, *The New York Times* 15 oktober 2018

P. Mozur, 'A Genocide Incited on Facebook, With Posts From Myanmar's Military', *The New York Times* 15 oktober 2018, nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.

Mulder, Klingenberg & Mifsud Bonnici 2017

T. Mulder, A.M. Klingenberg & G.P. Mifsud Bonnici, *Datamining en privacybescherming: Onderzoek voor het CJIB*, Groningen: Rijksuniversiteit Groningen 2017 (document niet openbaar).

Müller & Schwarz 2018

K. Müller & C. Schwarz, *Fanning the Flames of Hate: Social Media and Hate Crime* (Centre for Competitive Advantage in the Global Economy Working Paper Series nr. 373), University of Warwick 2018, warwick.ac.uk/fac/soc/economics/research/centres/cage/manage/publications/373-2018_schwarz.pdf.

Munger, *Political Behavior* 2017, p. 629-649

K. Munger, 'Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment', *Political Behavior* (39) 2017, afl. 3, p. 629-649.

Narayanan & Shmatikov 2008

A. Narayanan & V. Shmatikov, *Robust De-anonymization of Large Sparse Datasets* (Proceedings of the IEEE Symposium on Security and Privacy), 2008, p. 111-125, doi.org/10.1109/SP.2008.33.

Nguyen e.a. 2019

T.T. Nguyen e.a., *Deep Learning for Deepfakes Creation and Detection*, 25 september 2019, arXiv:1909.11573.

Nikitas e.a., *Urban Science* 2017

A. Nikitas e.a., 'How Can Autonomous and Connected Vehicles, Electromobility, BRT, Hyperloop, Shared Use Mobility and Mobility-As-A-Service Shape Transport Futures for the Context of Smart Cities?', *Urban Science* (1) 2017, afl. 4, doi.org/10.3390/urbansci1040036.

Nissenbaum, *Science and Engineering Ethics* 1996, p. 25-42

H. Nissenbaum, 'Accountability in a computerized society', *Science and Engineering Ethics* (2) 1996, afl. 1, p. 25-42.

Van Noort, *NRC* 13 februari 2018

W. van Noort, 'Unilever eist meer actie van big tech', *NRC*.nl 13 februari 2018.

Noy & Givoni, *Sustainability* 2018

K. Noy & M. Givoni, 'Is 'Smart Mobility' Sustainable? Examining the Views and Beliefs of Transport's Technological Entrepreneurs', *Sustainability* (10) 2018, afl. 2, doi.org/10.3390/su10020422.

***NVvR-rechterscode* 2011**

Nederlandse Vereniging voor Rechtspraak, *NVvR-rechterscode* 2011, nvvr.org/uploads/documenten/nvvr-rechterscode.pdf.

Ohm, *UCLA Law Review* 2010, p. 1701-1777

P. Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization', *UCLA Law Review* (57) 2010, p. 1701-1777.

Van Orshoven 2001

P. van Orshoven, 'De onafhankelijkheid van de rechter naar Belgisch recht', in: P. van Orshoven, L.F.M. Verhey & K. Wagner, *De onafhankelijkheid van de rechter. Preadviezen voor de Vereniging voor de vergelijkende studie van het recht van België en Nederland*, Deventer: W.E.J. Tjeenk Willink 2001, p. 76-121.

Pater e.a. 2016

J.A. Pater e.a., *Characterizations of Online Harassment: Comparing Policies Across Social Media Platforms* (Proceedings of the 19th International Conference on Supporting Group Work), 2016, p. 369–374, doi.org/10.1145/2957276.2957297.

Van der Parre, nos.nl 9 augustus 2019

H. van der Parre, 'Zelfrijdende auto kan gedrag van voetgangers voorspellen' nos.nl 9 augusutus 2019.

Parkinson e.a., *IEEE Transactions on Intelligent Transportation Systems* 2017, p. 2898-2915

S. Parkinson e.a., 'Cyber Threats Facing Autonomous and Connected Vehicles: Future Challenges', *IEEE Transactions on Intelligent Transportation Systems* (18) 2017, afl. 11, p. 2898-2915.

