



Universiteit Utrecht

Bachelorthesis Kunstmatige Intelligentie
7,5 ECTS

Verantwoordelijkheidsinvloeden op bestuurders van zelfrijdende auto's

Auteur: Merel van den Bos (6501346)

Datum: 16 april 2021

Begeleider: dr. Baptist Liefoghe

Tweede lezer: dr. Aleksei Nazarov

Abstract

Ondanks dat deels zelfrijdende auto's al zijn toegestaan in Nederland, zijn er nog veel ongedefinieerde aspecten. De huidige wetgeving omtrent ongevallen met dergelijke voertuigen biedt te veel ruimte voor speling. Maar naast de rol van de overheid daarin, zijn ook eindfabrikant en gebruiker erg belangrijk in zowel aansprakelijkheid als morele verantwoordelijkheid. Uit dit onderzoek blijkt dat de invloeden op de subjectieve ervaring van verantwoordelijkheid in het verkeer met name bestaan uit controlevermindering, keuzevrijheid en gebalanceerd vertrouwen. Voor de toekomst van de volledig zelfrijdende auto's, die nu nog niet zijn toegestaan, is het van belang het verantwoordelijkheidsgevoel van bestuurders hoog te houden en in combinatie met een heldere wetgeving kunnen de determinanten op de juiste manier beïnvloed worden. Verder praktisch onderzoek moet uitwijzen of de speculatieve methoden die dit onderzoek biedt voor onder andere het vertrouwen in de technologie dusdanig volstaan voor genoeg verantwoordelijkheidsgevoel. Uiteindelijk moet er een juridisch en gebruiksvriendelijk kader ontstaan waarbinnen volledig zelfsturende auto's over een aantal decennia een vervanging kunnen zijn van de huidige auto's.

Kernwoorden: zelfsturende auto's, kunstmatige intelligentie, ethiek, morele verantwoordelijkheidsgevoel

Inhoudsopgave

Hoofdstuk 1 Inleiding	4
Hoofdstuk 2 Morele verantwoordelijkheid	6
2.1 <i>Betrokken partijen</i>	6
2.1.1 Verantwoordelijkheidsgroepen	6
2.1.2 Schuldig in ongevallen	6
2.2 <i>Juridische achtergrond [11]</i>	6
2.2.1 Wetgeving aansprakelijkheid.....	6
2.2.2 Wetgeving functionaliteiten	8
2.3 <i>De kant van de fabrikant</i>	8
2.4 <i>Conclusie</i>	8
Hoofdstuk 3 Determinanten van verantwoordelijkheidsgevoel	9
3.1 <i>Human-on-the-loop</i>	9
3.2 <i>Vertrouwen</i>	10
3.2.1 Vertrouwen in technologieën.....	10
3.2.2 Balans.....	11
3.3 <i>Autonomie</i>	11
3.4 <i>Meaningful Human Control</i>	12
3.5 <i>Conclusie</i>	14
Hoofdstuk 4 Verantwoordelijkheidsgevoel optimaliseren	15
4.1 <i>Vaardigheden</i>	15
4.2 <i>Vertrouwen</i>	15
4.2.1 Vertrouwen in technologieën.....	16
4.2.2 Kalibratie	17
4.3 <i>Meaningful Human Control</i>	17
Hoofdstuk 5 Conclusie & Discussie	19
Hoofdstuk 6 Bibliografie	21

Hoofdstuk 1 Inleiding

De auto is anno 2021 niet meer uit de maatschappij weg te denken. In 2010 werd zelfs tot bijna 50% van de tijd gekozen om met de auto te gaan en bezat bijna 25% van de huishoudens twee of meer auto's in Nederland [1]. Inmiddels, ruim 10 jaar later, zijn auto's met een zelfsturende functie veelal verkrijgbaar en aanwezig in het verkeer, met als groot voorbeeld Tesla. De verkoopcijfers van Tesla laten zien dat Nederland sinds 2017 een steeds grotere afnemer is geworden van hun desbetreffende auto's [2]. De definitie van een zelfsturende auto luidt: "Een zelfsturende auto, autonome auto of zelfrijdende auto is een voertuig dat wordt bestuurd door een automatische chauffeur die in staat is om de wagen met behulp van kunstmatige intelligentie van één punt naar een ander punt te brengen zonder menselijke tussenkomst" [3]. Om dit secuur uit te kunnen voeren is een zelfsturende auto uitgerust met sensoren zoals radar, gps en computervisie. Deze geavanceerde controlesystemen kunnen met de hulp van elektronische sensorinformatie een geschikt navigatiepad identificeren. Daarnaast worden hiermee obstakels en relevante zaken als een stopbord geïdentificeerd [4]. De mate van kunstmatige intelligentie in zelfrijdende auto's wordt hiermee steeds groter. Een grote factor hiervan bestaat uit 'decision-making', een welbekend fenomeen binnen de kunstmatige intelligentie. Gebruikmakend van een 'agent-based' methode, worden keuzes gemaakt en gewisseld van feedback tussen verschillende controlemechanismen. Doordat de sensoren en de processen, bestaande uit een diep neurale netwerk, daarvan alles om hun heen filteren en gebruiken voor volgende acties, is het netwerk continu aan het leren. Op basis van gebruik wordt het netwerk dus getraind om op deze wijze steeds meer situaties te herkennen en minder fouten te maken [5]. Daarnaast geeft bijvoorbeeld Tesla zelf aan, die momenteel auto's met Autopilot-functies aanbiedt, software-updates uit te voeren, zodat het systeem steeds beter wordt naarmate gebruik [6].

Echter zijn er diverse gradaties binnen de zelfsturende auto. Op dit moment zitten we in de fase waar veelal de zogenoemde Autopilot auto's op de weg te vinden te zijn en bevinden ons in de transitie naar volledig zelfrijdend. Auto's met een Autopilot-functie zijn auto's met een werking die gebruikmakend van kunstmatige intelligentie de bestuurder extra functionaliteiten aanbieden. Hierbij moet de bestuurder wel blijven oplettend, maar de auto kan in principe zonder aanraking sturen, gas geven en remmen [7]. Op dit moment zijn deze deels zelfrijdende voertuigen in heel Europa toegestaan [8]. Voor de rest van dit onderzoek wordt verwezen naar de zelfsturende auto met bovengenoemde functionaliteiten. Hiervoor is gekozen om het onderzoek actueel te houden en juist mogelijke adviezen te bieden die gericht op de toekomst zijn, waarin voorspeld wordt dat er in volledig zelfsturende auto's gereden zal worden op de openbare weg. Daarnaast zullen alle contexten zoveel mogelijk op de Nederlandse samenleving toegespitst zijn.

