



# Morele overwegingen voor het inzetten van predictive policing algoritmes

Bachelorthesis Kunstmatige Intelligentie (7,5 ECTS)

**Auteur: Jolien van Vliet (6198724)**

**Begeleidster: dr. Janneke van Lith**  
**Tweede lezer: dr. Baptist Liefoghe**

## Abstract

Ondanks dat 'predictive policing' algoritmes de criminaliteit doen dalen, brengen ze de nodige morele problemen met zich mee. Person-based 'predictive policing' algoritmes kunnen bijdragen aan het versterken van etnisch profileren en de schending van privacy. Location-based 'predictive policing' algoritmes kunnen stigmatisering tot gevolg hebben. Bij beide soorten treden er 'feedback loops' op, waardoor de nadruk op bepaalde personen of wijken blijft liggen. Dit literatuuronderzoek brengt de morele aspecten die spelen rondom 'predictive policing' algoritmes in kaart en verkent hoe 'predictive policing' algoritmes verantwoord zouden kunnen worden ingezet in de maatschappij. Uit dit onderzoek komt dat 'predictive policing' algoritmes nog niet ver genoeg ontwikkeld zijn om toegepast te kunnen worden in de praktijk zonder ethische bezwaren. Algoritmes blijken het meest te worden vertrouwd wanneer ze te modificeren zijn. Het is dus nodig voor 'predictive policing' algoritmes om aanpassingen te kunnen maken. Hier is transparantie voor nodig, dat ook bijdraagt aan het waarborgen van privacy en het scheppen van duidelijkheid naar de samenleving toe. Eventueel vervolgonderzoek zou zich kunnen richten op de statistische overwegingen die ook meespelen rondom dit vraagstuk, zoals de betrouwbaarheid van de 'predictive policing' algoritmes.

**Keywords:** 'predictive policing' algoritmes, 'feedback loops'

## Inhoudsopgave

Inleiding	3
Sectie 1: Algemeen vertrouwen algoritmes	5
Sectie 2: Soorten predictive policing algoritmes	7
Sectie 3: Positieve en negatieve aspecten	8
Sectie 4: Morele gevolgen van feedback loops	11
Sectie 5: Eisen voor een veilig gebruik van 'predictive policing' algoritmes	14
Sectie 6: Conclusie	17
Bibliografie	19

## Inleiding

De politie in Los Angeles heeft tussen 2011 en april 2020 het algoritme PredPol gebruikt. Dit is een van de vele algoritmes die op basis van data van eerdere criminaliteit voorspellingen doen over waar het waarschijnlijk is dat criminaliteit gaat plaatsvinden. Ook kunnen sommige van deze algoritmes uitspraken doen over welke personen een hogere kans hebben om een misdaad te begaan (Miller, 2020). Een dergelijk algoritme wordt ook wel een 'predictive policing' (vanaf nu PP) algoritme genoemd. Tijdens het gebruik van PredPol is er een daling van 20% geconstateerd in de hoeveelheid misdaden tussen januari 2013 en januari 2014 bij de Los Angeles Police Department's Foothill Division. Er was zelfs een dag zonder enige criminaliteit op 13 februari 2014 (PredPol, z.d.). Een PP algoritme lijkt een goed instrument om criminaliteit te verminderen, omdat criminaliteit voorkomen kan worden als de plaats waar deze gaat plaatsvinden kan worden voorspeld.

Toch wordt het algoritme PredPol niet meer gebruikt in Los Angeles. Als formele reden wordt de slechte financiële situatie in de stad door Covid-19 genoemd, maar het is waarschijnlijk dat de protesten van activisten een belangrijke rol hebben gespeeld. Zo is er een petitie getekend door 450 academici uit de hele Verenigde Staten waarin de ethische aspecten van het algoritme worden betwijfeld (Haskins, 2020). PP algoritmes zijn bijvoorbeeld zo ontworpen dat ze stigmatiseren, of discrimineren op basis van ras of geslacht. Uit het onderzoek van O'Donnell (2019) blijkt dat er 5.1 keer zoveel zwarte Amerikanen worden opgesloten in gevangenissen als witte Amerikanen. Doordat deze data wordt meegenomen in de berekeningen van de PP algoritmes, kan het algoritme bevooroordeeld zijn op een negatieve wijze, waardoor er racisme kan optreden.

Een ander voorbeeld is dat criminaliteit op bepaalde locaties wordt voorspeld zonder dat daar directe aanleiding voor is, met als gevolg dat er op die locatie ook meer politie wordt ingeschakeld. Hierdoor volgen er meer arrestaties omdat er meer toezicht is, en niet per se omdat daar meer criminaliteit is dan op andere plekken (Lum & Isaac, 2016). Dit resulteert in het probleem dat de politie zich focust op slechts een bepaald gebied, wat niet in lijn is met de misdaadcijfers (Selbst, 2017), wat ook weer kan uitmonden in bevooroordeelde uitkomsten.

