



Ongelukken Met Autonome Voertuigen:

Wanneer het verschil tussen causale verantwoordelijkheid en juridische aansprakelijkheid problematisch blijkt

Auteur: Rosa Eggink (6447384)

Begeleider: Dr. Jespersen, B.T.F. (Bjørn)

Tweede lezer: Prof. dr. ir. Broersen, J.M. (Jan)

Universiteit Utrecht, Geesteswetenschappen

Bachelor Scriptie Kunstmatige Intelligentie (7,5 ECTS)

Abstract Autonome voertuigen beginnen een steeds realistischer begrip te worden. Echter, autonome voertuigen brengen nieuwe vraagstukken met zich mee. Deze scriptie richt zich op het vraagstuk: *Onder welke omstandigheden is het verschil tussen causale verantwoordelijkheid en juridische aansprakelijkheid problematisch voor ongelukken met autonome voertuigen?* Verschillende scenario's van ongelukken met autonome voertuigen en scenario's van causale verantwoordelijkheid en juridische aansprakelijkheid worden besproken en geanalyseerd. Vervolgens bespreekt deze scriptie de problemen met betrekking tot juridische aansprakelijkheid die ontstaan bij de relevante scenario's. Denkbare oplossingen van dragers van juridische aansprakelijkheid voor de aansprakelijkheidskloof worden geanalyseerd. Tenslotte wordt besproken waarom denkbare oplossingen binnen het huidige model van verantwoordelijkheid en aansprakelijkheid ethisch onjuist blijken en welk vervolgonderzoek hierdoor nodig is.

Sleutelwoorden Ethiek · Verantwoordelijkheid · Aansprakelijkheid · Aansprakelijkheidskloof · AI · Autonoom voertuig

Inhoud

Inleiding

1	Theoretisch kader	3
1.1	Causale verantwoordelijkheid en juridische aansprakelijkheid	3
1.2	Ethisch en technisch juist functioneren	4
2	Scenario's	7
2.1	Mogelijke scenario's van ongelukken met autonome voertuigen	7
2.2	Mogelijke scenario's van causale verantwoordelijkheid en juridische aansprakelijkheid	12
3	Ontstane problematiek en pogingen tot oplossingen	14
4	Discussie	17
	Conclusie	19
	Referenties	20

Inleiding

Autonome voertuigen beginnen een steeds realistischer begrip te worden. Daarnaast krijgt de Artificiële Intelligentie van autonome voertuigen steeds meer mogelijkheden. Vele bedrijven zoals Waymo en Tesla hebben inmiddels level 4 autonome voertuigen bereikt (SAE International, 2021). Level 4 autonome voertuigen zijn bijna volledig autonome voertuigen waarbij nog wel een mogelijkheid is tot menselijk ingrijpen. Er wordt gezegd dat Tesla tegen het einde van dit jaar een level 5 autonoom voertuig af heeft. Level 5 autonome voertuigen zijn volledig autonoom. In tegenstelling tot level 4 autonome voertuigen is er bij level 5 autonome voertuigen geen mogelijkheid meer voor menselijk ingrijpen. Volledig autonome voertuigen zoals level 5 voertuigen brengen hierdoor nieuwe vraagstukken met zich mee, binnen zowel de kunstmatige intelligentie als de ethiek. Een van deze vraagstukken gaat over de verantwoordelijkheid en aansprakelijkheid in het geval van ongelukken met deze voertuigen. Er is geen menselijke actor meer die de auto bestuurt. De bestuurder van de auto is in dit geval de AI. Is er in dit geval nog een morele actor die drager is van juridische aansprakelijkheid? Het huidige model van juridische aansprakelijkheid is gebaseerd op morele actoren die zo genoemde dragers kunnen zijn juridische aansprakelijkheid (Bivins, 2006). Wanneer de moreel verantwoordelijke menselijke actor wegvalt heeft dit grote gevolgen voor de juridische aansprakelijkheid. Er is immers juridische aansprakelijkheid nodig voor onder andere verzekeringen en schadevergoedingen in het geval van ongelukken. Het is daarom noodzakelijk om te bespreken hoe autonome voertuigen binnen ons huidige model van verantwoordelijkheid en aansprakelijkheid passen. Hierbij zal gekeken moeten worden naar wat er gebeurt wanneer een drager van causale verantwoordelijkheid niet een drager van juridische aansprakelijkheid blijkt (Sio & Nucci, 2017). Het deel van deze vraagstukken waar ik mij op zal richten heeft daarmee betrekking op ongelukken met autonome voertuigen. Ik zal dit bespreken door middel van een drieluik. In sectie 1 zal ik mij richten op causale verantwoordelijkheid, juridische aansprakelijkheid en het juist functioneren van de AI van een autonoom voertuig. Vervolgens zal ik in sectie 2 redeneren over mogelijke scenario's van ongelukken met autonome voertuigen en mogelijke scenario's van causale verantwoordelijkheid en juridische aansprakelijkheid. Hierbij zal ik ook een analyse geven van de relevantie van verschillende scenario's voor mijn these. Uiteindelijk bespreek ik in sectie 3 de ontstane ethische en juridische

problematiek en zal ik een poging doen tot het oplossen van deze problematiek. De these die ik in dit stuk zal beargumenteren is als volgt: het verschil tussen causale verantwoordelijkheid en juridische aansprakelijkheid is problematisch voor ongelukken met autonome voertuigen. De onderzoeksvraag die hierbij centraal staat is: *Onder welke omstandigheden is het verschil tussen causale verantwoordelijkheid en juridische aansprakelijkheid problematisch voor ongelukken met autonome voertuigen?* Voor het schrijven van deze scriptie heb ik 6669 woorden gebruikt.

1 Theoretisch kader

Voor mijn these kijk ik naar de problematiek met betrekking tot causale verantwoordelijkheid en op morele verantwoordelijkheid gebaseerde juridische aansprakelijkheid, die mogelijk ontstaat bij ongelukken met autonome voertuigen. Voordat ik mogelijke scenario's van ongelukken kan bespreken en voordat er gekeken kan worden naar mogelijke dragers van causale verantwoordelijkheid en juridische aansprakelijkheid, is het noodzakelijk om een notie van *causale verantwoordelijkheid* en een notie van *juridische aansprakelijkheid* te bespreken.