Naast de functionaliteiten van zelfbesturende voertuigen, mag het maatschappelijke dilemma dat eraan verbonden is niet onderbelicht worden. Zo ontstaan er diverse vraagstukken omtrent verantwoordelijkheid bij bijvoorbeeld ongevallen waar een zelfsturende auto betrokken bij is. Vaak is niet alleen de bestuurder betrokken, maar ook de fabrikant en programmeur bijvoorbeeld. De vraag die hierbij wordt opgeroepen, is dus hoe wij als mensen tegen morele verantwoordelijkheid aankijken en wie wij als schuldig zien. Een grote literaire basis is hiervoor al gelegd door een groot scala aan onderzoekers. Een overzichtelijke review hiervan is bijvoorbeeld geschreven door S. Nyholm in 2018 [9]. In veel van de onderzoeken werd het zogeheten trolleyprobleem voorgelegd waarbij een ethisch dilemma met een zelfrijdende auto verschillende uitkomsten zou kunnen hebben. Een voorbeeld dilemma is dat de auto een splitsing tegemoet gaat en moet kiezen tussen twee kwaden; linksaf ligt een baby en rechtsaf staan drie 75-plussers. Het gaat in dit geval om een kwestie tussen leven en dood. Het bleek dat diverse ethische stromingen oplossingen bieden voor mogelijke ongevallen en dergelijke hypothetische situaties. Een van de wat traditionelere stromingen wordt door veel mensen als juist beschouwd; het utilitarisme, een stroming die streeft naar een zo groot mogelijk behaal van geluk, ofwel waarbij de meeste mensen gered kunnen worden [10]. Toch bleek ook dat deze methode een keerzijde heeft en dat het niet in alle situaties even goed te rechtvaardigen is voor alle betrokken partijen.

Echter is er nog weinig onderzoek gedaan naar de directe link tussen het verantwoordelijkheidsgevoel van de bestuurder zelf en wat de invloeden van kunstmatige intelligentie daarop zijn. Een vernieuwend vervolg is dus om te kijken of er een manier zou zijn waarop je het morele verantwoordelijkheidsgevoel kan beïnvloeden, zodanig dat bestuurders exact weten en ervaren in hoeverre de verantwoordelijkheid bij hen zelf ligt.

Opbouw

De opzet van deze bijdrage is als volgt. Voor een verdere introductie over morele verantwoordelijkheid wordt in Hoofdstuk 2 het ethische aspect uitgewerkt van de zelfrijdende auto. Alle betrokken partijen komen hier aan bod en hoe zij op dit moment in de maatschappij gerechtvaardigd worden. In Hoofdstuk 3 worden een aantal van de determinanten van het verantwoordelijkheidsgevoel, met de toenemende kunstmatige intelligentie in acht nemend, uitvoerig besproken. In Hoofdstuk 4 wordt vervolgens naar manieren gekeken om dit gevoel te beïnvloeden, uitgaande van twee hoofdcategoryën; één zijnde een sociaal maatschappelijke aanpak en de ander betreft een aanpak die de technische kant van het systeem verandert. Het concluderende hoofdstuk met daarin ook een discussie wordt beschreven in Hoofdstuk 5.

Hoofdstuk 2 Morele verantwoordelijkheid

Dit hoofdstuk bevat de nodige achtergrondinformatie over de onderzoeken die tot dusver zijn gedaan over de morele verantwoordelijkheid van zelfrijdende auto's.

2.1 Betrokken partijen

2.1.1 Verantwoordelijkheidsgroepen

In een verkeerssituatie kunnen zich talloze deelnemers bevinden. Zo begint de eerste verantwoordelijkheidsgroep bij de gebruikers van het verkeer. In dit geval zijn dat de bezitters van de zelfrijdende auto. Zoals de context al aanduidt, zit hier gelijk ook de tweede verantwoordelijkheidsgroep in; de auto zelf. De zelfrijdende auto suggereert zichzelf van A naar B te kunnen bewegen zonder menselijke interactie, dus heeft hiermee een bijkomend verantwoord. Dit type verantwoordelijkheid gaat dus eigenlijk uit naar de 'agency'. Daarnaast is een auto, zelfrijdend of niet, deel van een groter geheel. Dit is de verantwoordelijkheid van de zogenoemde opdrachtgevers, ofwel de overheid of de fabrikant zelf. Als laatste is er nog de mogelijkheid om geen verantwoordelijkheid toe te schrijven aan wie dan ook [10].

2.1.2 Schuldig in ongevallen

In situaties waar ongevallen met zelfrijdende auto's betrokken zijn, wordt aan de zelfrijdende auto minder schuld toegeschreven dan aan een auto door mensen bereden. Daarnaast blijkt dat mensen een zelfrijdende auto niet als een onafhankelijk moreel verantwoord beschouwen, maar bestempelen de fabrikant en overheid als eindverantwoordelijk voor de technologie [10].

2.2 Juridische achtergrond [11]

Om het maatschappelijk belang te behartigen en te spreken van ethisch verantwoordelijkheden, kan daarbij het juridische aspect niet ontbreken. Het is een vraagstuk dat steeds uitdagender wordt om te beantwoorden in bijvoorbeeld het aansprakelijkheidsrecht. De grootste adders onder het gras zullen de situaties zijn die niet voorgeprogrammeerd zijn; uit onwetendheid en niet uit onkunde.

2.2.1 Wetgeving aansprakelijkheid

Er zijn momenteel veel tegenstrijdige wetten die diverse uitkomsten voor verantwoordelijkheid aantonen. Volgens de huidige Nederlandse wet, voortkomend uit *artikel 6:185 e.v. BW - Fabrikant, leveranciers en importeurs* is de fabrikant van een zelfrijdende auto, ook wel productaansprakelijkheid genoemd, eindverantwoordelijk voor de eventuele gebreken aan de zelfrijdende auto. Binnen deze regelgeving, kan het voorkomen dat de eindfabrikant de schuld vervolgens bij de leverancier van de losse onderdelen van de auto legt. Echter zal in dit geval het Hof

van Justitie uiteindelijk moeten beslissen welke partij aansprakelijk wordt gesteld.

Daarnaast is de definitie van een 'gebrekkig' voertuig ook nodig onder een vergrootglas te plaatsen. Een gebrekkig voertuig houdt volgens de wet in dat het voertuig niet de veiligheid biedt die men er, gelet op alle omstandigheden, van mag verwachten. Dit houdt wel rekening met moment van marktplaatsing.

De tweede wet die van belang is om te benoemen, is *artikel 6:162 BW - Gebruiker, bezitter van een gebrekkige zaak (art. 6:173 BW) en houder/eigenaar van een motorrijtuig (art. 185 WvW)*. Op basis van bovenstaande wet, kan de gebruiker van een zelfrijdende auto zelf aansprakelijk gesteld worden. Dit is mogelijk indien hij verweten wordt op een verkeerde manier gebruik te maken van de functionaliteiten. Ook kan het zijn dat de auto slecht onderhouden wordt en hierdoor gebrekkig gaat handelen; een ontstaan risico na marktplaatsing. Een bestuurder zal pas echt kans maken op productaansprakelijkheid als overmacht aangetoond kan worden. Echter ligt de lat om daar in aanmerking voor te komen hoog. In context van de zelfrijdende auto komt dit erop neer dat de gebruiker 'rechtens geen enkel verwijt' moet treffen en dat er een oorzaak van buiten het voertuig moet zijn geweest. Daarbij komt ook nog eens kijken dat desbetreffende oorzaak geheel buiten diens zicht en bewustzijn volstaat. Gebreken in de zelfrijdende auto zelf, inclusief programmeerfouten in de software of hardware, worden vooralsnog niet als grond van overmacht getekend.

De volgende wet genaamd *Andere weggebruikers (art. 6:162 BW) en de wegbeheerder (art. 6:174 en 6:162 BW)* beslaat de verschillen in beoordelingen tussen aansprakelijkheid van traditionele motorrijtuigen en aansprakelijkheid van zelfrijdende voertuigen. Hier wordt bijvoorbeeld meegenomen dat de gebruiker van de zelfrijdende auto zich geheel volgens de regels van de auto gedraagt en daarmee net zoveel deel uitmaakt van het verkeer als elk ander voertuig. Ook neemt deze wet de wegbeheerder in acht. Dit hangt samen met *artikel 6:174 BW risicoaansprakelijkheid* voor gebreken aan de weg. Doordat de overheid zelf toestaat de zelfrijdende auto's op de markt te brengen, brengt zij haar eigen wegbeheerders in een lastige positie.