Pavlopoulos, Malakasiotis & Androutsopoulos 2017

J. Pavlopoulos, P. Malakasiotis & I. Androutsopoulos, *Deeper Attention to Abusive User Content Moderation* (Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing) 2017, p. 1125-1135, doi.org/10.18653/v1/D17-1117.

Pearl, *Communications of the ACM* 2019, p. 54-60

J. Pearl, 'The Seven Tools of Causal Inference, with Reflections on Machine Learning', *Communications of the ACM* (62) 2019, afl. 3, p. 54-60.

Pennachin & Goertzel 2007

C. Pennachin & B. Goertzel 2007, 'Contemporary Approaches to Artificial General Intelligence', in: B. Goertzel & C. Pennachin (red.), *Artificial General Intelligence*, Berlijn: Springer 2007, p. 1-30.

Peppet, *Texas Law Review* 2014, p. 85-176

S.R. Peppet, 'Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent', *Texas Law Review* (93) 2014, p. 85-176.

Philipsen & Themeli, *Rechtstreeks* 2019, p. 46-49

S. Philipsen & E. Themeli, 'Een introductie op de robotrechter', *Rechtstreeks* 2019, afl. 2, p. 46-49.

Philipsen & Themeli 2020

S. Philipsen & E. Themeli, 'AI as the Court: Assessing AI deployment in Civil Cases', in: *GLAIC conference proceedings 2020* (*nog te verschijnen*).

Pleiss e.a. 2017

G. Pleiss e.a., *On Fairness and Calibration*, 3 november 2017, arXiv:1709.02012.

Van de Poel e.a., *Science and Engineering Ethics* 2012, p. 49-67

I. van de Poel e.a., 'The Problem of Many Hands: Climate Change as an Example', *Science and Engineering Ethics* (18) 2012, afl. 1, p. 49-67.

Position paper CJIB 2018

Position paper CJIB ten behoeve van het rondetafelgesprek 'De overheid als schuldeiser' op 1 oktober 2018, CJIB 2018, tweedekamer.nl/debat_en_vergadering/commissievergaderingen/details?id=2018A03389#.

Post 2009

R. Post, 'Hate Speech', in: I. Hare & J. Weinstein (red.), *Extreme Speech and Democracy*, New York: Oxford University Press 2009, p. 123-138.

Prakken, *NJB* 2018, p. 269-274

H. Prakken, 'Komt de robotrechter er aan?', *NJB* 2018/207, afl. 4, p. 269-274.

Prins, *NJBlog* 29 september 2015

C. Prins, 'Heeft U nog de controle over uw auto?', *NJBlog* 29 september 2015.

Prins & Van der Roest, *NJB* 2018, p. 260-268

C. Prins & J. van der Roest, 'AI en de rechtspraak. Meer dan alleen de 'robotrechter'', *NJB* 2018/206, afl. 4, p. 260-268.

Van der Put, *Rechtstreeks* 2019, p. 50-60

M. van der Put, 'Kan artificiële intelligentie de rechtspraak betoveren?', *Rechtstreeks* 2019, afl. 2, p. 50-60.

Qu, towardsdatascience.com 22 oktober 2018

J. Qu, 'Training Self Driving Cars using Reinforcement Learning', towardsdatascience.com 22 oktober 2018.

Raiyn, *Transport and Telecommunication* 2018, p. 325-334

J. Raiyn, 'Data and Cyber Security in Autonomous Vehicle Networks', *Transport and Telecommunication* (19) 2018, afl. 4, p. 325-334.

Rammert 2008

W. Rammert, *Where the action is: Distributed agency between humans, machines, and programs* (The Technical University Technology Studies Working Papers) 2008.

Rapport Decisio & DSP 2019

Onderzoek naar het gebruik van betaalprofielen (rapport Decisio & DSP), bijlage bij *Kamerstukken II* 2019/20, 24515, nr. 496.