1.1 Causale verantwoordelijkheid en juridische aansprakelijkheid

De notie die ik aan zal houden van causale verantwoordelijkheid geldt als volgt. Iets of iemand *A* is een drager van causale verantwoordelijkheid over uitkomst *X* desda *X* causaal afhankelijk is van *A* (Sartorio, 2017). Dit wil zeggen dat aan iets of iemand causale verantwoordelijkheid voor een uitkomst wordt toegeschreven wanneer we weten dat deze iets of iemand de uitkomst veroorzaakt heeft (Willemsen, 2019). Stel een auto rijdt door een wijk heen en er steekt een vrouw over. De vrouw heeft geen voorrang en wordt hierdoor aangereden. We weten dat de vrouw causaal gezien haar eigen aanrijding veroorzaakt heeft, omdat ze zich als het ware als object voor causale impact door de auto heeft aangeboden. We kunnen dus stellen dat de vrouw een drager is van causale verantwoordelijkheid. Een ander voorbeeld betreft de roekeloze bestuurder. Wanneer hij op de weg rijdt en in zijn volledige roekeloosheid verscheidene ongelukken veroorzaakt, dan is hij causaal verantwoordelijk voor deze ongelukken.

Naast causale verantwoordelijkheid zal ik ook kijken naar juridische aansprakelijkheid. De notie van juridische aansprakelijkheid die ik aan zal houden geldt als volgt. Juridische aansprakelijkheid wordt in het algemeen gezien als het geven van schuld of eer aan iets of iemand voor zijn acties (Bivins, 2006). Het stellen van dragers van juridische aansprakelijkheid hangt nauw samen met morele verantwoordelijkheid, waarbij morele verantwoordelijkheid gebaseerd is op het hebben van een morele verplichting bij een actie. Dit wil zeggen dat wij meestal iets of iemand A zien als drager van juridische aansprakelijkheid over een uitkomst X , desda A drager is van morele verantwoordelijkheid over de actie Y die leidt tot deze uitkomst. Echter, morele verantwoordelijkheid en juridische aansprakelijkheid kunnen uit elkaar liggen. Zo kan een kind moreel verantwoordelijk zijn voor zijn/haar acties, maar zullen de ouders aansprakelijk gesteld worden voor de uitkomst. Ik zal voor het beargumenteren van mijn these aanhouden dat A drager is van juridische aansprakelijkheid over uitkomst X , desda A drager is van morele verantwoordelijkheid, een morele verplichting heeft, van de actie Y die leidt tot uitkomst X .

1.2 Ethisch en technisch juist functioneren

Ik zal nu focussen op het in kaart brengen van de mogelijke scenario's van ongelukken met autonome voertuigen. Hiervoor lijkt het nodig te zijn om drie aspecten op te stellen waar de mogelijke scenario's op gebaseerd zijn. Allereerst zal ik mij focussen op het ontwerp en de productie van het voertuig. Er kan hiervoor gesteld worden of het voertuig wel of niet goed ontworpen en/of geproduceerd is. Vervolgens kunnen we kijken naar de AI van het autonome voertuig. In dit geval zou je kijken of de AI functioneert naar behoren. Ten derde kunnen we kijken naar de context van het ongeluk: is er iets of iemand buiten het autonome voertuig om die een fout begaat?

Wanneer we kijken naar het tweede aspect, lijkt deze nog vrij ambigu. Ik stel dat de AI van een autonoom voertuig wel of niet goed functioneert. Echter, wanneer functioneert de AI van een autonoom voertuig wel of niet goed? We kunnen dit bekijken vanuit twee perspectieven. Aan de ene kant is er de ethische as en aan de andere kant is er de technische as. Op basis hiervan kunnen we twee vraagstukken analyseren:

- i. *Wanneer functioneert de AI van een autonoom voertuig technisch gezien wel of niet goed?*
- ii. *Wanneer functioneert de AI van een autonoom voertuig ethisch gezien wel of niet goed?*

Om te analyseren welke definitie van aspect twee relevant is voor mijn these, zal ik beide perspectieven analyseren. Allereerst zal ik kijken naar vraagstuk (i). Het is voor mijn these belangrijk om te weten wanneer de AI van een autonoom voertuig technisch gezien wel of niet goed functioneert. Laten we kijken naar het algoritme van de AI van het autonome voertuig. Het algoritme is zo ontworpen dat het dient als AI van het autonome voertuig. Echter, dit algoritme kan een onbedoelde fout in de uitwerking bevatten. We kunnen hierbij een onderscheid tussen type/token maken, waarbij het ontwerp van de AI het abstracte type van de AI is en de productie van dit ontwerp een aantal concrete tokens van dit type creëert. Zowel op type niveau als op token niveau kan er een technische fout optreden. Wanneer er op type niveau een technische fout is, zit er een technische fout in het ontwerp van het algoritme van de AI en geldt dit voor alle token AI's van dit type. Echter, wanneer er op token niveau een technische fout is, dan zit er een technische fout in de productie van het algoritme van een specifieke token van dit type AI. Er blijkt in dit geval een fout te zitten in het functioneren van het algoritme van de AI op type niveau; er zit een fout in het ontwerp van het algoritme. Het algoritme is ondanks deze fout nog wel een algoritme van de AI van het autonome voertuig. De technische fout treedt namelijk op in de uitvoering van het algoritme. We kunnen dus stellen dat we spreken van een AI van een autonoom voertuig, ondanks dat de AI niet goed functioneert. Daarnaast kan er een technische fout optreden op token niveau. In dit geval zit er een fout in de productie van de specifieke AI van dit autonome voertuig. Ook hiervoor geldt dat we nog altijd kunnen spreken van een AI van een autonoom voertuig, ondanks dat de AI technisch niet goed functioneert. Een AI die technisch niet goed functioneert bevat hiermee dus een fout in de uitvoering van haar algoritme, op type en/of token niveau. Voor het vervolg van mijn these zal ik, wanneer ik spreek over het technisch wel of niet goed functioneren van de AI van een autonoom voertuig, aanhouden dat de AI goed functioneert wanneer haar algoritme geen fout bevat in de uitvoering van de AI op zowel type als token niveau.