De laatste wet die bijdraagt aan dit vraagstuk is *Contentverstrekkers/analisten; hackers (art. 6:162 BW)*. Aangezien een zelfrijdende auto functioneert op informatie van buitenaf, bijvoorbeeld weer en locatie informatie, schemert hier dus ook de mogelijkheid dat leveranciers van deze informatie beperkt verantwoordelijk kunnen zijn voor schade die veroorzaakt is door of aan een zelfrijdende auto. Echter zal onderaan de streep de eindfabrikant hier toch vaak aansprakelijk gesteld voor worden.

2.2.2 Wetgeving functionaliteiten

Zoals de introductie van dit artikel aangaf, is de Autopilot-functie in Nederland toegestaan op dit moment. Echter staat de huidige wetgeving nog niet toe om gebruik te maken van de volledige zelfrijdende functies. Doordat Tesla zelf bepaalt welke software er wanneer gedraaid wordt, is het plausibel dat deze functionaliteiten momenteel staan uitgeschakeld [8].

2.3 De kant van de fabrikant

Terugkijkend naar de Autopilot-functie in Tesla's Model S, die momenteel veelal op de weg gebruikt worden, en de risicoaansprakelijkheid die erbij komt kijken, is het vanzelfsprekend dat Tesla met uitermate doordachte bijschriften levert. Tesla zelf zegt dan ook de bestuurder te willen helpen met de vermoeiende aspecten van het autorijden en kan met Autopilot de auto automatisch laten sturen, accelereren en remmen binnen de rijstrook. Echter geven ze hierbij wel aan dat er actief toezicht van de bestuurder vereist is en zegt ook hiermee niet volledig zelfrijdend te zijn [6].

Naast de Autopilot-functionaliteiten, zegt Tesla ook dat elke nieuwe Tesla wel geleverd is met 'Full Self-Driving Capability' en de benodigde hardware hiervoor. Dit maakt het dus mogelijk, zodra de wetgeving dit ook toelaat, om de software gelijk uit te voeren en over te gaan op volledige zelfrijdend. Echter komen hier weer nieuwe gebruiksaanwijzingen bij kijken, bijvoorbeeld hoe er gehandeld moet worden als de bestuurder alcohol in zijn bloed heeft.

2.4 Conclusie

Zoals de huidige juridische achtergrond demonstreert, zijn er momenteel geen concrete wetten die technologieën als autonome auto's betrekken in aansprakelijkheidsrisico bijvoorbeeld. Daarnaast blijkt dat het vaak onduidelijk is of een ongeval of inschatting binnen of buiten de macht ligt van de bestuurder in combinatie met functionaliteiten die gebruikt zijn van een (deels) zelfrijdende auto. Hierdoor blijken zaken uiteindelijk vaak onbewijsbaar en blijft schuld tussen veel partijen hangen. Tot op de dag van vandaag is er dus geen juridisch kader waarin een bestuurders verantwoordelijkheid begint en eindigt, noch voor de fabrikant.

Hoofdstuk 3 Determinanten van verantwoordelijkheidsgevoel

Nu duidelijk is wat feitelijk en juridisch aspecten van aansprakelijkheid zijn, is de volgende stap het onderzoeken van de subjectieve aspecten. Dit hoofdstuk bevat een aantal van de determinanten die bijdragen aan het verantwoordelijkheidsgevoel van een bestuurder in een zelfrijdende auto.

3.1 Human-on-the-loop

Eerder in deze paper is gesproken over de vermindering van uitvoerende operaties door mensen bij het gebruik van Autopilot-functies of volledige zelfbesturing van een auto. Omdat de auto dus eigenlijk gaat meedenken en bepalen met de mens, wordt geïmpliceerd dat de mens minder directe controle over de auto heeft. Omdat de controle van de mens echter enkel verminderd wordt en niet totaal buitengesloten, behoort de zogenoemde Autopilot van Tesla tot de 'human-on-the-loop' (HOTL) systemen [12]. Human-on-the-loop systemen zijn, zoals de naam ook al deels onthult, modellen of systemen die menselijke interactie vereisen [13]. Doordat de mens op deze wijze eigenlijk het neurale netwerk van de agent of het algoritme aan het trainen is, wordt het hiermee continu verbeterd. Doordat de bestuurder tijdens het gebruiken van Autopilot bijvoorbeeld nog wel zijn handen aan het stuur moet houden en moet ingrijpen waar nodig, gebeurt deze 'machine learning' dus tijdens gebruik van het systeem. Er ontstaat op deze manier een samenwerking tussen mens en machine. Doordat de auto dusdanig met de mens mee gaat denken, is de controle uitoefening van de mensen minder noodzakelijk.

De toekomst zal hierdoor echter wel inhouden dat het algoritme geoptimaliseerd zal zijn doordat het nu continu getraind wordt, ook onbewust, door de huidige gebruikers ervan. Dit betekent ook dat het systeem uiteindelijk kan functioneren waarbij menselijke bestuurders 'out-of-the-loop' geplaatst worden [14]. Een risico dat dat mee zich brengt is mogelijke kans van verlies van bepaalde kwaliteiten en handmatige vaardigheden die men normaliter gebruikt om auto te rijden. Een bestuurder gebruikt bij een normale auto actieve hersenprocessen om te rijden en op te letten, maar bij gebruik van automatisering blijkt het brein over te gaan op passievere informatieverwerking. Onderzoek van Endsley et al. (1995) toonde aan dat, ondanks in de eerste instantie handmatige training, proefpersonen die opereerden als toezichhoudende controllers van een automatisering in een simulatie langzamer en inefficiënter waren in het systeem onder controle houden dan de proefpersonen die enkel in een handmatige modus hadden gewerkt. Ook bleek dat handmatige vaardigheden niet alleen dienen voor de uitvoering van een taak, maar ook voor de noodzakelijke detectie ervan wanneer uitvoering vereist is [15]. In het geval dat de zelfrijdende auto bijvoorbeeld een plas niet op zijn sensor detecteert en daar met volle vaart op afgaat, is het van belang dat de bestuurder weet hoe je verstandig remt om niet te gaan slippen.

De bijdrage van dit model aan het verantwoordelijkheidsgevoel zit hem dus met name in de manier waarop de mens interacteert met het systeem en in hoeverre de 'human in the loop' blijft en wat de effecten daarvan zijn. Een factor met een grote bijdrage aan hoe deze interactie tussen mens en machine goed verloopt, is vertrouwen.

3.2 Vertrouwen

3.2.1 Vertrouwen in technologieën

De subjectieve ervaring van handelingen overdragen aan een systeem gaat hand in hand met een dusdanige hoeveelheid vertrouwen in een systeem hebben. Vertrouwen bestaat uit een set overtuigingen bestaande uit onder andere integriteit en voorspelbaarheid, de bereidheid van een partij afhankelijk van een andere partij in een situatie die risico meebrengt of juist de combinatie hiervan. Daarnaast bestaat er, wat ook erg van belang is bij de promotie van nieuwe technologieën, een nuance tussen twee soorten vertrouwen. Het eerste wordt 'initial trust' genoemd, beter bekend als de eerste indruk. Bij kunstmatige intelligentie draagt representatie, perceptie en meningen van anderen veel bij aan de eerste indruk. Hoe de eerste indruk wordt waargenomen ligt dus deels bij de eindfabrikant en verkoper, maar kan anderzijds dus ook via andere gebruikers verspreid worden. De inhoud van een positieve eerste indruk gaat met name om de begrijpbaarheid van de technologie. Mensen vinden het over het algemeen prettig om goed te snappen hoe de technologie is opgebouwd en welke functionaliteiten voor welke gevolgen zorgen.