Rapport nationale ombudsman 2019

Invorderen vanuit het burgerperspectief: onderzoek naar knelpunten die burgers ervaren bij het invorderen van schulden door de overheid (rapport nationale ombudsman 2019/005), Den Haag: Nationale ombudsman 2019.

Regan, *Information, Communication & Society* 2002, p. 382-405

P.M. Regan, 'Privacy as a Common Good in the Digital World', *Information, Communication & Society* (5) 2002, afl. 3, p. 382-405.

Reiling 2009

A.D. Reiling, *Technology for Justice. How Information Technology can support Judicial Reform*, Leiden: Leiden University Press 2009.

Reiling, *Computerrecht* 2020, p. 40-45

A.D. Reiling, 'De rechtspraktijk: toepassing van AI in de rechtspraak', *Computerrecht* 2020/6, afl. 2, p. 40-45

Van Rhee 2018

G. van Rhee, *Samen werken aan recht en veiligheid. Rechtspraak in zwaar weer* (Brief van 14 september 2018 aan de minister voor Rechtsbescherming en de leden van de vaste commissie voor Justitie & Veiligheid van de Eerste en Tweede Kamer), Den Haag: NVvR 2018, nvvr.org/uploads/afbeeldingen/20180914-brief-TK-en-MvRB-over-rechtspraak.pdf.

Ribeiro, Singh & Guestrin 2016

M.T. Ribeiro, S. Singh & C. Guestrin, *"Why should I trust you?" Explaining the Predictions of Any Classifier* (Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining), 2016, p. 1135-1144, doi.org/10.1145/2939672.2939778.

Roberts 2014

S.T. Roberts, *Behind the screen: the hidden digital labor of commercial content moderation* (diss. Illinois Urbana-Champaign), 2014.

Roberts 2016

S.T. Roberts, 'Commercial Content Moderation: Digital Laborers' Dirty Work', in: S.U. Noble & B.M. Tynes (red.), *The Intersectional Internet: Race, Sex, Class and Culture Online*, New York: Peter Lang Publishing 2016.

Roberts, *VICE* 2 mei 2018

M. Roberts, 'Instagram Is Using AI to Filter Out Toxic Comments', *VICE*.com 2 mei 2018.

Roberts 2019

S.T. Roberts, 'Content Moderation', in: L.A. Schintler & C.L. McNeely (red.), *Enyclopedia of Big Data*, Cham (Zwitserland): Springer International Publishing 2019.

Roff, brookings.edu 7 december 2018

H.M. Roff, 'The folly of trolleys: Ethical challenges and autonomous vehicles', brookings.edu 7 december 2018

Rosenfeld, *Cardozo Law Review* 2003, p. 1523-1567

M. Rosenfeld, 'Hate Speech in Constitutional Jurisprudence: A Comparative Analysis', *Cardozo Law Review* (24) 2003, afl. 4, p. 1523-1567.

Rössler e.a. 2018

A. Rössler e.a., *FaceForensics: A Large-scale Video Dataset for Forgery Detection in Human Faces*, 24 maart 2018, arXiv:1803.09179.

Rudin, *Nature Machine Intelligence* 2019, p. 206-215

C. Rudin, 'Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead', *Nature Machine Intelligence* (1) 2019, p. 206-215.

Russell & Norvig 2010

S.J. Russell & P. Norvig, *Artificial Intelligence. A Modern Approach*, Upper Saddle River (VS): Prentice Hall 2010.

Santos e.a. 2018

C.N. Santos, I. Melnyk & I. Padhi, *Fighting Offensive Language on Social Media with Unsupervised Text Style Transfer*, 20 mei 2018, arXiv:1805.07685v1.

Sap e.a. 2019

M. Sap e.a., *The Risk of Racial Bias in Hate Speech Detection* (Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics), 2019, p. 1668-1678.

SAPAI 2019

Strategisch Actieplan voor Artificiële Intelligentie (Beleidsnota), bijlage bij *Kamerstukken II* 2019/20, 26643, nr. 640.