Nu we vraagstuk (i) hebben gedefinieerd, kunnen we kijken naar de definitie van vraagstuk (ii). Ik zal hierbij kijken naar het functioneren van de AI van een autonoom voertuig langs de ethische as. Als we willen definiëren wanneer de AI ethisch gezien wel of niet goed functioneert, zullen we wederom enkele voorwaarden moeten stellen. Een van deze voorwaarden zal ik baseren op het minimaliseren, oftewel het verminderen, van het aantal fouten gemaakt door een autonoom voertuig, ten opzichte van het elimineren, oftewel het helemaal wegnemen van fouten gemaakt door een autonoom voertuig. Wanneer we spreken over elimineren, stellen we dat een autonoom voertuig foutloos zal functioneren volgens bepaalde standaarden. We kunnen dan ook wel spreken van een ideologisch beeld, het ideale foutloze beeld, van het functioneren van autonome voertuigen. Hier tegenover staat het minimaliseren. Bij minimaliseren stellen we niet dat een autonoom voertuig ideologisch functioneert. We gaan er in dit geval vanuit dat er mogelijke situaties zijn waarbij foutloos functioneren onmogelijk is. Er zal hierbij altijd een situatie technisch gezien mogelijk zijn waarbij een ongeluk plaats vindt, ondanks dat het de AI van het autonome voertuig wel ethisch juist handelt. Natuurlijk is het niet ideaal dat ongelukken niet uitgesloten kunnen worden. Echter blijkt dit theoretisch wel gedoogd te moeten worden. Wanneer we kijken naar een ideologisch beeld van autonome voertuigen is er geen ruimte voor fouten en daarmee voor ongelukken. Gezien er geen ruimte is voor ongelukken, is een ideologisch beeld dan ook van autonome voertuigen niet relevant voor mijn these.

Ik zal me nu uitsluitend richten op een minimalistisch beeld van autonome voertuigen. Hiervoor is het noodzakelijk dat ik stel dat de AI van het voertuig ethisch juist handelt precies wanneer de veiligheid van de inzittenden van het voertuig een primair belang is in het algoritme. De veiligheid van mensen of dingen in de directe omgeving van het voertuig zijn hierbij altijd van secundair belang. Ondanks dat de AI van het voertuig precies juist handelt naar zijn primaire belang, is er nog wel een mogelijkheid tot ongelukken. Een voorbeeld hiervan is als volgt. Stel autonoom voertuig A en B hebben inzittenden en rijden op dezelfde snelweg. We nemen aan dat er iets gebeurt waardoor zowel de inzittenden van A, als de inzittenden van B mogelijk in gevaar zullen komen. Zowel de AI van A als de AI van B zal hierdoor handelen naar hun respectievelijke primaire belang. Beide autonome voertuigen zullen hun inzittenden beschermen en handelen hiermee ethisch juist. Als gevolg hiervan blijken de voertuigen niet te kunnen voorzien in hun secundaire belang, de veiligheid van de directe omgeving. Er gebeurt in

deze situatie een ongeluk, ondanks het ethisch juist handelen van de AI van de autonome voertuigen. Een technische oplossing hiervoor zou zijn dat zowel *A* als *B* voor de botsing tot stilstand komt. Echter, gaat deze oplossing in tegen een ander belang van de AI van de voertuigen; namelijk, dat een autonoom voertuig zijn inzittenden verplaatst en daarom juist niet stilstaat. Om die reden zal ik niet verder ingaan op deze oplossing en blijft de mogelijkheid tot ongelukken in dit geval bestaan. Het is dus wel mogelijk dat er bij een ethisch juist functionerende AI van een autonoom voertuig ongelukken voor kunnen komen.

Ik heb nu aangetoond dat een minimalistisch beeld van ongelukken met autonome voertuigen wel relevant is voor mijn these. Voor het gevolg van mijn argumentatie zal ik daarom een ethisch minimalistisch beeld van ongelukken met autonome voertuigen aanhouden. Hierbij heeft de AI van een autonoom voertuig altijd als primair belang de veiligheid van de inzittenden. Als secundair belang is er altijd de veiligheid van de directe omgeving van het autonome voertuig.

2 Scenario's

Nu ik het ethisch en technisch juist functioneren van de AI en de werking van causale verantwoordelijkheid en juridische aansprakelijkheid in kaart heb gebracht, zal ik mij focussen op het in kaart brengen van nieuw ontstane problematiek met betrekking tot ongelukken bij autonome voertuigen. Om deze problematiek in kaart te brengen, is het nodig dat we de mogelijke scenario's van een ongeluk afgaan om te zien wanneer deze problematiek optreedt. Ik zal hiervoor eerst een overzicht schetsen van de mogelijke scenario's met betrekking tot ongelukken bij autonome voertuigen. Vervolgens zal ik de verschillende scenario's afgaan om te onderzoeken welke scenario's voor mijn these relevant zijn. Hierna zal ik niet verder ingaan op de irrelevante scenario's en zal ik mijn focus volledig leggen op een analyse van de relevante scenario's aan de hand van causale verantwoordelijkheid en juridische aansprakelijkheid.

2.1 Mogelijke scenario's van ongelukken met autonome voertuigen

Allereerst is het van belang om de mogelijke scenario's van ongelukken op te stellen. Deze scenario's zullen we baseren op de volgende 3 condities die correct of incorrect

kunnen zijn: (i) het chassis van het autonome voertuig is goed gebouwd; (ii) de AI van het autonome voertuig functioneert zowel ethisch als technisch gezien juist; (iii) buiten het autonome voertuig om wordt er tijdens het ongeluk geen fout gemaakt (bijvoorbeeld, iemand steekt over waar deze persoon geen voorrang heeft en wordt hierdoor aangereden).