Vervolgens ontstaat een continu vertrouwen wat verder ontwikkeld en gevoed kan worden. Dit bestaat uit onder andere de betrouwbaarheid van een product en de mate van omgangbaarheid. Het is van belang dat automatisering juist gemakkelijk is om te gebruiken en niet een voor de mens simpele taak vermoelijk. Daarnaast is er bij bijvoorbeeld Autopilot-functionaliteiten sprake van een samenwerking tussen technologie en de mens. Om vertrouwen te voeden, is het van belang om de communicatie en samenwerking soepel uit te dragen. Een wellicht logische bijdrage aan de continuïteit van vertrouwen, is de veiligheid en privacybescherming van de technologie. Als veiligheid beschadigd raakt, verliest men vertrouwen in de kwaliteiten ervan. Als laatste is transparantie en de uitlegbaarheid van de systemen belangrijk. Zodra een product of concept te onbegrijpelijk wordt, daalt het vertrouwen daarmee. Transparantie is daarmee dus een essentiële factor voor zowel de eerste indruk als de continuïteit van vertrouwen. Daarnaast is het doel van de kunstmatige intelligentie belangrijk om ten alle tijden in acht te nemen om het vertrouwen gestimuleerd te houden; vervanging van de menselijke taken. Als er namelijk te veel zorg is door mensen dat kunstmatige intelligentie bijvoorbeeld hun baan gaat vervangen, belemmert dit het vertrouwen [16].

3.2.2 Balans

Als de mensheid echter te veel vertrouwen in systemen heeft, kan dit ook weer voor een negatief effect zorgen binnen het morele verantwoordelijkheidsgevoel. Door te veel vertrouwen in een systeem te hebben, wordt het passieve gedrag vergroot. De balans hierin is dus een essentiële determinant. Binnen de wetenschap is deze balans definieerbaar als 'optimal trust', het gezonde evenwicht tussen vertrouwen en wantrouwen in de passende context [17]. De 'optimal trust' bepalen is alleen zo makkelijk nog niet. In veel vakgebieden waar kunstmatige intelligentie gebruikt wordt, worden hiervoor diverse modellen en algoritmes gebruikt. Aan de Macquarie University in Australië namen vier onderzoekers het aanbevelingssysteem voor de selectie van betrouwbare serviceproviders in complexe sociale netwerken als voorbeeld [18]. Zij onderzochten bijvoorbeeld het heuristische algoritme H_MCOP waarbij waarden met meerdere beperkingen en waarden van QoS-attributen werden geaggregeerd op basis van een logische vergelijking. Op deze manier werden optimale waarden berekend die de decision-making vanuit logica gebruikten. Echter heeft het tot nu toe nog niet geresulteerd in een foutloos algoritme en zal een grotere database nodig zijn in de toekomst om meer gewenste resultaten te behalen.

Daarop inhakend zijn er ook manieren die meer op de daadwerkelijke wijze van menselijke decision-making meer of minder vertrouwen in een systeem kunnen genereren. Door vertrouwen niet correct te kalibreren met systeemcapaciteiten, kan dat ertoe leiden dat automatisering niet meer gebruikt of misbruikt wordt [19]. Niet-gebruik ontstaat wanneer de operator minder vertrouwen in het systeem heeft dan nodig. Zodoende gebruikt de operator het systeem niet ten volle capaciteiten. Anderzijds treedt misbruik op wanneer de operator de automatisering overtreft en kan weer leiden tot fouten doordat de operator acties van de automatisering verwacht die te hoog gegrepen zijn. Menselijk bewustzijn veranderen of aanbrengen, ook wel 'sensemaking' genoemd, wordt gedefinieerd als een opzettelijk proces waar besluitvormers op vertrouwen, om hun bewustzijn te verbeteren van onzekere en dubbelzinnige situaties. In hoofdstuk 4 komt aan bod hoe hier op een oplossende wijze gebruik gemaakt van kan worden.

3.3 Autonomie

Naast de juiste balans van vertrouwen, heeft autonomie ook een dubbelzijdig aspect. Aan de ene kant wordt het voertuig steeds autonomer, maar aan de andere kant verliest de mens hiermee autonomie. Onderzoek naar het gevoel van autonomie in het tijdperk van kunstmatige intelligentie betreft diverse wetenschappen: marketing, economie, filosofie, neurowetenschappen en psychologie. Al deze wetenschappen samen bieden een perspectief op hoe wij als mensen omgaan met kunstmatige intelligentie en in welke maten wij nog handelen vanuit vrije wil, eigen beslissingen en voorwaarden; ook wel autonomie genoemd.

Als eerste is het belangrijk om te benoemen dat mensen plezier uit autonomie halen. Het gevoel dat hun acties impact hebben op de wereld, wordt als positief ervaren. Als vrije wil als perceptie wordt ervaren, ontwikkelen mensen ook een bepaalde mate van moreel verantwoordelijkheidsgevoel [20]. In een normale auto hebben alle handmatige handelingen van de mens direct invloed op hoe het voertuig zich manifesteert en daarmee de wereld beïnvloedt. Dit gevoel van keuzevrijheid werkt positief op het gevoel van responsabiliteit. Door restricties op keuzevrijheid te leggen, ervaren mensen minder motivatie en wekt dit psychologische reactantie op, een term uit de psychologie die gebruikt wordt wanneer mensen sterk emotioneel reageren als hun bewegingsvrijheid wordt bedreigd [21]. In het geval van de (deels) zelfrijdende auto gebeurt dat door bepaalde handelingen aan de automatisering over te laten en weg te nemen van de menselijke operator. Daarnaast is het wel zo dat juist het bewust zijn van het hebben van vrije keuze een grote bijdrager is aan het positieve gevoel van autonomie. Het is dus niet zozeer dat de belemmering op keuzevrijheid op zichzelf een probleem hoeft te zijn, maar de bewustwording van de restricties maakt dat de verandering van subjectieve ervaring van autonomie kenbaar is. De determinant bestaat dus uit zowel vermindering van autonomie als het bewustzijn waarmee het ervaren wordt.

Gevoel van controle en vrijheidskeuze leidt dus tot positieve resultaten op subjectieve ervaringen van autonomie. Zodra hier restricties bij komen kijken en bewustwordingen daarvan aangetoond zijn, dreigt de motivatie en het verantwoordelijkheidsgevoel te dalen.