Schapire 2003

R. Schapire, *COS 511: Foundations of Machine Learning* (Lecture notes van lezing 1 van 4 februari 2003 in het kader van de cursus Computer Science aan Princeton University), 2003, cs.princeton.edu/courses/archive/spr03/cs511/scribe_notes/0204.pdf.

Schellekens, *Computer Law & Security Review* 2016, p. 307-315

M. Schellekens, 'Car hacking: Navigating the regulatory landscape', *Computer Law & Security Review* (32) 2016, afl. 2, p. 307-315.

Schermer 2013

B.W. Schermer, 'Risks of Profiling and the Limits of Data Protection Law', in: B. Custers e.a. (red.), *Discrimination and Privacy in the Information Society. Data Mining and Profiling in Large Databases*, Berlijn/Heidelberg: Springer 2013, p. 137-152.

Schermer, Hagenauw & Falot 2018
B.W. Schermer, D. Hagenauw & N. Falot, *Handleiding Algemene verordening gegevensbescherming en Uitvoeringswet Algemene verordening gegevensbescherming*, Den Haag: Ministerie van Justitie en Veiligheid, 22 januari 2018, rijksoverheid.nl/avg.

Schmidt & Wiegand 2017
A. Schmidt & M. Wiegand, *A Survey on Hate Speech Detection using Natural Language Processing* (Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media), Association for Computational Linguistics 2017, p. 1-10, doi.org/10.18653/v1/W17-1101.

Schölkopf 2019
B. Schölkopf, *Causality for Machine Learning*, 23 december 2019, arXiv:1911.10500.

Schoneveld e.a. 2018
D.-J. Schoneveld e.a., *Effecten van bijzondere incassobevoegdheden en overheidspreferenties* (Eindrapport Berenschot van 4 mei 2018), bijlage bij *Kamerstukken II* 2017/18, 24515, nr. 446.

Schoonen, Trouw.nl 28 april 2019
W. Schoonen, 'Oorzaak en gevolg: is het de haan die de zon oproept?', Trouw.nl 28 april 2019.

Schreuder, *AV&S* 2014, p. 131-136
A.I. Schreuder, 'Aansprakelijkheid voor 'zelfdenkende' apparatuur', *AV&S* 2014/20, afl. 5/6, p. 131-136.

Schuurmans 2015
Y.E. Schuurmans, 'Van bestuursrechtelijke detailhandel naar maakindustrie' (oratie Leiden), Leiden: Universiteit Leiden 2015.

Sebastian, *Feminist Media Studies* 2017, p. 1107-1111
M. Sebastian, 'Privacy and consent: the trouble with the label of "revenge porn"', *Feminist Media Studies* (17) 2017, afl. 6, p. 1107–1111.

Shoham, *Communications of the ACM* 2016, p. 47-49
Y. Shoham, 'Why Knowledge Representation Matters', *Communications of the ACM* (59) 2016, afl. 1, p. 47-49.

Sivakumar, Gordo & Paluri, engineering.fb.com 11 september 2018
V. Sivakumar, A. Gordo, M. Paluri, 'Rosetta: Understanding text in images and videos with machine learning', engineering.fb.com 11 september 2018

Spronken, *NJB* 2018, p. 791
T. Spronken, 'A court with no face and no place', *NJB* 2018/586, afl. 12, p. 791.

Stilgoe, *Social Studies of Science* 2018, p. 25-56
J. Stilgoe, 'Machine learning, social learning and the governance of self-driving cars', *Social Studies of Science* (48) 2018, afl. 1, p. 25-56.

Stolk, Boot & Spanninga, montesquieu-instituut.nl 26 november 2018
R. Stolk, J. Boot & H. Spanninga, 'De data beslissen? Een kwestie van waarden!', montesquieu-instituut.nl 26 november 2018

Surden, *Georgia State University Law Review* 2019, p. 1305-1337
H. Surden, 'Artificial Intelligence and Law: An Overview', *Georgia State University Law Review* (35) 2019, afl. 4, p. 1305-1337.