De mogelijke scenario's bestaan uit de mogelijke combinaties van waardes die we kunnen toekennen aan proposities voor condities (i), (ii) en (iii). Aan ieder van de proposities kan een waarde *waar* of *onwaar* worden toegekend in overeenstemming met de erbij horende waarheidswaarden. Zo kan het chassis van het autonome voertuig *wel/niet* goed geproduceerd zijn en hiermee is de propositie voor conditie (i) *waar*, waardoor er aan conditie (i) voldaan wordt. De waarde die een conditie toegekend krijgt is logisch onafhankelijk van de waarde van de andere condities in het desbetreffende scenario. Wanneer de propositie voor (i) waar is, dan heeft dit geen directe consequentie voor de waarde van (ii) en (iii). Zo geldt dit ook voor de andere condities, wier waardes tevens geen directe consequenties voor elkaar hebben. Wanneer we deze mogelijke combinaties van waardes maken krijgen we zo de volgende acht scenario's:

1. i) Het chassis van het autonome voertuig is goed geproduceerd. *waar*
ii) De AI van het autonome voertuig functioneert zowel ethisch als technisch gezien juist. *waar*
iii) Buiten het autonome voertuig om wordt er geen fout gemaakt. *waar*
2. i) Het chassis van het autonome voertuig is goed geproduceerd. *waar*
ii) De AI van het autonome voertuig functioneert zowel ethisch als technisch gezien juist. *onwaar*
iii) Buiten het autonome voertuig om wordt er geen fout gemaakt. *waar*
3. i) Het chassis van het autonome voertuig is goed geproduceerd. *waar*
ii) De AI van het autonome voertuig functioneert zowel ethisch als technisch gezien juist. *waar*
iii) Buiten het autonome voertuig om wordt er geen fout gemaakt. *onwaar*
4. i) Het chassis van het autonome voertuig is goed geproduceerd. *waar*

- ii) De AI van het autonome voertuig functioneert zowel ethisch als technisch gezien juist. *onwaar*
 - iii) Buiten het autonome voertuig om wordt er geen fout gemaakt. *onwaar*
5. i) Het chassis van het autonome voertuig is goed geproduceerd. *onwaar*
- ii) De AI van het autonome voertuig functioneert zowel ethisch als technisch gezien juist. *waar*
 - iii) Buiten het autonome voertuig om wordt er geen fout gemaakt. *waar*
6. i) Het chassis van het autonome voertuig is goed geproduceerd. *onwaar*
- ii) De AI van het autonome voertuig functioneert zowel ethisch als technisch gezien juist. *onwaar*
 - iii) Buiten het autonome voertuig om wordt er geen fout gemaakt. *waar*
7. i) Het chassis van het autonome voertuig is goed geproduceerd. *onwaar*
- ii) De AI van het autonome voertuig functioneert zowel ethisch als technisch gezien juist. *waar*
 - iii) Buiten het autonome voertuig om wordt er geen fout gemaakt. *onwaar*
8. i) Het chassis van het autonome voertuig is goed geproduceerd. *onwaar*
- ii) De AI van het autonome voertuig functioneert zowel ethisch als technisch gezien juist. *onwaar*
 - iii) Buiten het autonome voertuig om wordt er geen fout gemaakt. *onwaar*

Ik heb nu de mogelijke scenario's opgesteld met betrekking tot ongelukken bij autonome voertuigen. Nu wil ik bepalen welke scenario's relevant zijn voor onze these. Echter, wanneer kunnen we spreken over een *relevant* scenario? Mijn these baseert zich op nieuw ontstane problematiek bij ongelukken wanneer er een autonoom voertuig betrokken raakt. We kunnen spreken van nieuw ontstane problematiek met betrekking tot ongelukken met autonome voertuigen, wanneer we in ons huidige model geen manier van omgaan hebben die een oplossing biedt voor het desbetreffende scenario. In andere woorden, een scenario is relevant desda het autonome voertuig zorgt voor een vraag naar een nieuwe manier van omgaan met verantwoordelijkheid bij ongelukken. Voor onze scenario's betekent dit dat

we op zoek moeten gaan naar welke van deze scenario's een nieuwe situatie creëert die buiten het huidige model valt.

Laten we nu terugkeren naar de aspecten waarop we onze scenario's gebaseerd hebben. Allereerst hebben we *(i)* Het chassis van het autonome voertuig is goed geproduceerd. Stel we nemen aan dat de propositie voor *(i)* onwaar is en dat *(ii)* en *(iii)* waar zijn. Ons autonome voertuig rijdt door een wijk heen en komt bij een zebrapad aan waar een klein meisje oversteekt. Het autonome voertuig probeert te remmen, maar de remmen werken niet. Als gevolg hiervan wordt het meisje aangereden. In dit geval is het chassis niet goed geproduceerd. Geldt er voor dit scenario dat het scenario anders behandeld moet worden omdat het een autonoom voertuig is? Nee, dit ongeluk kan immers op dezelfde manier om dezelfde oorzaken plaatsvinden wanneer we over een niet-autonoom voertuig spreken. We kunnen dus stellen dat alle scenario's waarbij *(i)* onwaar is, niet relevant zijn voor onze these. Scenario's 5, 6, 7 en 8 kunnen we nu afstrepen.

Ik kan nu verder gaan naar aspect *(ii)*: de AI van het autonome voertuig functioneert zowel ethisch als technisch gezien juist. Ik neem wederom aan dat de propositie onwaar is en dat de proposities voor *(i)* en *(iii)* waar zijn. Hierbij ontstaan er drie mogelijkheden. De AI van het autonome voertuig functioneert ethisch juist en technisch onjuist. De AI van het autonome voertuig functioneert ethisch onjuist en technisch juist. De AI van het autonome voertuig functioneert zowel ethisch als technisch onjuist. Ik focus eerst op de situatie waarbij de AI ethisch juist functioneert, maar technisch onjuist functioneert. Ons autonome voertuig rijdt weer door een wijk heen en komt bij een zebrapad aan waar een klein meisje oversteekt. Het autonome voertuig zou volgens het algoritme van de AI moeten remmen. Echter, de uitvoering van het algoritme zorgt ervoor dat het autonome voertuig niet remt. Er blijkt namelijk een fout te zitten in het ontwerp van het algoritme, op type niveau, waardoor de AI van het autonome voertuig technisch gezien niet juist functioneert. Iedere token van dit type AI zou dus onder dezelfde materiele omstandigheden hetzelfde rijgedrag vertonen. Als gevolg hiervan wordt het meisje wederom aangereden. Wanneer we kijken naar de ontstane situatie lijkt het op het eerste oog dat deze situatie niet binnen het huidige model van ongelukken en verantwoordelijkheid past. In een situatie waarbij er geen sprake was van een autonoom voertuig, was er immers ook geen sprake van een AI waarvan het algoritme niet doet wat het moet doen. Echter, als we kijken naar de oorzaak van de fout, dan zouden we dit toch