3.4 Meaningful Human Control

Alle bovengenoemde determinanten beïnvloeden het verantwoordelijkheidsgevoel van mensen. De zorg hierom wordt steeds groter door de toename van kunstmatige intelligentie in alles om de mens heen. Deze zorg kan geuit worden in het zogenoemde 'meaningful human control' (MHC) principe. Dit principe is afkomstig van de discussie omtrent autonome wapensystemen. Autonome wapensystemen beschikken over de capaciteiten om zonder menselijk contact tussen leven en dood te kiezen, bijvoorbeeld door een terrorist uit te schakelen [22]. Doordat dit soort wapens een grote ethische zorg teweeg hebben gebracht, werd er gesproken van een 'verantwoordelijkheidskloof' veroorzaakt door deze systemen. Het principe 'meaningful human control' werd in het juridisch-politieke debat geïntroduceerd om zorgen weg te nemen. Volgens het MHC-principe zouden mensen, niet de computers en de algoritmes die daarbij horen, de uiteindelijke controle moeten behouden en dus verantwoordelijk blijven voor relevante beslissingen over militaire operaties, ofwel dodelijk of niet. Over het algemeen zijn er twee voorwaarden waar een autonoom systeem aan moet voldoen om onder het MHC-principe te blijven: ten eerste, een 'tracking'-voorwaarde en ten tweede een 'tracing'-voorwaarde. De 'tracking'-voorwaarde houdt in dat het systeem moet kunnen reageren

op zowel de morele redenen van de mensen die het ontwerpen en gebruiken evenals de relevante feiten die zich in de omgeving bevinden waar het systeem in handelt. Deze omvat dus zowel mentale toestanden van menselijke gebruikers als kenmerken van de buitenwereld. Een voorbeeld in een zelfrijdende auto zou als volgt zijn: bestuurder A rijdt in een zelfrijdende auto B, die zich dus tussen A en de wereld bevindt. Dit systeem B moet overtuigingen verkrijgen of er iets het geval is of niet, bijvoorbeeld of er een persoon staat waar de auto wil parkeren (waarbij C symboliseert of dat zo is). De 'tracking'-voorwaarde impliceert in dit geval:

1. Als C niet het geval is en A systeem B gebruikt om tot een overtuiging te komen of C wel of niet aanwezig is, dan zou B niet geloven dat er een persoon staat
2. Als C het geval is en A systeem B gebruikt om tot een overtuiging te komen of C wel of niet aanwezig is, dan zou B geloven dat er een persoon staat

Concreet houdt het in dat het systeem, ongeacht zijn werking, in staat moet zijn om op een succesvolle manier te reageren op de kenmerken van de wereld. Wanneer het systeem gebruikt wordt door mensen als besluitvormingsmethode, is het noodzaak dat het systeem de (relevante) morele redenen volgt van de mens. Als de mens bijvoorbeeld ziet dat er een hond op de parkeerplaats staat en de machine dit aanziet als een persoon, zou een wijziging in gedrag van het systeem gecorrigeerd moeten worden. Een mens zou volgens eigen morele redenen de hond namelijk niet willen aanrijden. Hetzelfde geldt voor bijvoorbeeld een wezenlijk moreel verschil op een andere plek op de wereld; in Nederland rijdt men rechts en in Engeland zou de auto aan de linkerkant van de weg moeten rijden. Concluderend moet een systeem, om onder zinvolle menselijke controle te staan, aantoonbaar reageren op de menselijke morele redenen die relevant zijn aan de omstandigheden en omgeving.

De 'tracing'-voorwaarde houdt in dat een systeem zodanig ontworpen moet worden dat het de mogelijkheid biedt om de resultaten van zijn handelingen altijd te kunnen herleiden tot ten minste één persoon in de opvolging van het ontwerp en de exploitatie. Een voorbeeld is dat iemand achter het stuur gaat zitten en ziet dat de auto niet goed is geüpdatet. Toch kiest de bestuurder ervoor om de volledige controle aan de zelfrijdende auto te geven en veroorzaakt een ongeluk. Deze situatie is herleidbaar naar het moment dat deze persoon ervoor koos om te gaan rijden zonder zelf het stuur over te nemen, wetend dat de software niet goed was geüpdatet. Echter geeft de fabrikant altijd aan om nooit met een halve update van de software te gaan rijden en de controle volledig uit handen te geven. Het ongeval is dus terug te herleiden naar een specifieke keuze van één persoon [23].

Deze voorwaarden samen maken een geheel bruikbare vuistregel om verantwoordelijkheidsgevoel op peil te houden. Samenvattend houdt de

MHC-theorie in, zowel in de autonome wapensystemen als de zelfrijdende auto's: mensen maken bewuste beslissingen, zijn voldoende geïnformeerd over rechtmatigheid en zijn als laatste goed opgeleid om controle over hun autonome voertuig te hebben [24].

3.5 Conclusie

Er is een grote samenhang tussen de determinanten van het verantwoordelijkheidsgevoel. Als eerste moet de samenwerking tussen een automatisering en de mens goed verlopen. Een grote factor tussen deze twee is vertrouwen, dat gebaseerd kan zijn op diverse percepties. Echter moet er wel een gezonde hoeveelheid wantrouwen blijven bestaan om mensen niet te passief te laten handelen. De balans kan worden behaald door invloed uit te oefenen op bewustwording, wat weer samenhangt met de mate waarin autonomie ervaren wordt. Als laatste is er dus een mogelijkheid om het MHC-concept in acht te nemen, wat net als 'optimal-trust' algoritmes door middel van logica de optimale besluitvormingen van automatiseringen vergelijkt.

Hoofdstuk 4 Verantwoordelijkheidsgevoel optimaliseren

Uit hoofdstuk 3 blijkt dat verschillende determinanten invloed uitoefenen op het morele verantwoordelijkheidsgevoel van mensen. Echter is het zaak om ondanks controlevermindering het gevoel van morele verantwoordelijkheid in balans te houden. Hiervoor wordt in dit hoofdstuk ingezoomd op de determinanten en met welke toepassing hier iets aan gedaan kan worden.

4.1 Vaardigheden

De determinant 'human-on-the-loop' benoemde de keerzijde van mensen uit de loop plaatsen bij een automatisering; de vermindering van vaardigheden die toch wel nodig blijken te zijn, ondanks dat ze niet de gehele tijd uitvoerend zijn. Het is dus van groot belang dat men wel getraind blijft worden in vaardigheden die normaliter nodig zijn om een auto te rijden, bijvoorbeeld fileparkeren. Hiervoor is de invloed van de overheid het grootst, ofwel de sociaal maatschappelijke aanpak. Deze pedagogische methode slaat namelijk op de manier hoe Nederland haar bestuurders opleidt tot een geschikte autorijder en weggebruiker. Op dit moment kan een rijbewijs behaald worden in een normale auto en wordt er geen rekening gehouden met eventuele overnames van vaardigheden van de autonome auto. Met de kennis van mogelijk verlies van menselijke handmatige vaardigheden zou de focus dus nog steeds op het aanleren ervan moeten liggen. Echter bleek ook uit het theoretisch kader van de HOTL-systemen dat een juiste samenwerking tussen mens en machine ook essentieel is voor een soepele implementatie. Wellicht moet er dus juist extra tijd geïnvesteerd worden in rijopleidingen om naast handmatige vaardigheden, samenwerkingsverbanden te ontwikkelen tussen automatiseringen en mensen. Een mogelijkheid om dit beter aan te leren, zou kunnen betekenen dat mensen naast een regulier rijexamen ook een test in een simulator moeten afnemen om de effecten van een zelfrijdende auto te ervaren en te kunnen controleren.

Daarnaast biedt de simulator een eventuele oplossing voor het verantwoordelijkheidsgevoel dat minder wordt ervaren bij het dalen van controle. Als op het punt van leren autorijden al gekeken wordt naar dergelijke zaken, kan op dit punt ook de verantwoordelijkheid worden 'aangeleerd'. Ook hierin zou de overheid de overhand moeten nemen.