Surden & Williams, *Cardozo Law Review* 2016, p. 121-181
H. Surden & M.-A. Williams, 'Technological Opacity, Predictability, and Self-Driving Cars', *Cardozo Law Review* (38) 2016, afl. 1, p. 121-181.

Susskind 2019
R. Susskind, *Online Courts and the Future of Justice*, Oxford: Oxford University Press 2019.

Suzor e.a., *International Journal of Communication* 2019, p. 1526-1543
N.P. Suzor e.a., 'What Do We Mean When We Talk About Transparency? Toward Meaningful Transparency in Commercial Content Moderation', *International Journal of Communication* (13) 2019, p. 1526-1543.

Sweeney 2000
L. Sweeney, *Simple Demographics Often Identify People Uniquely* (Data Privacy Working Paper 3), Carnegie Mellon University 2000

Theil, *Verfassungsblog* 8 februari 2018
S. Theil, 'The German NetzDG: A Risk Worth Taking?', *Verfassungsblog* 8 februari 2018.

Tichelaar, *TAV* 2018, p. 29-33

B.E. Tichelaar, 'Aansprakelijkheidsvraagstukken bij zelfrijdende auto's', *TAV* 2018/24, afl. 1, p. 29-33.

Timmer & Kool 2014

J. Timmer & L. Kool (red.), *Tem de robotauto - De zelfsturende auto voor publieke doelen*, Den Haag: Rathenau Instituut 2014.

Titley, Keen & Földi 2014

G. Titley, E. Keen & L. Földi, *Starting points for combating hate speech online*, Council of Europe 2014.

Tjong Tjin Tai & Boesten, *NJB* 2016, p. 656-664

E. Tjong Tjin Tai & S. Boesten, 'Aansprakelijkheid, zelfrijdende auto's en andere zelfbesturende objecten', *NJB* (91) 2016/496, afl. 10, p. 656-664.

Urban, Karaganis & Schofield 2016

J.M. Urban, J. Karaganis & B.L. Schofield, *Notice and Takedown in Everyday Practice*, BerkeleyLaw, University of California en The American Assembly of Columbia University 2016, illusionofmore.com/wp-content/uploads/2016/04/Berkeley_Columbia-on-512-takedown.pdf.

Uzman & Boogaard, *Overheid & Aansprakelijkheid* 2017, p. 63-70

J. Uzman & G. Boogaard, 'Onrechtmatige rechtspraak en de rechtsstaat', *Overheid & Aansprakelijkheid* 2017/30, afl. 2, p. 63-70.

Veale & Edwards, *Computer Law & Security Review* 2018, p. 398-404

M. Veale & L. Edwards, 'Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling', *Computer Law & Security Review* (34) 2018, afl. 2, p. 398-404.

Vedder 2000

A.H. Vedder, 'Discriminatiegronden in het informatietijdperk', in: R. Holtmaat (red.), *De toekomst van gelijkheid. De juridische en maatschappelijk inbedding van de gelijkebehandelingsnorm*, Deventer: Kluwer 2000, p. 161-175.

Veledar, Damjanovic-Behrendt & Macher 2019

O. Veledar, V. Damjanovic-Behrendt & G. Macher, 'Digital Twins for Dependability Improvement of Autonomous Driving', in: A. Walker, R. O'Connor & R. Messnarz (red.), *Systems, Software and Services Process Improvement. EuroSPI 2019. Communications in Computer and Information Science,* Cham (Zwitserland): Springer 2019, p. 415-426.

Vellinga, *VR* 2014, p. 370-377

N.E. Vellinga, 'De civielrechtelijke aansprakelijkheid voor schade veroorzaakt door een autonome auto', *VR* 2014/151, afl. 10, p. 370-377.

Vellinga, *VR* 2020, p. 66-71

N.E. Vellinga, 'Juridische aspecten van het gebruik van autonome auto's', *VR* 2020/42, afl. 3, p. 66-71.