Dit zijn een aantal problemen rondom PP algoritmes die ervoor zorgen dat er bezwaren kleven aan het gebruik ervan in de samenleving. Deze problemen hebben zowel statistische als ethische aspecten. De statistische aspecten hebben betrekking op de betrouwbaarheid van de PP algoritmes, dus in hoeverre deze juiste resultaten genereren. Wegens de omvang van dit onderzoek wordt er enkel aandacht gegeven aan de morele aspecten. Dit heeft geleid tot de volgende onderzoeksvraag:

Wat zijn de morele overwegingen om een algoritme, dat wordt gebruikt voor het voorspellen van criminaliteit, verantwoord in te kunnen zetten in de praktijk?

Om antwoord te vinden op de hoofdvraag, zijn de volgende deelvragen opgesteld. Deze worden per sectie apart behandeld.

1. Worden algoritmes in het algemeen vertrouwd?

Als mensen helemaal geen vertrouwen hebben in algoritmes, zouden deze niet in de praktijk kunnen worden toegepast. Met deze vraag wordt onderzocht in hoeverre algoritmes worden vertrouwd, om vast te stellen of dit ook voor PP algoritmes geldt.

2. Wat voor soorten 'predictive policing' algoritmes zijn er?

Er zijn verschillende soorten PP algoritmes die voorspellingen doen over verschillende aspecten, bijvoorbeeld een algoritme dat uitspraken doet over de locatie van een potentiële toekomstige misdaad, of over welke individuen er een grotere kans hebben om een misdaad te plegen. Dit zal in sectie 2 verder worden toegelicht. Het is van belang om onderscheid te maken in de

verschillende soorten PP algoritmes, omdat elke soort andere problemen en vraagstukken met zich meebrengt. Deze zullen apart behandeld worden.

3. Wat zijn de positieve en negatieve aspecten van de 'predictive policing' algoritmes die al in de praktijk zijn toegepast?

Naast het PredPol algoritme dat in Los Angeles werd gebruikt, zijn dit en ook andere algoritmes al eerder toegepast in de praktijk, en worden in sommige steden nog steeds gebruikt. Het is nodig de positieve en negatieve aspecten van de PP algoritmes in de praktijk in kaart te brengen, zodat er kan worden gekeken naar hoe een PP algoritme eruit zou moeten zien om moreel verantwoord ingezet te kunnen worden in de praktijk.

4. Wat zijn de morele gevolgen van feedback loops bij 'predictive policing' algoritmes?

Feedback loops zorgen ervoor dat waar meer criminaliteit is voorspeld door het PP algoritme, er op die locatie ook meer politie wordt ingeschakeld. Hierdoor volgen er meer arrestaties omdat er meer toezicht is, en niet per se omdat daar meer criminaliteit is dan op andere plekken (Lum & Isaac, 2016). Dit brengt een aantal morele gevolgen met zich mee, wat in deze sectie behandeld zal worden.

5. Aan welke eisen zou een 'predictive policing' algoritme moeten voldoen om op een juiste en veilige manier toegepast te kunnen worden in de maatschappij?

Het doel van deze vraag is om met potentiële oplossingen te komen voor de vraagstukken die bij de vorige deelvragen naar boven zijn gekomen.

Deze vragen zal ik onderzoeken aan de hand van een literatuuronderzoek. Er zijn al meerdere onderzoeken gedaan naar PP algoritmes. Zo heeft Ferguson (2017) het boek 'The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement' geschreven. Hij gaat hier vooral in op hoe de aanpak van de politie is veranderd dankzij de PP algoritmes. De cijfers en feiten die hij gebruikt zijn interessant voor dit onderzoek. Het verschil is echter dat hij ingaat op de gevolgen die PP algoritmes hebben voor de politie, terwijl dit onderzoek meer focus legt op de morele problemen in de samenleving. Ook het onderzoek van Stop LAPD Spying Coalition (2018) geeft nuttige informatie over de argumenten van de tegenstanders van PP algoritmes. Er zijn veel verklaringen te lezen over inwoners van Los Angeles die nare ervaringen hebben met de politie als gevolg van het gebruik van de uitkomsten van PP algoritmes. De meeste onderzoeken naar PP algoritmes zijn uitgevoerd in Amerika, omdat daar ook het meest gebruik wordt gemaakt van PP algoritmes. Om deze reden wordt er in mijn onderzoek vooral data gebruikt van Amerikaanse steden en wijken.

Het doel van deze scriptie is om de morele problemen in kaart te brengen, in te gaan op de morele aspecten die spelen rondom PP algoritmes en om te verkennen hoe PP algoritmes op een juiste en veilige manier zouden kunnen worden ingezet.