kunnen vergelijken met een niet- autonoom voertuig. In onze situatie zit er een fout in het algoritme van de AI. De AI is in deze situatie de bestuurder van het voertuig. In een situatie met een niet-autonoom voertuig zouden we de oorzaak van het ongeluk bij de bestuurder kunnen zoeken. De bestuurder zou immers in dat geval niet gedaan hebben wat deze in theorie zou moeten doen. Als we nu weer kijken naar onze situatie met het autonome voertuig, lijkt onze situatie opeens toch erg op een situatie zonder autonoom voertuig. We zouden in principe binnen het huidige model dan de oorzaak kunnen zoeken bij de AI van het voertuig. Wanneer erbij technisch onjuist functioneren de oorzaak gezocht kan worden bij de AI van het voertuig volgens het huidige model van aansprakelijkheid, dan geldt dit automatisch ook voor situaties waarbij de AI en/of ethisch onjuist functioneert. Het onjuist functioneren is hier tevens oorzakelijk te vinden bij de AI van het autonome voertuig. Er geldt dus dat voor alle situaties waarvoor *(ii)* onwaar is, dat we het huidige model van verantwoordelijkheid en aansprakelijkheid kunnen toepassen. Nu we het scenario met de propositie van *(ii)* als onwaar kunnen plaatsen binnen het huidige model van ongelukken, kunnen we stellen dat alle scenario's waarvoor geldt dat de propositie van *(ii)* onwaar is, irrelevant zijn voor mijn these. Daarom strepen we nu ook scenario 2 en 4 af.

Tenslotte kijken we naar aspect *(iii)*: buiten het autonome voertuig om wordt er geen fout gemaakt. We nemen tevens voor conditie *(iii)* aan dat de propositie onwaar is en dat de proposities voor *(i)* en *(ii)* waar zijn. Ons autonome voertuig rijdt voor dit scenario weer door een wijk heen. In dit geval is er geen sprake van een zebrapad met een klein meisje. Daarentegen is er nu een volwassen vrouw die oversteeft zonder dat zij hier voorrang heeft. De vrouw wordt aangereden door het autonome voertuig. De oorzaak is in deze situatie niet te herleiden tot het chassis of de AI van het autonome voertuig. Wel kunnen we de oorzaak herleiden tot de volwassen vrouw die een fout maakte door over te steken. In een situatie met een niet-autonoom voertuig hadden we de oorzaak op dezelfde manier kunnen herleiden tot de volwassen vrouw. We kunnen dus stellen dat we deze situatie niet anders hoeven aan te pakken ondanks dat er een autonoom voertuig bij het ongeluk betrokken is. Dit scenario past binnen ons huidige model van ongelukken en verantwoordelijkheid en is daarmee irrelevant voor mijn these. We kunnen nu dus ook scenario 3 afstrepen.

Het blijkt nu dus dat er nog alleen scenario 1 relevant is voor mijn these: louter een scenario, dus, waarbij (i) het chassis van het voertuig van het autonome voertuig is goed geproduceerd, (ii) de AI van het autonome voertuig functioneert zowel ethisch als technisch gezien juist, en (iii) buiten het autonome voertuig om wordt er geen fout gemaakt allen waar zijn, is relevant voor de these die ik beargumenteer. Voor het vervolg van mijn argumentatie neem ik dan ook scenario 1 aan. De situatie volgend uit scenario 1 die ik hiervoor aan zal nemen is als volgt. Ons autonome voertuig rijdt door een wijk. Er gebeurt een ongeluk waarbij een volwassen vrouw wordt aangereden, zonder dat de vrouw in kwestie een fout maakt. Het chassis van het autonome voertuig is goed geproduceerd. Het algoritme van de AI van het voertuig doet wat het zou moeten doen. En, er is geen iets of iemand die een fout maakt gedurende het ongeluk. Toch is het noodzakelijk dat er wordt uitgezocht wie of wat er *juridisch aansprakelijk* is. Er is immers een drager van juridische aansprakelijkheid nodig voor onder andere verzekeringen en schadevergoedingen, maar ook voor het toekomstige minimaliseren van ongelukken en het verbeteren van verkeersveiligheid. Ik zal in het volgende deel op zoek gaan naar de mogelijke scenario's voor verantwoordelijkheid en juridische aansprakelijkheid binnen de aangenomen context.

2.2 Mogelijke scenario's van causale verantwoordelijkheid en juridische aansprakelijkheid

Ik zal nu proberen te onderzoeken wie of wat er mogelijk juridisch aansprakelijk gesteld zouden kunnen worden. Hiervoor is het nodig om een relevant scenario van causale verantwoordelijkheid en juridische aansprakelijkheid vast te stellen voor de argumentatie van mijn these. Allereerst zal ik hiervoor alle mogelijke scenario's van verantwoordelijkheid en juridische aansprakelijkheid opstellen. Hierna zal ik analyseren welke scenario's relevant zijn binnen de aangenomen context.

Wanneer we spreken over dragers van causale verantwoordelijkheid en juridische aansprakelijkheid, dan hoeven dit niet dezelfde dragers te zijn. Het is mogelijk dat iets of iemand wel een drager van causale verantwoordelijkheid is, maar niet juridisch aansprakelijk gesteld kan worden. Tevens is het mogelijk dat we iets of iemand juridisch aansprakelijk stellen, ondanks dat dit niet een drager van causale verantwoordelijkheid

is. Er ontstaan bij verantwoordelijkheid en juridische aansprakelijkheid daarom drie scenario's:

1. De drager van causale verantwoordelijkheid is tevens de drager van juridische aansprakelijkheid.
2. De drager van juridische aansprakelijkheid is niet een drager van causale verantwoordelijkheid.
3. De drager van causale verantwoordelijkheid is niet een drager van juridische aansprakelijkheid.

Voor een scenario om relevant te zijn voor mijn these, is van belang dat er een probleem optreedt met betrekking tot het vaststellen van de aansprakelijkheid. Laten we als eerst scenario 1 aannemen: de drager van de causale verantwoordelijkheid is hier tevens de drager van juridische aansprakelijkheid. Stel er gebeurt een ongeluk waarbij een autonoom voertuig betrokken is. Laten we alleen voor het doel van de these er vanuit gaan dat de drager van causale verantwoordelijkheid en tevens de drager van juridische aansprakelijkheid in dit geval de producent van het autonome voertuig is. We kunnen dus in dit geval iets of iemand aansprakelijk stellen voor de gevolgen van het ongeluk. We hebben immers een drager van juridische aansprakelijkheid. In dit geval is er dan ook geen probleem met betrekking tot de aansprakelijkheid bij een ongeluk met een autonoom voertuig. Dit scenario blijkt hiermee dus niet relevant voor mijn these en zal ik nu dus ook afstrepen.