Vellinga & Vellinga, *VR* 2015, p. 82-90

N.E. Vellinga & W.H. Vellinga, 'Enkele verkeersrechtelijke aspecten van toelating van (deels) zelfrijdende of autonome auto's tot het wegverkeer', *VR* 2015/35, afl. 3, p. 82-90.

Venice Commission 2012

Venice Commission, *Opinion on Act CLXII of 2011 on the legal status and remuneration of judges and Act CLXI of 2011 on the organisation and administration of courts of Hungary* (Opinion 663/2012, aangenomen door de Venice Commission op haar 90ste plenaire sessie, CDL-AD(2012)001), Straatsburg: 19 maart 2012.

Venice Commission 2019

Venice Commission, *Hungary Opinion on the law on administrative courts and on the law on the entry into force of the law on administrative courts and certain transitional rules* (Opinion 943/2018, aangenomen door de Venice Commission op haar 118de plenaire sessie, CDL-AD(2019)004), Straatsburg: 19 maart 2019.

Vetzo & Gerards, *Computerrecht* 2019, p. 10-19

M.J. Vetzo & J.H. Gerards, 'Algoritme-gedreven technologieën en grondrechten', *Computerrecht* 2019/3, afl. 1, p. 10-19.

Vetzo, Gerards & Nehmelman 2018

M.J. Vetzo, J.H. Gerards & R. Nehmelman, *Algoritmes en grondrechten*, Den Haag: Boom juridisch 2018.

De Vey Mestdagh & Lubbers, *AA* 2015, p. 267-280

C.N.J. de Vey Mestdagh & J. Lubbers, ‘‘Nee hoor, u wilt helemaal niet naar Den Haag…’ Over de techniek, het recht en de toekomst van de zelfrijdende auto’, *AA* (64) 2015, afl. 4, p. 267-280.

***Visie tenuitvoerlegging financiële sancties* 2015**

CJIB, *Visie tenuitvoerlegging financiële sancties – financiële claim*, 28 oktober 2015 (document niet openbaar).

***Visie op de rechtspraak* 2010**

Visie op de rechtspraak, Raad voor de rechtspraak, Den Haag, 24 mei 2010

Voigt & Von dem Bussche 2017

P. Voigt & A. von dem Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide,* Cham (Zwitserland): Springer International Publishing 2017.

Waaldijk 2005

C. Waaldijk, ‘Wanneer telt (on)gelijke behandeling van (on)gelijke gevallen als indirect onderscheid?’, in: S.D. Burri e.a. (red.), *Gelijke behandeling: oordelen en commentaa*r *2004*, Deventer: Kluwer 2005, p. 149-160.

Wachter, Mittelstadt & Floridi, *International Data Privacy Law* 2017, p. 76-99

S. Wachter, B. Mittelstadt & L. Floridi, ‘Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation’, *International Data Privacy Law* (7) 2017, afl. 2, p. 76-99.

Waldron 2012

J. Waldron, *The Harm in Hate Speech*, Cambridge: Harvard University Press 2012.

Walker Smith, cyberlaw.stanford.edu/blog 18 november 2013

B. Walker Smith, ‘Human error as a cause for vehicle crashes’, cyberlaw.stanford.edu/blog 18 november 2013

Warner & Hirschberg 2012

W. Warner & J. Hirschberg, *Detecting Hate Speech on the World Wide Web* (Proceedings of the Second Workshop on Language in Social Media), Association for Computational Linguistics 2012, p. 19-26.

Waseem 2016

Z. Waseem, *Are You a Racist or Am I Seeing Things? Annotator Influence on Hate Speech Detection on Twitter* (Proceedings of the First Workshop on NLP and Computational Social Science), Association for Computational Linguistics 2016, p. 138-142.

Washington, *Colorado Technology Law Journal* 2018, p. 131-160

A.L. Washington, 'How to Argue with an Algorithm: Lessons from the COMPAS-ProPublica Debate', *Colorado Technology Law Journal* (17) 2018, afl. 1, p. 131-160.