Dit zal ik gaan doen vanuit een interdisciplinaire invalshoek: ethische aspecten worden verbonden met het onderzoeksgebied van de kunstmatige intelligentie. Het is van belang dat de morele aspecten goed worden afgewogen voordat een nieuwe technologie, zoals een PP algoritme, kan worden geïntroduceerd in de maatschappij. Dit is belangrijk voor zowel de bescherming van privacy en mensenrechten, als de veiligheid. Met mijn onderzoek wil ik inzicht geven in hoe we met deze aspecten om kunnen gaan bij PP algoritmes.

## Sectie 1: Algemeen vertrouwen algoritmes

Voordat we gaan kijken naar de morele overwegingen om een PP algoritme verantwoord in te kunnen zetten in de praktijk, is het nodig om te onderzoeken of algoritmes in het algemeen wel worden vertrouwd. In het geval van de PP algoritmes gaat het om vertrouwen van verschillende partijen. Ten eerste moeten de agenten de PP algoritmes vertrouwen. Dit is belangrijk omdat zij de informatie afkomstig van de PP algoritmes ook daadwerkelijk moeten gebruiken om bijvoorbeeld naar een verdachte locatie of persoon te gaan. Wanneer dit vertrouwen er niet is en er meer op eigen intuïtie wordt gehandeld, heeft het geen zin om PP algoritmes in te zetten. Ten tweede is het vertrouwen vanuit de samenleving belangrijk, aangezien een overheidsinstantie handelt vanuit de gegevens van een algoritme. Met name de rechtspraak moet kunnen vertrouwen op PP algoritmes, zodat deze als hulpmiddel kunnen worden ingezet voor het bepalen of een persoon wel of niet schuldig is. Een voorbeeld hiervan is dat wanneer PP algoritmes in de praktijk worden gebruikt, er vaker huiszoekingen plaats zouden kunnen vinden. Dit is alleen toegestaan met een huiszoekingsbevel, dat enkel op basis van gegronnd bewijs kan worden gegeven (Maxius.nl, 2020). Het vertrouwen in een PP algoritme zou dan zo goed moeten zijn, dat de uitkomsten van het algoritme ook onder gegronnd bewijs valt. Er zijn verschillende onderzoeken gedaan naar het vertrouwen in algoritmes. De conclusies van die onderzoeken zijn niet eenduidig.

Uit het onderzoek van Logg et al. (2019) wordt de conclusie getrokken dat mensen een inschatting van een statistisch algoritme verkiezen boven een menselijke inschatting. Ze noemen dit verschijnsel 'algorithm appreciation'. Logg et al. (2019) hebben onderzoek gedaan naar dit fenomeen door middel van zes verschillende experimenten, waar de proefpersonen een eigen inschatting moesten maken in verschillende situaties. Zo werden de proefpersonen gevraagd om een inschatting van het gewicht te maken van een persoon op een foto. Vervolgens kreeg de ene helft advies van een mens en de andere helft van een statistisch model. Uit het onderzoek bleek dat de eerste inschatting vaker werd aangepast na het advies van het algoritme, dan na het advies van een mens (Logg et al., 2019). Een mogelijke verklaring hiervan is dat mensen denken dat algoritmes objectievere en rationelere inschattingen kunnen maken dan mensen (Koll, 2020, p.57).

Haaks op deze conclusie staat die van Dietvorst et al. (2015). Zij beweren juist dat er juist sprake is van 'algorithm aversion': het verkiezen van de inschatting van een mens boven die van een statistisch model. Ze hebben onderzoek gedaan naar dit fenomeen door middel van vijf experimenten. Deze experimenten hadden een vergelijkbare opzet zoals bij het onderzoek van Logg et al. (2019). De proefpersonen moesten inschattingen maken van verschillende situaties, zoals bijvoorbeeld de volgorde van de vijftig Verenigde Staten op basis van de meeste passagiers die met het vliegtuig waren vertrokken. Een verschil met de experimentele opzet van Logg et al. (2019) is dat de proefpersonen eerst een oefenronde kregen, waarbij ze ook inzicht kregen in de inschattingen van het statistische model en die van een groep mensen. Het statistische model had in elke conditie een betere inschatting dan de mensen, maar gaf niet helemaal het exacte antwoord. De proefpersonen kozen vervolgens in elke conditie voor het menselijke advies (Dietvorst et al., 2015). Een mogelijke verklaring voor dit fenomeen is dat mensen verwachten dat algoritmes perfect werken. Wanneer zij in de proefronde hebben gezien dat deze ernaast kunnen zitten, is een groot deel van het vertrouwen weg. Dit zou kunnen zijn omdat mensen denken dat wanneer een algoritme fout zit, dit een systematische fout is, terwijl bij een menselijke fout het logisch gevonden wordt dat mensen niet perfect zijn en fouten maken. Ook worden menselijke fouten gezien als random, en wordt hier niet de conclusie getrokken dat dit per se de volgende keer weer gebeurt (Koll, 2020, p.18). Mensen zijn dus meer geneigd om een menselijke fout te vergeven, dan een fout van een algoritme.