4.2 Vertrouwen

HOTL-systemen vereisen voor een optimale samenwerking niet alleen onderhouden vaardigheden, maar ook vertrouwen. In de paragraaf over vertrouwen, werd duidelijk dat menselijk vertrouwen in automatisering geoptimaliseerd kan worden door heldere communicatie en transparantie. Zolang de communicatie van het systeem naar de mens goed blijft, zou

hiermee het vertrouwen in automatisering kunnen blijven groeien. Communicatie bestaat uit de uitleg over het systeem, maar ook over het gebruik ervan.

4.2.1 Vertrouwen in technologieën

Een invloed uitoefenen op de eerste indruk is een lastige determinant om op te sturen. Het kan nou eenmaal net gebeuren dat iemand toevallig een artikel op een nieuwswebsite tegenkomt waar een zelfrijdende auto in slecht daglicht wordt gezet. De voeding van vertrouwen in technologieën is daarom interessanter om te beschouwen. Er zijn twee kernwaarden die dusdanig controleerbaar zijn om te aan te halen.

De eerste factor gaat over de communicatie die tussen mens en machine speelt. De communicatie beslaat die van het gebruik, wat inhoudt dat de fabrikant naar de gebruiker helder moet zijn over hoe op de juiste manier met het product omgegaan moet worden. Op deze manier voorkom je ook misbruik of niet-gebruik wat inhaakt op het 'sensemaking' uit Hoofdstuk 3. Als eerste is het van belang om in de 'human-machine communication' (HMC) technologie als een middel te gaan zien; het is een onderdeel geworden van alle communicatie [25]. De daadwerkelijke discussie over HMC gaat dus op in een veel groter geheel; de hele communicatiewetenschap. Als duidelijk is hoe mensen onderling interacteren, kan pas in kaart worden gebracht hoe invloeden als technologie dat beïnvloeden. Onderzoek naar HMC gaat dus niet puur over de interactie die mensen met een machine hebben, maar benadrukt meer de rol van technologie binnen de communicatie die nu bestaat.

Die informatie in acht nemend, zijn er veel mogelijke communicatiemogelijkheden tussen mens en machine. Omdat een bestuurder ondertussen ook de auto moet besturen, actief ofwel passief, biedt stemondersteuning in dit geval de meeste mogelijkheden en zal in dit onderzoek ook enkel belicht worden. Uit empirisch onderzoek blijkt dat stem-gebaseerde interactie met de focus op gesproken taal door de machine en de mogelijkheid vragen te stellen aan de machine geprefereerd wordt [26]. Voor een zelfrijdende auto die samenwerkt met mensen biedt vocale interactie dus een optie voor bijvoorbeeld uitleg over een nieuwe functionaliteit na een software-update.

Zoals kort benoemd werd, werkt het ook andersom; vanuit de mens naar de machine. De andere kernwaarde van continuïteit van vertrouwen was transparantie. Mensen neigen een technologie sneller te vertrouwen als ze weten hoe het werkt. Een mogelijk hulpstuk bij de uitleg van de auto en het mechanisme erachter biedt de stemondersteuning. Uitleg zou uiteraard ook kunnen bij aankoop, maar dit zal snel vergeten worden en tevens niet garanderen dat degene die de uitleg krijgt ook altijd de bestuurder zal zijn. Een potentieel alternatief is dus om tijdens gebruik of na een software-

update stemondersteuning in te schakelen voor transparantie van het systeem.

Zoals bovengenoemde aangaf, ervaren mensen een gesproken taal interactie als prettig en voor de transparantie zou de mens bijvoorbeeld vragen kunnen stellen aan het systeem. Het systeem zal van een vragendatabase beschikken om zodoende mensen te ondersteunen en transparantie te behalen.

4.2.2 Kalibratie

Er zijn daarnaast ook oplossingen die de 'sensemaking', een determinant uit het vorige hoofdstuk, van mensen in extreme vorm kunnen influenceren.

Enerzijds valt er te beargumenteren dat de verkoper gebruikers van een zelfrijdende auto verzekert van de controle die ze juist nog wel uitoefenen. Dit kunnen ze doen door sommige functionaliteiten bijvoorbeeld uit te schakelen of door een aangepast algoritme te leveren wat persoonlijk is aangepast per bestuurder. Een aangepast algoritme is echter meer een technische verandering dan een maatschappelijke. Het is dan ook een oplossing die uiteindelijk bij de fabrikant te zoeken is, maar toch heeft dit op meerdere determinanten een invloed die van grote kracht kan zijn. Door persoonlijke voorkeuren door te voeren, zal de bestuurder nog steeds wel handmatige acties overdragen aan het systeem, maar is er toename in keuzevrijheid, wat positief kan werken op het gevoel van responsabiliteit. De fabrikant geeft hiermee de restrictie op keuzevrijheid een draai. Daarnaast krijgen bestuurders het gevoel dat er een sterker samenwerkingsverband tussen mens en machine doordat ze een eigen zeggen hebben in het functioneren van het algoritme.

Anderzijds, als het knelpunt het gevoel van keuzevrijheid is, zou de eindfabrikant ook het tegenstrijdige kunnen opwekken. Mocht de wet het ooit toelaten, zou de fabrikant juist de gebruiker actief kunnen aansporen het voertuig geheel zelf te laten rijden. Het doel hiervan zou dan zijn om de gedachte op te wekken dat de gebruiker hiermee de wegen en het transport juist veiliger maakt. Hiermee wordt het verantwoordelijkheidsgevoel over de algehele veiligheid van het verkeer en de samenleving gedreven en niet louter over het voertuig zelf. Echter sluit je hiermee de mens volledig buiten de loop en kan dit gepaard gaan met het eerder besproken verlies van specifieke vaardigheden.

4.3 Meaningful Human Control

De drie principes van 'meaningful human control' blijven centraal staan in veel, mogelijk alle, oplossingen: mensen maken bewuste beslissingen, zijn voldoende geïnformeerd over rechtmatigheid en zijn goed opgeleid om controle over hun autonome voertuig te hebben. Een aspect dat wellicht al aangepakt kan worden tijdens rijopleiding betref de voldoende informatie voorzien aan bestuurders over rechtmatigheid en verantwoordelijkheid.

Ook bleek dat een gesproken taal interactie tussen mens en machine als prettig wordt ervaren. Daarnaast wordt de huidige zelfrijdende auto steeds slimmer doordat mensen in de huidige omgeving van de Autopilot wel moeten blijven opletten en het systeem corrigeren. In de huidige human-on-the-loop omgeving is het dus eigenlijk zo dat het systeem feedback krijgt van de mens. Maar omgekeerde feedback zou weer een ander perspectief kunnen bieden voor de samenwerking tussen mens en machine. Dit kan bijvoorbeeld gedaan worden door sensoren in de auto te plaatsen die bijhouden hoe vaak de handen het stuurwiel aanraken en zodoende gelijk feedback geven aan de bestuurder of dit voldoende is of niet. De feedback kan vervolgens ook weer op verschillende manieren aan de bestuurder zicht- of hoorbaar gemaakt kunnen worden. Met een sprekend systeem of alarmerend geluid zou dit bijvoorbeeld mogelijk gemaakt kunnen worden.

Het zou in de toekomst ook nog een stap verder kunnen om zo te voorkomen dat mensen geen misbruik van automatisering maken, een probleem dat kan ontstaan als de kalibratie van vertrouwen in een automatisering niet correct gebeurt. Het kan gebeuren dat een bestuurder denkt wel een alcoholische versnapering te kunnen nuttigen voordat hij een zelfrijdende auto instapt. Hier zou een controlemechanisme vergelijkbaar met een blaastest wederom een feedback systeem kunnen bieden om dit soort gedragingen en misbruiken te voorkomen. In het begin van de zelfrijdende auto tijdperk is het voorbarig om mensen dan alsnog te kunnen laten rijden, maar mocht de technologie voldoende betrouwbaar zijn, is dit een goede gedragscontrole. Het zou dan kunnen dat de ingebouwde blaastest de zelfrijdende functies van de auto juist aanzet als de bestuurder te veel alcohol heeft gedronken en het handmatige rijden activeert wanneer dit niet het geval is [27].