Weber, *Computer Law & Security Review* 2015, p. 618-627

R.H. Weber, 'Internet of things: Privacy issues revisited', *Computer Law & Security Review* (31) 2015, afl. 5, p. 618-627.

Van Wees, *AV&S* 2015, p. 170-180

K.A.P.C. van Wees, 'Aansprakelijkheidsaspecten van (deels) zelfrijdende auto's', *AV&S* 2015/28, afl. 5, p. 170-180.

Van Wees, *Maandblad voor Vermogensrecht* 2018, p. 112-122

K.A.P.C. van Wees, 'Voertuigautomatisering en productaansprakelijkheid', *Maandblad voor Vermogensrecht* 2018, afl. 4, p. 112-122.

Van de Weijer, *De Volkskrant* 30 november 2018

B. van de Weijer, 'Zo wordt de zelfrijdende auto steeds minder bijziend', *De Volkskrant* 30 november 2018.

Van de Weijer, *De Volkskrant* 31 oktober 2019

B. van de Weijer, 'Visioen van zelfrijdende auto is voorlopig verdampt', *De Volkskrant* 31 oktober 2019.

Van de Weijer, *De Volkskrant* 27 december 2019

B. van de Weijer, 'Zelfrijdende auto? Rij nog maar even zelf', *De Volkskrant* 27 december 2019.

Van de Weijer & Van der Bent, *De Volkskrant* 26 juni 2015

B. van de Weijer & I. van der Bent, 'Hoe de autonome auto alles anders maakt', *De Volkskrant* 26 juni 2015.

Weiß, *Computer Networks* 2011, p. 3103-3119

C. Weiß, 'V2X communication in Europe - From research projects towards standardization and field testing of vehicle communication technology', *Computer Networks* (55) 2011, afl. 14, p. 3103-3119.

Widdershoven 2011

R.J.G.M. Widdershoven, 'Het beginsel van effectieve rechtsbescherming', in: A. van den Brink, S.A. de Vries & M.P.A. de Koninck (red.), *Beginselen bouwen Burgerschap*, Den Haag: Boom juridisch 2011, p. 103-114.

Wilks 2019

Y. Wilks, *Artificial Intelligence. Modern Magic or Dangerous Future?*, Londen: Icon Books 2019.

Wilmink, tno.nl 17 juli 2018

I. Wilmink, 'Hoe gaan zelfrijdende auto's ons verkeer en vervoer veranderen?', tno.nl 17 juli 2018.

Wilson 2012

R. Wilson, *Indignity, indifference, indignation: tackling hate speech online* (Report of the conference "Tackling hate speech: Living together online" in Budapest, 27-28 November 2012), Council of Europe 2012.

Wilson, Hoffman & Morgenstern 2019

B. Wilson, J. Hoffman & J. Morgenstern, 'Predictive Inequity in Object Detection', arXiv:1902.11097.

WODC 2017

Wetenschappelijk Onderzoek- en Documentatiecentrum, *Second Opinion Beslisboom afloop deurwaarderstraject CJIB*, Den Haag: WODC 2017 (niet openbaar; samenvatting op wodc.nl).

World Economic Forum 2018

World Economic Forum, *Harnessing Artificial Intelligence for the Earth* (Fourth Industrial Revolution for the Earth Series), januari 2018, www3.weforum.org/docs/Harnessing_Artificial_Intelligence_for_the_Earth_report_2018.pdf.

Wood e.a., *Work, Employment and Society* 2019, p. 56-75

A.J. Wood e.a., 'Good Gig, Bad Gig: Autonomy and Algorithmic Control in the Global Gig Economy', *Work, Employment and Society* (33) 2019, afl. 1, p. 56–75.

WRR 2016

Wetenschappelijke Raad voor het Regeringsbeleid, *Big Data in een vrije en veilige samenleving* (rapport nr. 95), Den Haag/Amsterdam: WRR/Amsterdam University Press 2016.