Op basis van deze uitkomsten, lijkt het alsof mensen sceptisch zijn tegenover het gebruik van algoritmes. Een PP algoritme zou dus perfect moeten zijn voordat het in gebruik kan worden genomen. Dit is erg opmerkelijk, aangezien het duidelijk is dat mensen ook niet perfect zijn. Kennelijk wordt het de mens eerder vergeven een fout te maken dan een algoritme. Dit kan als

een probleem worden gezien, omdat er dus perfectie verwacht wordt van een algoritme voordat de uitkomsten worden vertrouwd. Dit is vooralsnog niet realistisch, aangezien PP algoritmes gebruik maken van kansberekeningen, en dus niet 100% garantie kunnen geven. Een oplossing voor dit probleem zou kunnen zijn om de agenten meer inspraak te geven in de manier waarop berekeningen van PP algoritmes worden uitgevoerd. Deze opvatting is gebaseerd op een ander onderzoek van Dietvorst et al. (2016), waaruit blijkt dat mensen sneller de uitkomst van een algoritme vertrouwen als ze de mogelijkheid hebben om het algoritme te sturen als ze denken dat deze een fout maakt. Het onderzoek is uitgevoerd met een algoritme dat per definitie al imperfect was en fouten maakte. Maar zo lang de proefpersonen ook maar een beetje in staat waren om het algoritme aan te passen, ging het vertrouwen in de uitkomsten aanzienlijk omhoog (Dietvorst et al., 2016).

We kunnen hieruit dus concluderen dat er voor beide kanten wat te zeggen valt. Algoritmes zijn juist wel te vertrouwen, omdat deze in staat zijn om objectievere en exactere inschattingen te maken dan mensen door hun vermogen om exacte kansberekeningen te maken. Mensen verwachten dat algoritmes perfect zijn, en daarom vertrouwen ze ook op ze. Het gevaar hierbij ligt dat wanneer een algoritme eenmaal imperfect blijkt te zijn, het vertrouwen sneller verdwijnt dan het vertrouwen in de mens. Dit effect is dan weer minder wanneer er enige vorm van controle is over de werking van het algoritme. Met betrekking tot de PP algoritmes, zou het dus het beste zijn als deze zodanig werken dat er de mogelijkheid is voor de gebruikers om ze te modificeren. Om inzicht te krijgen over hoe deze te modificeren zijn, is er ook transparantie over de werking van het algoritme nodig. Dit wordt verder besproken in sectie 5. Dit vergroot het vertrouwen in PP algoritmes, wat ervoor zorgt dat deze daadwerkelijk ook in de praktijk kunnen worden gebruikt.

## Sectie 2: Soorten predictive policing algoritmes

Eerder zagen we al dat er twee soorten PP algoritmes zijn: location-based PP algoritmes die uitspraken doen over de locatie waar een misdaad zou gaan plaatsvinden en person-based PP algoritmes die uitspraken doen over individuen. Er zijn ook verschillende soorten person-based PP algoritmes. Zo zijn er person-based PP algoritmes die uitspraken doen over zowel mensen die een hoge kans hebben om een misdaad te plegen, als over mensen die een hoog risico hebben om slachtoffer te worden van misdaad. De verschillende soorten PP algoritmes zullen in deze sectie verder toegelicht worden. Deze informatie is van belang voor het bestuderen van de morele overwegingen om een PP algoritme in te kunnen zetten, omdat elke soort andere problemen en vraagstukken met zich meebrengt. Het is nodig om eerst te begrijpen welke soorten er zijn, om de bijbehorende problemen en vraagstukken te begrijpen. De onderstaande informatie komt hoofdzakelijk van Perry et al. (2013).