Hoofdstuk 5 Conclusie & Discussie

In dit onderzoek stond de vraag centraal of er een manier zou zijn waarop het morele verantwoordelijkheidsgevoel beïnvloed kan worden, zodanig dat bestuurders van (deels) zelfrijdende auto's exact weten en ervaren in hoeverre de verantwoordelijkheid bij hen zelf ligt.

Als eerste is besproken in welke mate de overheid op dit moment regels en wetten heeft opgesteld die vaststellen wie wanneer als schuldig verklaard kan worden bij ongevallen met een betrokken zelfrijdende auto. Er zijn veel wetten die elk een klein beetje bijdragen aan de uiteindelijke rechtvaardiging. Toch maakte de huidige literatuur [11] duidelijk dat er een concreet artikel ontbreekt dat over de functies van (een deels) zelfrijdende auto oordeelt en functionaliteiten ervan aanvecht. De huidige regelgeving biedt uitkomsten voor simpele ongevallen waarbij het slachtoffer gehalte minimaal is. Maar dit onderzoek toont aan dat, om genoeg rekening te houden met wat komen gaat betreft de snelle ontwikkelingen van de technologie, er strikte regels nodig zijn die te handhaven zijn voor zowel de overheid, de eindfabrikant als de bestuurder.

Hoofdstuk 4 liet een andere determinant zien die de overheid onder invloed heeft, namelijk de huidige opleidingsomgeving van bestuurders. Zoals het onderzochte risico naar de samenwerking tussen mens en machine duidelijk maakte en daarmee vaardigheid verliezen aantoonde [15], moeten handmatige vaardigheden nog wel aangeleerd blijven. Dit onderzoek bevestigt deze theorie, maar ziet wel gelegenheid voor uitbreiding op de huidige rijopleiding. Een optie is om additionele training aan te bieden waarbij nieuwe vaardigheden aangeleerd worden die bij een samenwerking met automatisering komen kijken. Bij de aanvullende training kan ook extra aandacht geschonken worden aan de subjectieve beleving van de bestuurder in combinatie met de automatisering. Echter zijn dit speculatieve potentiële oplossingen en zou het effect hiervan in de praktijk getoetst moeten worden. Dit zou gedaan kunnen worden door een aantal rijlessen in Nederland pilots uit te laten voeren waarbij naast de reguliere rijlessen ook simulatorritjes 'gereden' worden in autonome auto's. Een voorwaarde is dan wel dat de rijinstructeur hier eerst zelf voor bijgeschoold moet worden. Door er ook nog een pilot-rijexamen aan toe te voegen zou dat ons kunnen helpen om te begrijpen of extra scholing daadwerkelijk zin zou hebben, mocht de volledig zelfrijdende auto ooit volgens wetgeving in het verkeer gebruikt gaan worden.

De determinant vertrouwen heeft wederom nog vervolgonderzoek nodig wat praktijkgericht moet zijn om daadwerkelijke adviezen op te baseren. Uit huidig onderzoek [18] bleek dat de 'optimal trust' algoritmes tot dusver nog niet tot de ultieme resultaten leiden. Echter zullen we onszelf moeten afvragen of een complete database ooit zal leiden tot het perfecte algoritme. Aangezien er jaarlijks overal ter wereld nieuwe wegen worden

aangelegd en de mensheid altijd gebrekkig zal zijn, zullen er altijd nieuwe situaties bijkomen of ontstaan die nog niet gevangen zijn in een database.

Toch impliceert dit onderzoek dat het streven naar de juiste balans van vertrouwen niet onrealistisch is. Met stimulans op factoren als communicatie en transparantie van de overheid, fabrikant en gebruiker is er wel degelijk een goede kans op de bewustwording van morele verantwoordelijkheid. Of iedereen het prettig vindt om door zijn/haar auto aangesproken te worden over nieuwe updates, is een tweede vraag. In combinatie met een eventueel gepersonaliseerd algoritme biedt dit een uitmuntende toetsing of stemondersteuning daadwerkelijk het vaakst geprefereerd wordt als manier van communiceren met een systeem. Aan de kant van de fabrikant ligt de optie om een persoonlijk algoritme aan te bieden, mocht dat binnen hun capaciteiten liggen. De verstrekte informatie uit dit onderzoek kan worden gebruikt om door fabrikanten gerichte praktische testen te ontwikkelen om de haalbaarheid hiervan te onderzoeken.

Waar een wellicht struikelpunt doorschemert, is de oplossing die transparantie van de technologie - ten goede aan het vertrouwen - biedt. Ondanks dat de wetenschap er weinig invloed op heeft, rust het 'black-box effect' op besluitvorming van geautomatiseerde kunstmatige intelligentie systemen [28]. Dit houdt in dat het algoritme data gebruikt en daar besluitvorming op baseert, maar eigenlijk zonder redenen hoe of waarom te onthullen. Hierdoor hebben mensen niet alleen moeite om vooroordelen die het systeem overneemt van mensen eruit te filteren, maar is mensen van een uitleg voorzien ook een lastige opgave. Daarnaast is de gebruikte terminologie om het wel uit te leggen voor de gemiddelde Nederlander geen dagelijks jargon. De combinatie van veel terminologie uit de logica en het 'black-box effect' kan transparantie een lastig streven maken. Toch draagt de kennis over de kracht van transparantie bij aan het belang ervan en zou er onderzocht kunnen worden of er makkelijkere manieren zijn het mechanisme uit te leggen.

Al met al illustreert dit onderzoek dat een futuristische zelfrijdende auto in de praktijk nog niet zo snel geïmplementeerd kan worden, ondanks dat de technologie het zelf al wellicht zou toestaan. Dit onderzoek biedt implicaties voor zowel de sociaal maatschappelijke aanpak als de technische kant van automatisering van auto's. Er zitten veel adders onder het gras en er zijn veel rampscenario's die vermeden kunnen worden door van tevoren strikte, concrete, beschermde, in stand te houden en te handhaven regels en systemen zijn opgesteld. Maar als dat eenmaal steevast staat, zou het verkeer er over een aantal decennia wel eens heel anders uit kunnen zien.