Laten we nu scenario 2 bekijken. We nemen hiervoor aan dat de drager van juridische aansprakelijkheid niet een drager van causale verantwoordelijkheid is. Stel er gebeurt weer een ongeluk met een autonoom voertuig. We nemen puur en alleen voor het doel van de these aan dat de drager van juridische aansprakelijkheid in dit geval de producent van het autonome voertuig is. De producent van het autonome voertuig blijkt in de aangenomen situatie echter niet een drager van causale verantwoordelijkheid te zijn. De drager van juridische aansprakelijkheid is hiermee geen drager causale van verantwoordelijkheid in deze situatie. Maar, toch zorgt dit niet voor problematiek met betrekking tot het aansprakelijk stellen. We hebben een drager van juridische

aansprakelijkheid, dus kunnen we ook iets of iemand juridisch aansprakelijk stellen. Bij scenario 2 ontstaat er daarom geen problematiek met betrekking tot aansprakelijkheid bij een ongeluk met een autonoom voertuig. Ook scenario 2 blijkt hiermee niet relevant voor mijn these en zal ik daarom afstrepen.

Tenslotte kunnen we scenario 3 aannemen. De drager van causale verantwoordelijkheid is bij dit scenario niet een drager van juridische aansprakelijkheid. We stellen hiervoor wederom dat er een ongeluk gebeurt waarbij een autonoom voertuig betrokken is. Tevens stellen we, puur en alleen voor het doel van de these, dat de drager van causale verantwoordelijkheid in dit geval de producent van het autonome voertuig is. Echter, de producent van het autonome voertuig blijkt in deze situatie niet de drager van juridische aansprakelijkheid. Gezien er geen drager van juridische aansprakelijkheid is, kan er niet iets of iemand aansprakelijk gesteld worden voor de gevolgen van het ongeluk. Dit scenario is het punt waarbij de problematiek omtrent aansprakelijkheid bij ongelukken met autonome voertuigen naar voren lijkt te komen. Bij scenario 3 kunnen we dus zeggen dat er problematiek op lijkt te treden in het geval van ongelukken met autonome voertuigen. Hierdoor kunnen we stellen dat scenario 3 wel relevant is voor mijn these.

Voor het vervolg van de argumentatie voor mijn these zal ik dan ook scenario 3 aannemen. Op basis van scenario 1 van *scenario's van ongelukken met autonome voertuigen* en scenario 3 van *scenario's van causale verantwoordelijkheid en juridische aansprakelijkheid*, zal ik de mogelijk ontstane vormen van problematiek met betrekking tot verantwoordelijkheid en aansprakelijkheid aanhalen. Hierna zal ik enkele oplossingen voor deze problematiek aandragen.

3 Ontstane problematiek en pogingen tot oplossingen

Met de analyse van mogelijke scenario's van ongelukken met autonome voertuigen en de analyse van mogelijke scenario's met causale verantwoordelijkheid en juridische aansprakelijkheid heb ik een relevant scenario aangenomen voor mijn these. Het doel is om de nieuw ontstane problematiek met betrekking tot causale verantwoordelijkheid en juridische aansprakelijkheid bij ongelukken met autonome voertuigen in kaart te brengen. Het scenario dat hiervoor relevant bleek bestaat uit twee subscenario's met betrekking tot

het type ongeluk en de causale verantwoordelijkheid met juridische aansprakelijkheid. Ten eerste is er het scenario met betrekking tot het mogelijk type ongeluk. Hiervoor heb ik aangetoond relevant te zijn en daarmee aangenomen scenario 1 waarbij (i) *Het chassis van het autonome voertuig is goed geproduceerd* waar is, (ii) *De AI van het autonome voertuig functioneert zowel ethisch als technisch gezien juist* waar is en (iii) *Buiten het autonome voertuig om wordt er geen fout gemaakt* waar is. Vervolgens heb ik, met betrekking tot causale verantwoordelijkheid en juridische aansprakelijkheid, aangetoond relevant te zijn en daarmee aangenomen scenario 3 waarvoor geldt: *de drager van causale verantwoordelijkheid is niet een drager van juridische aansprakelijkheid*. Het scenario dat hiermee gecreëerd is ziet er nu voor het vaststellen van dragers van juridische aansprakelijkheid als volgt uit: de drager van juridische aansprakelijkheid is niet, volgens het huidige model van aansprakelijkheid en verantwoordelijkheid, vast te stellen. Echter, hoe ziet een situatie binnen dit scenario er dan uit? Ik zal hiervoor terug keren naar een eerdergenoemd voorbeeld.

Laten we weer kijken naar een situatie waarbij er sprake is van autonoom voertuig A en autonoom voertuig B. We weten dat voor zowel A als B geldt dat (i) *Het chassis van het autonome voertuig is goed geproduceerd*, (ii) *De AI van het autonome voertuig functioneert zowel ethisch als technisch gezien juist* en (iii) *Buiten het autonome voertuig om wordt er geen fout gemaakt*. Stel A en B rijden wederom over een snelweg. Er zijn op dit moment meerdere dingen die kunnen gebeuren waardoor A en B in een ongeluk belanden. Ieder van de mogelijke gebeurtenissen behoort tot de categorie *onvermijdelijke factoren*. Zo kan er door het weer aquaplaning ontstaan of valt er een boom over de weg als gevolg van bliksem. In ieder van deze gevallen zal zowel A als B het primaire belang hebben om zijn inzittenden te beschermen. Het beschermen van de directe omgeving is hierbij secundair belang. Als gevolg hiervan is het mogelijk dat er ongelukken zijn waarbij A en B betrokken zijn. A en B zijn in deze situatie dragers van causale verantwoordelijkheid. Het ongeluk is immers causaal afhankelijk van de acties van A en B. Echter, gezien zowel A als B niet gezien kan worden als drager van juridische aansprakelijkheid, zal er niet iets of iemand aansprakelijkheid gesteld kunnen worden. Het huidige model van verantwoordelijkheid en aansprakelijkheid biedt hiermee geen ruimte voor het omgaan met nieuw ontstane scenario's van ongelukken met autonome voertuigen. We kunnen hierbij dus spreken van een probleem: namelijk het probleem dat er niet voor ieder scenario van ongelukken met autonome voertuigen een drager van

juridische aansprakelijkheid vast te stellen is; en daarmee, dat er niet voor ieder scenario van ongelukken met autonome voertuigen iets of iemand aansprakelijk gesteld kan worden. Dit ondanks dat we het wel noodzakelijk stellen dat er iets of iemand aansprakelijk gesteld wordt. Het is voor ongelukken met autonome voertuigen dan ook onoverkomelijk dat er een oplossing aangedragen zal moeten worden voor scenario's die niet binnen het huidige model van verantwoordelijkheid en juridische aansprakelijkheid vallen.