1. Een algoritme dat zich focust op het voorspellen van misdaden. Dit is een location-based PP algoritme: hiermee worden de meest waarschijnlijke locatie en tijd voorspeld waarop een misdaad gaat plaatsvinden. Deze voorspellingen worden gedaan op basis van gegevens over het voorkomen van eerdere misdaden. Hiermee worden berekeningen gedaan over hoe waarschijnlijk het is dat er in bepaalde gebieden misdaden gaan plaatsvinden. Door het toevoegen van gegevens over tijd en datum van de eerder misdaden, kunnen ook uitspraken worden gedaan over wanneer misdaden plaatsvinden.
2. Een person-based algoritme dat een voorspelling doet over potentiële daders. Hierbij worden personen met een hoog risico op het in de toekomst overtreden van de wet geïdentificeerd. Dit wordt gedaan op basis van de voorgeschiedenis: mensen met psychische problemen, ex-gevangenen en mensen met een strafblad hebben bijvoorbeeld een hogere kans om (weer) de wet te overtreden, en zullen dus eerder als een risico worden beschouwd. Hierbij wordt soms zelfs door sommige ontwikkelaars gebruik gemaakt van data uit de sociale media, zoals Facebook of Twitter (Selbst, 2017).
3. Een algoritme dat zich concentreert op het schetsen van de identiteit van dader. Dit is een person-based PP algoritme: hiermee worden uitspraken gedaan over het profiel van de dader, wat helpt om gericht op zoek te gaan. Dit wordt gedaan door de gegevens van de misdaad te koppelen aan daders van eerdere misdaden.
4. Een person-based algoritme dat zich focust op het voorspellen van de slachtoffers van misdaad. Hierbij worden personen of groepen die een hoge kans hebben om slachtoffer te zijn van misdaad geïdentificeerd. Dit wordt op dezelfde methode gedaan als het voorspellen van daders. Daarnaast wordt er ook gebruik gemaakt van de locaties waar er misdaden worden voorspeld, aangezien mensen in hoge risicogebieden een grotere kans hebben om slachtoffer te worden (Perry et al., 2013, pp.36-42). Omdat deze soort nog maar weinig wordt gebruikt, wordt deze verder niet behandeld in mijn onderzoek.

Deze verschillende soorten PP algoritmes zijn al toegepast in de praktijk. De positieve en negatieve aspecten hiervan zijn in de volgende sectie te lezen.

## Sectie 3: Positieve en negatieve aspecten

Zoals eerder vermeld, zijn er zowel location- als person-based PP algoritmes al eerder in de praktijk toegepast. Voorbeelden hiervan zijn PredPol en LASER. Het is interessant om te kijken naar hoe deze algoritmes het in de praktijk hebben gedaan en een beeld te krijgen van de positieve en negatieve aspecten. Veel PP algoritmes worden niet meer in de praktijk gebruikt omdat men tegen problemen aan liep. Het is nodig om dieper in te gaan naar hoe deze problemen zich hebben gemanifesteerd in de praktijk, zodat er uiteindelijk kan worden gekeken naar hoe een PP algoritme eruit zou moeten zien om moreel verantwoord ingezet te kunnen worden in de praktijk.

Ten eerste zorgen de zogenaamde person-based PP algoritmes ervoor dat potentiële daders eerder in beeld komen. Met de informatie die een dergelijk algoritme levert kunnen hulpdiensten aan potentiële daders hulp bieden om zo te voorkomen dat zij de fout in gaan, wat dan weer als gevolg heeft dat de algemene criminaliteit daalt. Een voorbeeld hiervan is van toepassing in Kansas. In 2014 begon de lokale politie gebruik te maken van een person-based PP algoritme. Hieruit kwamen 884 namen die als een hoog risico werden gezien, dus mensen die een verhoogde kans hebben om in de toekomst de wet te overtreden. Overheidsinstanties gingen persoonlijk langs bij die burgers om diensten aan te bieden om hun leven te beteren. Ze werden dus gewaarschuwd dat ze in de gaten werden gehouden, maar er werd ook hulp aangeboden om te voorkomen dat zij (weer) de fout in zouden gaan. Er was een daling van de criminaliteit: eind 2014 was er in Kansas een daling te zien in criminaliteit van 26,5% ten opzichte van de vorige drie jaren waarin geen algoritmes werden gebruikt (Ferguson, 2017, p.37).

Ten tweede kunnen er met behulp van location-based PP algoritmes gerichtere surveillances worden uitgevoerd. In wijken waar vaker misdaden worden gepleegd, kunnen meer agenten worden ingezet om uiteindelijk een veiligere omgeving te creëren met minder criminaliteit. Een voorbeeld van een location-based PP algoritme is PredPol. Dit algoritme gebruikt data van misdrijven van de afgelopen tien jaar. Met deze informatie worden er uitspraken gedaan over waar en wanneer het het meest waarschijnlijk is dat de komende twaalf uur misdaden gaan plaatsvinden. PredPol markeert gebieden met een hoog risico op criminaliteit als een PredPol hotspot. Deze hotspots zijn ongeveer 150 vierkante meter. Twee keer per dag krijgen de agenten in de desbetreffende gebieden een rapport over de PredPol hotspots van de dag. Wanneer agenten niets te doen hebben, worden ze geacht om naar PredPol hotspots te gaan voor extra toezicht (Smith, 2019, p.25). Bij de Santa Cruz, CA Police Department is er gebruik gemaakt van PredPol. Het resultaat was dat er een daling te zien was in de algemene criminaliteit van 25% in juni 2013 en 29% in juli 2013 vergeleken met diezelfde maanden in 2012 (PredPol, z.d.).

Naast de bovengenoemde positieve resultaten, kleven er ook nadelen aan het gebruik van de PP algoritmes.