WRR 2017

Wetenschappelijke Raad voor het Regeringsbeleid, *Weten is nog geen doen. Een realistisch perspectief op redzaamheid* (rapport nr. 97), Den Haag: WRR 2017.

Yang e.a., *Science China Technological Sciences* 2018, p. 1446-1471

D.G. Yang e.a., 'Intelligent and connected vehicles: Current status and future perspectives', *Science China Technological Sciences* (61) 2018, afl. 10, p. 1446-1471.

Yin 2009

R.K. Yin, 'How to Do Better Case Studies', in: L. Bickman & D.J. Rog (red.), *The SAGE Handbook of Applied Social Research Methods*, Los Angeles: Sage 2009, p. 254-282.

Zuleta & Burkal 2017

L. Zuleta & R. Burkal, *Hate Speech in the Public Online Debate*, Kopenhagen: Danish Institute for Human Rights 2017, humanrights.dk.

Zwenne, in: *T&C Privacy- en telecommunicatierecht* 2018

G.J. Zwenne, 'Commentaar op art. 22 AVG', in: G.J. Zwenne & P.C. Knol (red.), *Tekst & Commentaar Privacy- en telecommunicatierecht*, Deventer: Kluwer 2018 (online, bijgewerkt 1 december 2019).

Zwenne & Steenbruggen 2017

G.J. Zwenne & W.A.M. Steenbruggen, 'Privacyrisico's en -waarborgen bij het gebruik van big data tegen zorgfraude: een verkenning', in: L. Ottes e.a., *Big data in de zorg*. *Preadvies* (Vereniging voor Gezondheidsrecht), Den Haag: Sdu 2017, p. 71-99.

Over het onderzoek

Het onderzoek is uitgevoerd door een team van onderzoekers, waarbij sprake is geweest van een werkverdeling. Het team als geheel draagt de verantwoordelijkheid voor de inhoud van het rapport.

De eerste auteurs van de algemene hoofdstukken van het rapport (H. 1-3 en H. 8-10) zijn:
mr. dr. S. Kulk & mr. S. van Deursen

De casestudy's zijn primair geschreven door:
mr. dr. S. Kulk & T. Snijders LL.B (Casestudy Contentmoderatie door online platformen (H. 4))
mr. dr. V.E. Breemen & mr. A.H.H. Wouters (Casestudy Zelfrijdende auto's (H. 5))
mr. S. van Deursen & mr. dr. S. Philipsen (Casestudy De rechtspraak (H. 6))
mr. dr. I.M. Boekema & mr. dr. S.E. Heeger (Casestudy Overheidsincasso bij verkeers-boetes (H. 7))

Het onderzoek als geheel is uitgevoerd onder begeleiding van:
prof. mr. J.H. Gerards
prof. mr. E. Bauw
prof. mr. drs. M. de Cock Buning
prof. dr. mr. H. Prakken
mr. dr. N.R. Koffeman
prof. mr. A. Gerbrandy

Met coördinatie door:
mr. dr. N.R. Koffeman & mr. A.H.H. Wouters

Prof. mr. drs. De Cock Buning is kort voor de afronding van het onderzoek, half januari 2020, in dienst getreden bij Netflix. Ondanks dat haar rol in het onderzoek beperkt is geweest en zij vanaf december 2019 daaraan geen inhoudelijke bijdrage meer heeft geleverd, hechten de onderzoekers eraan te vermelden dat deze overstap op geen enkele wijze van invloed is geweest op de inhoud of conclusies van dit in gezamenlijkheid door het onderzoeksteam tot stand gebrachte rapport. Hoewel het onderzoek bovendien geen betrekking heeft op het door het genoemde bedrijf gebruikte systeem van aanbevelingen/kijktips, is in overleg met de begeleidingscommissie uit overwegingen van integriteit en zorgvuldigheid besloten om De Cock Buning tijdens de laatste bijeenkomst van de begeleidingscommissie de overstap te laten toelichten en is zij vervolgens niet bij de inhoudelijke bespreking van het eindrapport aanwezig geweest.