Ik zal nu een poging doen tot het aandragen van oplossingen voor het vastgestelde probleem. Ik zal dit probleem proberen op te lossen door de aansprakelijkheid aan verschillende mensen/instaties toe te schrijven, ondanks dat deze geen dragers van juridische aansprakelijkheid zijn. Ik zal hiervoor kijken naar 3 mogelijke mensen/instaties om juridisch aansprakelijk te stellen. Zo is er de *producent* van het autonome voertuig, de *ontwikkelaar* van de AI van het autonome voertuig en de *eigenaar* van het autonome voertuig.

Ten eerste kunnen we overwegen de producent van het autonome voertuig juridisch aansprakelijk te stellen. De producent van het autonome voertuig zou dan juridische aansprakelijkheid toegeschreven worden, wanneer er geen drager van aansprakelijkheid vast gesteld kan worden. Echter, je kan stellen dat de producent van het autonome voertuig zijn of haar verantwoordelijkheid bij het verkopen van het autonome voertuig overdraagt aan een koper. De producent is immers geen gebruiker van het voertuig en zou daarmee geen ethische verplichting hoeven dragen over het gebruik. Daarnaast kan er niet van de producenten verwacht worden dat zij mogelijke kopers doorlichten. Mogelijke kopers en gebruikers zouden wellicht op een niet verantwoorde manier met het gevaarlijke artefact omgaan.

Ten tweede kunnen we overwegen de ontwikkelaar van de AI van het autonome voertuig juridisch aansprakelijk te stellen, ondanks dat deze geen drager is van juridische aansprakelijkheid. De ontwikkelaar is immers ethisch betrokken bij de implementatie en het ontwerp van het algoritme van de AI van het autonome voertuig. Maar ook hier zijn er overwegingen om deze oplossing niet aan te nemen. De ontwikkelaar van de AI is dan wel ethisch betrokken bij de implementatie van het algoritme van de AI, de ontwikkelaar draagt net als de producent van het autonome voertuig niet het risico voor het gebruik. Er

kan wederom niet verwacht worden dat de ontwikkelaar van de AI potentiële kopers/gebruikers doorlicht. Wellicht wordt er niet op een verantwoorde manier met het gevaarlijke artefact omgegaan.

We lijken nu alleen te kunnen focussen op de derde optie van mogelijke toeschrijvingen van juridische aansprakelijkheid. Dit zou betekenen dat de eigenaar van het autonome voertuig juridisch aansprakelijk gesteld wordt. De eigenaar van het autonome voertuig neemt namelijk het risico om het autonome voertuig te gebruiken. Echter, is het moreel juist om de eigenaar van het autonome voertuig volledig aansprakelijk te stellen voor het ongeluk? De eigenaar van het autonome voertuig heeft immers geen invloed gehad op zowel de productie van het autonome voertuig als de ontwikkeling van de AI van het autonome voertuig. De eigenaar van het autonome voertuig heeft hiermee geen morele keuzes gemaakt met betrekking tot het produceren van het autonome voertuig en het ontwikkelen van de AI. Wel heeft de eigenaar gekozen voor dit gevaarlijke artefact met bepaalde kenmerken. Bij het maken van deze keuze heeft de eigenaar hiermee de morele verplichting om een ethisch juiste keuze te maken.

Geen van de mogelijke oplossingen voor de drager van juridische aansprakelijkheid lijkt toereikend voor ons probleem. Voor zowel de producent, als de ontwikkelaar en de eigenaar geldt dat we het niet kunnen verantwoorden om ze aansprakelijk te stellen. Ze zijn immers geen dragers van morele verantwoordelijkheid en zullen daarmee geen ethisch verantwoorde dragers van juridische aansprakelijkheid kunnen zijn. Een oplossing voor het probleem van een drager van juridische aansprakelijkheid is daarom niet te vinden door juridische aansprakelijkheid toe te schrijven aan een mens of instantie zoals de producent, de ontwikkelaar of de eigenaar.

4 Discussie

Ik heb nu aangetoond dat er een kloof ontstaat met betrekking tot het verschil tussen causale verantwoordelijkheid en juridische aansprakelijkheid bij ongelukken met autonome voertuigen. Dit is ook wel een *aansprakelijkheidskloof* te noemen. Tevens heb ik aangetoond dat we geen ethisch juiste oplossing voor dit probleem kunnen vinden door iets/iemand aansprakelijkheid toe te schrijven die geen drager is van juridische

aansprakelijkheid. Wanneer we verantwoordelijkheid willen nemen voor een oplossing voor de mogelijke problematiek van ongelukken met autonome voertuigen, dan zullen we niet moeten kijken hoe we scenario's met autonome voertuigen kunnen aanpassen aan het huidige model van verantwoordelijkheid en aansprakelijkheid. Echter, we zullen moeten kijken naar de oorzaak van het probleem dat ontstaat met autonome voertuigen en aansprakelijkheid. Het huidige model van juridische aansprakelijkheid is gebaseerd op morele actoren. Het cruciale punt is dat met de substitutie van een AI voor een morele actor er een probleem ontstaat voor de juridische aansprakelijkheid. Wanneer we willen zoeken naar een oplossing voor dit probleem zullen we daarom moeten kijken naar de opbouw van het huidige model van verantwoordelijkheid en aansprakelijkheid.