Om te beginnen kunnen de PP algoritmes stigmatisering veroorzaken. Dit is te zien in zowel location- als person-based PP algoritmes. Het wonen in een bepaalde wijk of het hebben van een bepaalde afkomst kunnen redenen zijn om meer in de gaten gehouden te worden. Een voorbeeld hiervan is bijvoorbeeld bij PredPol. Wanneer je woont in een PredPol hotspot, kun je al snel als verdacht worden beschouwd. Schuldig tot het tegendeel is bewezen lijkt de nieuwe standaard te zijn. Een ander voorbeeld van stigmatisering is het Los Angeles Strategic Extraction and Restoration Program, oftewel LASER (Smith, 2019, p.3). Hierbij wordt gebruikt gemaakt van een location- en een person-based PP algoritme. Ik ga hier alleen in op het person-based algoritme. Dit PP algoritme werd gebruikt in Los Angeles van 2011 tot 2019, met als doel om geweld en misdaad met vuurwapens te verminderen. Dit algoritme werkt met de identificatie van zogeheten 'chronic offenders': personen die in het verleden verantwoordelijk zijn geweest voor (wapen)geweld. Deze 'chronic offenders' worden in de gaten gehouden met behulp van een puntensysteem, dat te zien is in tabel 1 (Smith, 2019, p.6).



Aantal punten	Actie
5	Lid zijn van een gang
5	Voorwaardelijk vrij zijn uit de gevangenis
5	Eerdere arrestatie met een handgeweer
5	Strafblad met gewelddadige misdaden
1	In contact geweest met de politie in de laatste twee jaar

Tabel 1: Het puntensysteem van LASER (Smith, 2019, p.6).

De 'chronic offenders' met het hoogste aantal punten worden extra in de gaten gehouden, en waar nodig verwijderd uit de wijk bijvoorbeeld door arrestatie. Er is hier sprake van stigmatisering, omdat er automatisch een etiket op iemands hoofd wordt geplaatst, enkel op basis van het verleden. Een persoon kan bijvoorbeeld nog niks fout gedaan hebben na uit de gevangenis te zijn gekomen, maar dan toch al een verdachte zijn (Stop LAPD Spying Coalition, 2018). Het geeft de indruk dat voor een 'chronic offender' geldt dat hij of zij schuldig is totdat diens onschuld is bewezen, in plaats van andersom. Het gebruik van LASER moest in 2019 worden stopgezet door vele protesten en uiteindelijk een rechtszaak, waarin de politie gedwongen werd om documenten vrij te geven. Hieruit bleek dat bijna de helft van de 'chronic offenders' nul of geen arrestaties hadden voor gewelddadige misdaden en dat bijna 10% geen contact heeft gehad met de politie (Baek & Mooney, 2020). Ook bleek dat honderd mensen die op de lijst van 'chronic offenders' staan, nul punten hadden (Stop LAPD Spying Coalition, 2018). Dit laat zien dat het als 'chronic offender' worden beschouwd niet in alle gevallen even gegrond is.

Ten tweede is er sprake van de schending van privacy. Wanneer iemand eenmaal als een 'chronic offender' wordt beschouwd, wordt het softwareprogramma Palentir gebruikt. Dit is een algoritme dat meerdere databases doorzoekt om de activiteiten van een persoon over de afgelopen twee jaar vast te stellen. Dit is nodig voor het puntensysteem van het person-based PP algoritme van LASER. Palentir focust zich op het criminele verleden van een persoon, met name op de soort van de misdaden. Ook maakt Palentir gebruik van een 'Automated License Plate Reader', waarmee er onderzoek wordt gedaan naar het sociale netwerk van een persoon (Stop LAPD Spying Coalition, 2018, p10). De privacy van de 'chronic offenders' lijkt dus ver te zoeken. Verder verrichten de agenten (een van) de volgende acties met de twaalf 'chronic offenders' met de hoogste punten (Smith, 2019, p.7):

- Het sturen van een brief met een waarschuwing
- Het uitvoeren van gerechtelijke controles, zoals fouilleringen of huiszoekingen
- Het controleren op de naleving van de voorwaardelijke vrijlating
- Het op de deur kloppen om advies te geven over beschikbare programma's en diensten om het risico op de herhaling van een misdrijf te verkleinen

Met name het tweede punt lijkt in tegenstrijd te zijn met het Vierde amendement van de Amerikaanse grondwet, dat verbiedt namelijk onredelijke huiszoekingen en inbeslagnemingen (National Constitution Center, 2021). Het recht op privacy betekent namelijk ook dat mensen het recht hebben om met rust gelaten te worden en afgeschermd te worden tegen bespieding. Dit werd in 1879 al beschreven door de Amerikaan Thomas Cooley: "the right to be let alone" (Warren & Brandeis, 1890, p.193). Er is een spanning tussen privacy en andere belangen, zoals garanderen van de veiligheid voor de burgers en het voorkomen van criminaliteit of terrorisme. Het is echter de vraag of het redelijk is om zonder directe aanleiding, alleen op basis van een puntensysteem, een dergelijke zwaarwegende inbreuk op de privacy te doen door middel van een huiszoeking.