Hoofdstuk 6 Bibliografie

- [1] CBS, van Beuningen, J., Molnar-in 't Veld, H., & Bouhuijs, I. (2012, april). *Personenautobezit van huishoudens en personen*. <https://www.cbs.nl/nl-nl/achtergrond/2012/14/personenautobezit-van-huishoudens-en-personen>
- [2] *Tesla verkoopcijfers Nederland*. (z.d.). AutoWeek. Geraadpleegd op 18 maart 2021, van <https://www.autoweek.nl/verkoopcijfers/tesla/>
- [3] Wikipedia-bijdragers. (2021, 27 februari). *Zelfsturende auto*. Wikipedia. https://nl.wikipedia.org/wiki/Zelfsturende_auto#:~:text=Een%20zelfsturende%20auto%2C%20autonome%20auto,te%20brengen%20zonder%20omenselijke%20tussenkomst
- [4] Howard, D. (2014). *Public Perceptions of Self-Driving Cars: The Case of Berkeley, California* | Semantic Scholar. <https://www.semanticscholar.org/>. <https://www.semanticscholar.org/paper/Public-Perceptions-of-Self-Driving-Cars%3A-The-Case-Howard-Dai/39a10ac3ad0ab01bce3aa5d323f9700e53f7f34e>
- [5] Veres, S. M., Molnar, L., Lincoln, N. K., & Morice, C. P. (2011). *Autonomous vehicle control systems — a review of decision making*. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 225(2), 155–195. <https://doi.org/10.1177/20413041110394727>
- [6] *Autopilot*. (z.d.). Tesla Nederland. Geraadpleegd op 21 maart 2021, van https://www.tesla.com/nl_NL/autopilot
- [7] Ingle, S., & Phute, M. (2016, september). *Tesla Autopilot : Semi Autonomous Driving, an Uptick for Future Autonomy*. <https://www.irjet.net/archives/V3/i9/IRJET-V3I969.pdf>
- [8] *Is de Zelfrijdende functie toegestaan in Nederland?* (z.d.). support.mistergreen.nl. Geraadpleegd op 22 maart 2021, van <https://support.mistergreen.nl/zelfrijdend>
- [9] Nyholm, S. (2018). The ethics of crashes with self-driving cars: A roadmap, I. *Philosophy Compass*, 13(7), e12507. <https://doi.org/10.1111/phc3.12507>
- [10] Li, J., Cho, M. J., Ju, W., Malle, B. F., & Zhao, X. (2016, april). *From Trolley to Autonomous Vehicle: Perceptions of Responsibility and Moral Norms in Traffic Accidents with Self-Driving Cars*.

<https://www.researchgate.net/publication/299640195> From Trolley to Autonomous Vehicle Perceptions of Responsibility and Moral Norms in Traffic Accidents with Self-Driving Cars

[11] Ars Aequi. (2019, 8 april). *Wetgever, pas op! De (vrijwel) autonome auto komt eraan*. [https://arsaequi.nl/product/wetgever-pas-op-de-vrijwel-autonome-auto-komt-eraan/#:%7E:text=De%20\(vrijwel\)%20autonome%20auto%20komt%20eraan,-In%20de%20technologiesector&text=Het%20zal%20niet%20lang%20meer,ook%20weer%20nieuwe%20schaderisico's%20introduceren](https://arsaequi.nl/product/wetgever-pas-op-de-vrijwel-autonome-auto-komt-eraan/#:%7E:text=De%20(vrijwel)%20autonome%20auto%20komt%20eraan,-In%20de%20technologiesector&text=Het%20zal%20niet%20lang%20meer,ook%20weer%20nieuwe%20schaderisico's%20introduceren).

[12] Nahavandi, S. (2017). *Trusted Autonomy Between Humans and Robots: Toward Human-on-the-Loop in Robotics and Autonomous Systems*. IEEE Systems, Man, and Cybernetics Magazine, 3(1), 10–17. <https://doi.org/10.1109/msmc.2016.2623867>

[13] Bisen, V. S. (2021, 9 februari). *What is Human in the Loop Machine Learning: Why & How Used in AI?* Medium. <https://medium.com/vsinghbisen/what-is-human-in-the-loop-machine-learning-why-how-used-in-ai-60c7b44eb2c0>

[14] Heikoop, D. D., Hagenzieker, M., Mecacci, G., Calvert, S., Santoni De Sio, F., & van Arem, B. (2019). *Human behaviour with automated driving systems: a quantitative framework for meaningful human control*. Theoretical Issues in Ergonomics Science, 20(6), 711–730. <https://doi.org/10.1080/1463922x.2019.1574931>

[15] Endsley, M. R., & Kiris, E. O. (1995). *The Out-of-the-Loop Performance Problem and Level of Control in Automation*. Human Factors: The Journal of the Human Factors and Ergonomics Society, 37(2), 381–394. <https://doi.org/10.1518/001872095779064555>

[16] Siau, K., & Wang, W. (2018). *Building Trust in Artificial Intelligence, Machine Learning, and Robotics*. https://www.researchgate.net/profile/Keng-Siau-2/publication/324006061_Building_Trust_in_Artificial_Intelligence_Machine_Learning_and_Robotics/links/5ab8744baca2722b97cf9d33/Building-Trust-in-Artificial-Intelligence-Machine-Learning-and-Robotics.pdf

[17] Wicks, A. C., Berman, S. L., & Jones, T. M. (1999). *The Structure of Optimal Trust: Moral and Strategic Implications*. The Academy of Management Review, 24(1), 99. <https://doi.org/10.2307/259039>

[18] Liu, G., Wang, Y., Orgun, M. A., & Lim, E. P. (2013). *Finding the Optimal Social Trust Path for the Selection of Trustworthy Service Providers in Complex Social Networks*. IEEE Transactions on Services Computing, 6(2), 152–167. <https://doi.org/10.1109/tsc.2011.58>

- [19] Fallon, C. K., Murphy, A. K. G., Zimmerman, L., & Mueller, S. (2010, januari). *The Calibration of Trust in an Automated System: A Sensemaking Process*.
https://www.researchgate.net/publication/251928358_The_calibration_of_trust_in_an_automated_system_A_sensemaking_process
- [20] André, Q., Carmon, Z., Wertenbroch, K., Crum, A., Frank, D., Goldstein, W., Huber, J., van Boven, L., Weber, B., & Yang, H. (2017). *Consumer Choice and Autonomy in the Age of Artificial Intelligence and Big Data*. *Customer Needs and Solutions*, 5(1-2), 28-37.
<https://doi.org/10.1007/s40547-017-0085-8>
- [21] B. (2020, 30 juli). *Psychologische reactantie*. Diekstra.NL.
<https://diekstra.nl/psychologische-reactantie/>
- [22] Homan, C. (2015, december). *Autonome wapens*. Armex.
https://www.clingendael.org/sites/default/files/2016-02/KNV_ARMEX_Artikel_P24-25.pdf
- [23] Santoni de Sio, F., & van den Hoven, J. (2018). *Meaningful Human Control over Autonomous Systems: A Philosophical Account*. *Frontiers in Robotics and AI*, 5, 1. <https://doi.org/10.3389/frobt.2018.00015>
- [24] Heikoop, D., Mecacci, G., Hagenzieker, M. P., & Santoni de Sio, F. (2018, juni). *Meaningful Human Control over Automated Driving Systems*.
https://www.researchgate.net/profile/Daniel-Heikoop/publication/326901327_Meaningful_Human_Control_over_Automated_Driving_Systems/links/5b6af4caa6fdcc87df6d9763/Meaningful-Human-Control-over-Automated-Driving-Systems.pdf
- [25] Guzman, A. L. (2018). "What is Human-Machine Communication, Anyway?" <https://www.pnas.org/content/pnas/92/22/9921.full.pdf>
- [26] Cohen, P. R., & Oviatt, S. L. (1995). *The role of voice input for human-machine communication*. *Proceedings of the National Academy of Sciences*, 92(22), 9921-9927. <https://doi.org/10.1073/pnas.92.22.9921>
- [27] Shubbak, M. H. (2013). *Self-Driving Cars: Legal, Social, and Ethical Aspects*. *SSRN Electronic Journal*, 9, <https://doi.org/10.2139/ssrn.2931847>
- [28] Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S., & Turini, F. (2019). *Meaningful Explanations of Black Box AI Decision Systems*. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33, 9780-9784. <https://doi.org/10.1609/aaai.v33i01.33019780>