Zoals reeds aangetoond, kan een oplossing niet binnen het huidige model gevonden worden. Daarom vereist een oplossing een wellicht verre gaande bijstelling van het model. Een bijstelling van dit model zou kunnen neerkomen op het toevoegen van nieuwe/extra factoren en/of actoren. Echter, wanneer we het huidige model van verantwoordelijkheid en aansprakelijkheid aanhouden, zal er geen ethisch juiste oplossing zijn voor het probleem. Dit lijkt in eerste instantie een harde conclusie. Echter, wanneer ik kijk naar oplossingen aangedragen in eerdere onderzoeken dan lijken ze tegen dezelfde problematiek aan te lopen. Verdiesen, Santoni de Sio en Dignum schrijven in hun onderzoek naar autonome wapens eveneens over aansprakelijkheidskloven door een tekort aan beheersing (Verdiesen, et al., 2021). Alhoewel hun onderzoek gericht is op autonome wapens in tegenstelling tot autonome voertuigen, is de theoretische problematiek wel vergelijkbaar. Dit is tevens terug te vinden in onderzoek van Santoni de Sio en Mecacci, gericht op AI en verantwoordelijkheidskloven (Santoni de Sio & Mecacci, 2020). Dit onderzoek toont aan dat er niet alleen voor autonome voertuigen en autonome wapens een verantwoordelijkheidskloof bestaat, maar ook voor AI in de algemenere zin. Daarnaast schrijven Hevelke en Nida-Rümelin dat we mogelijk de gebruikers van autonome voertuigen collectief verantwoordelijk moeten stellen voor schade aangericht door autonome voertuigen (Hevelke & Nida-Rümelin, 2014), en wel losstaand van de causale invloed die zij wel of niet hadden op het gedrag van het autonome voertuig. Echter, Hevelke en Nida-Rümelin stellen hierbij ook dat de morele verantwoordelijkheid nooit volledig weg moet worden genomen van de producent van het voertuig. Het is belangrijk om een deel van de verantwoordelijkheid bij de producent van het voertuig en de ontwikkelaar van de AI te houden om zo doorlopende ontwikkeling en verbetering van autonome voertuigen te stimuleren. Op basis van de bestaande

literatuur kan ik stellen dat mijn conclusie een harde maar zeker geen extreme conclusie is. Er is daarom een behoefte om de aansprakelijkheidskloof voor niet alleen autonome voertuigen, maar ook AI in de algemenere zin verder te onderzoeken. Vervolgonderzoeken zouden zich bijvoorbeeld kunnen richten op de opbouw van het huidige model van verantwoordelijkheid en aansprakelijkheid, met focus op hoe er hierbinnen ruimte gecreëerd kan worden voor AI.

Conclusie

De these die centraal stond in deze scriptie was: het verschil tussen causale verantwoordelijkheid en juridische aansprakelijkheid is problematisch voor ongelukken met autonome voertuigen. Ik heb in sectie 1.1 gedefinieerd wat het is om een drager van causale verantwoordelijkheid te zijn en wat het is om een drager van juridische aansprakelijkheid te zijn. In sectie 1.2 heb ik vervolgens gedefinieerd wanneer we kunnen spreken van het ethisch en technisch juist functioneren van een AI. Aan de hand van deze definities heb ik in sectie 2.1 scenario's opgesteld van mogelijke ongelukken met autonome voertuigen. Na een analyse op basis van relevantie heb ik één scenario aan kunnen nemen, waarna ik mij in 2.2 gefocust heb op het opstellen en analyseren van scenario's van causale verantwoordelijkheid en juridische aansprakelijkheid. Nadat ik aan de hand van deze analyse een definitief scenario had aangenomen, heb ik een poging gedaan de problematiek in kaart te brengen en oplossingen aan te dragen. Tenslotte heb ik beargumenteerd waarom er op dit moment (nog) geen mogelijke oplossing is voor de aansprakelijkheidskloof bij autonome voertuigen. In deze scriptie heb ik daarmee aangetoond dat het verschil tussen causale verantwoordelijkheid en juridische aansprakelijkheid problematisch en diepgeworteld is voor ongelukken met autonome voertuigen. Om een oplossing voor dit probleem te vinden blijkt er nader onderzoek nodig te zijn. Onderzoeken die hiervoor nodig zijn zouden zich kunnen richten op de invloed van AI als substitutie van een morele actor op morele verantwoordelijkheid en juridische aansprakelijkheid. Hierbij zal onderzocht moeten worden hoe er ruimte gemaakt kan worden voor autonome artefacten binnen het huidige model van verantwoordelijkheid en aansprakelijkheid.

Referenties

- Bivins, T. H. (2006). Responsibility and Accountability. In Fitzpatrick, K.R. & Bronstein, C. (Eds). *Ethics in public relations: Responsible advocacy*, (pp. 19–38). SAGE publications. <https://doi.org/10.4135/9781452204208.n2>
- 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>
- Hevelke, A., & Nida-Rümelin, J. (2014). Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis. *Science and Engineering Ethics*, 21(3), 619–630. <https://doi.org/10.1007/s11948-014-9565-5>
- Keeling, G. (2019). Why Trolley Problems Matter for the Ethics of Automated Vehicles. *Science and Engineering Ethics*, 26(1), 293–307. <https://doi.org/10.1007/s11948-019-00096-1>
- Nucci, E. D., & Sio, F. S. D. (2019). *Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on Remotely Controlled Weapons (Emerging Technologies, Ethics and International Affairs)* (1ste ed.). Routledge.
- SAE International. (2021, april). (R) Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. https://www.sae.org/standards/content/j3016_202104/. Geraadpleegd op 8 mei 2021
- Santoni De Sio, F. (2017). Killing by Autonomous Vehicles and the Legal Doctrine of Necessity. *Ethical Theory and Moral Practice*, 20(2), 411–429. <https://doi.org/10.1007/s10677-017-9780-7>
- Santoni De Sio, F., & Di Nucci, E. (2017). Pushing the Margins of Responsibility: Lessons from Parks' Somnambulistic Killing. *Neuroethics*, 11(1), 35–46. <https://doi.org/10.1007/s12152-017-9311-1>
- Santoni De Sio, F., & Mecacci, G. (2021). Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. *Philosophy & Technology*. Published. <https://doi.org/10.1007/s13347-021-00450-x>

- Santoni De Sio, F., & Van den Hoven, J. (2018). Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI*, 5(15). <https://doi.org/10.3389/frobt.2018.00015>
- Sartorio, C. (2007). Causation and Responsibility. *Philosophy Compass*, 2(5), 749–765. <https://doi.org/10.1111/j.1747-9991.2007.00097.x>
- Verdiesen, I., Santoni De Sio, F., & Dignum, V. (2020). Accountability and Control Over Autonomous Weapon Systems: A Framework for Comprehensive Human Oversight. *Minds and Machines*, 31(1), 137–163. <https://doi.org/10.1007/s11023-020-09532-9>