Het derde grote probleem bij een person-based PP algoritme zoals dat van LASER is dat het lastig is om als 'chronic offender' van de lijst af te komen. Er zijn geen minpunten te behalen, dus wanneer je eenmaal op de lijst staat, kom je er heel lastig weer vanaf. Dit wordt bovendien ook nog eens bemoeilijkt door een 'feedback loop' in het algoritme: wanneer bepaalde mensen zo streng in de gaten wordt gehouden door de politie dankzij de inschattingen van Palentir, worden

deze mensen ook eerder betrapt op misdaden dan andere misdadigers. Zo ligt de focus dus voornamelijk op de mensen die al in het systeem zitten, in plaats van dat er naar het algemene geheel wordt gekeken. Dit probleem komt ook voor bij location-based PP algoritmes: hoe vaker er in een bepaalde hotspot wordt gesurveilleerd, hoe meer arrestaties daar worden gedaan en hoe hoger de kans is dat dat weer als hotspot wordt gezien. Dit onderwerp wordt verder uitgewerkt in sectie 4.

Ten slotte blijkt uit onderzoek van de LA Times dat de politie vaker Afro-Amerikanen en latino's aanhouden dan witte Amerikanen, terwijl bij witte Amerikanen het vaakst drugs, wapens of andere strafbare voorwerpen worden gevonden. Uit het artikel van Poston & Chang (2019) bleek dat in Los Angeles in de eerste tien maanden van 2019 de auto van 24% van Afro-Amerikaanse automobilisten is doorzocht, terwijl bij latino's dit percentage 16% is en bij witte Amerikanen zelfs slechts 5%. Toch werden er bij 20% van de witte Amerikanen iets van drugs, wapens of ander smokkelwaar in de auto gevonden. Bij de Afro-Amerikanen was dit maar bij 17% en bij latino's 16% (Poston & Chang, 2019). Dit betekent dat de kans vier keer zo groot is voor een Afro-Amerikaan en voor een Latino drie keer zo groot om gestopt te worden door de politie dan een witte Amerikaan. Dit wordt ook wel etnisch profileren genoemd: "het gebruik door de politie van criteria of overwegingen omtrent ras, huidskleur, etniciteit, nationaliteit, taal en religie bij opsporing en rechtshandhaving – zowel op operationeel als organisatorisch niveau – terwijl daarvoor geen objectieve rechtvaardiging bestaat" (Amnesty International, 2020). Door het vaker aanhouden zullen Afro-Amerikanen en latino's dus vaker in de criminaliteitsdatabases voorkomen, waardoor de bias alleen maar versterkt wordt. Het gebruik van algoritmes versterkt een structureel institutioneel racisme (Poston & Chang, 2019). Uit de in 2018 vrijgegeven documenten van LASER bleek dan ook dat 84% van de 'chronic offenders' bestond uit latino's en Afro-Amerikaanse mensen (Baek & Mooney, 2020). De kwestie van etnisch profileren is echter een heel groot onderwerp, wat te groot is om volledig behandeld te worden in deze scriptie.

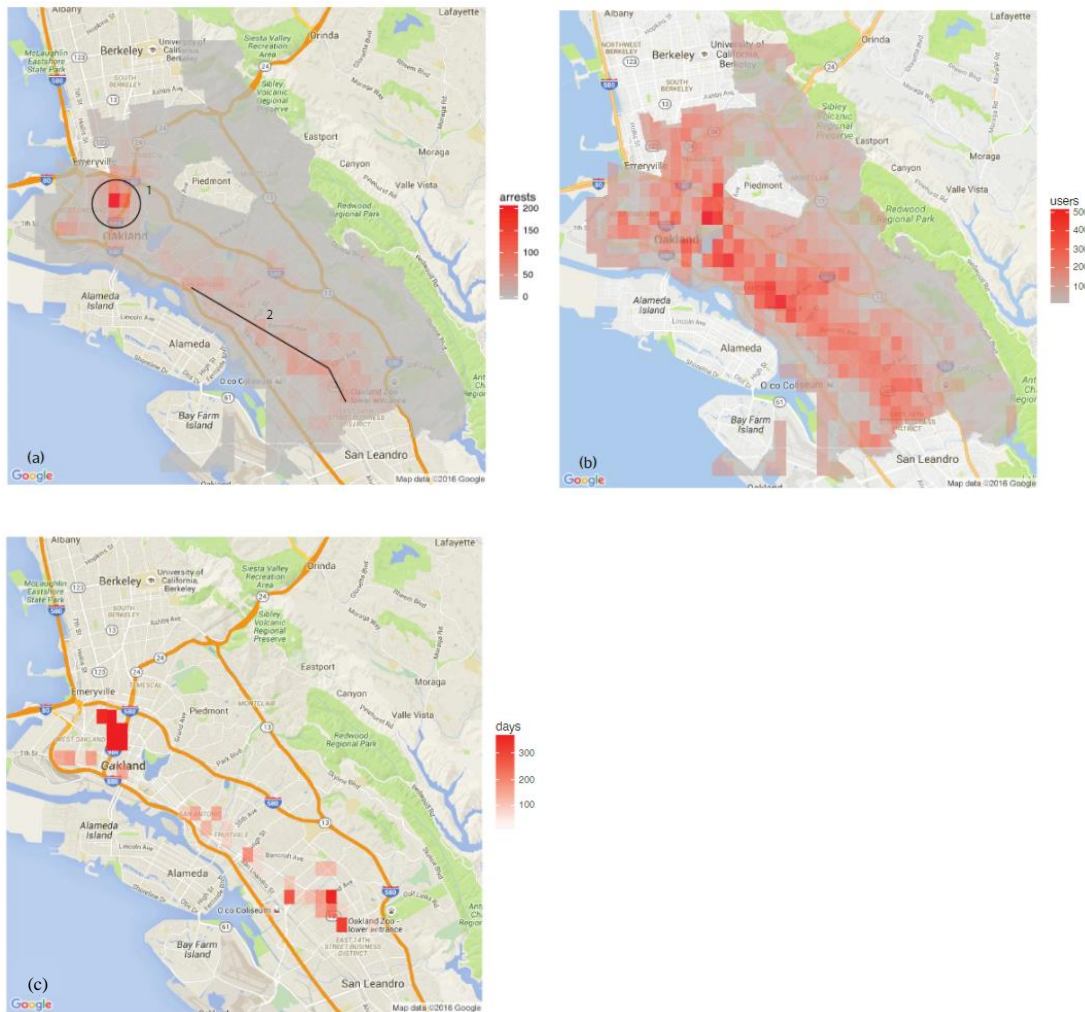
Aan de ene kant kleven er dus nadelen aan het gebruik van PP algoritmes, met als eerste punt stigmatisering: het enkel wonen in een bepaalde wijk of het hebben van een bepaalde afkomst kunnen redenen zijn om meer in de gaten gehouden te worden door de politie. Als tweede is er sprake van de schending van privacy door de 'chronic offenders' streng in de gaten te houden. Daarnaast bemoeilijken feedback loops het om van de lijst van 'chronic offenders' af te komen en wordt ook de nadruk van politietoezicht op bepaalde wijken versterkt. Tot slot worden door PP algoritmes het etnisch profileren bij de politie versterkt. Aan de andere kant hebben person-based PP algoritmes voor een sterke daling gezorgd in de criminaliteit, zo hebben ze dit in Kansas gedaan door middel van het aanbieden van hulp bij de mensen die een hoog risico hebben om het slechte pad op te gaan. Ook het location-based PP algoritme PredPol heeft voor een forse daling gezorgd in de criminaliteit in Los Angeles door vaker op de aangegeven hotspots te patrouilleren.

## Sectie 4: Morele gevolgen van feedback loops

In de vorige sectie zagen we dat feedback loops bij PP algoritmes een groot probleem zijn: uitkomsten van het algoritme zijn enkel afhankelijk van de data die al bekend zijn. Als die data eenzijdig zijn, levert dat eenzijdige uitkomsten, die vervolgens weer wordt opgenomen in de database. Daarmee wordt de eenzijdigheid versterkt. Wanneer er bijvoorbeeld de beslissing is genomen door een PP algoritme om in een bepaalde wijk te patrouilleren, dan komen daaruit enkel gegevens van eerdere misdaad uit die ene bepaalde wijk (Ensign et al., 2018, p.1). Door deze in te voeren, wordt de kans groot dat de uitkomst van het PP algoritme een volgende keer wederom leidt tot een patrouille in die wijk. Dit is niet zozeer een fout in het algoritme zelf, maar van de data die het ontvangt. Toch heeft het effect van feedback loops die het algoritme genereert grote gevolgen.

Aan de ene kant zorgen feedback loops ervoor dat bepaalde gebieden worden oververtegenwoordigd. Zoals hierboven al is aangegeven, hebben feedback loops in location-based PP algoritmes het gevolg dat er onevenredig veel data van misdaad beschikbaar komen uit bepaalde wijken omdat ze hun oordelen baseren op eenzijdige data. Op basis van het algoritme zou er vervolgens meer politie moeten worden ingezet in dat gebied, hetgeen leidt tot meer arrestaties. Deze arrestaties vormen vervolgens weer input voor het PP algoritme, waarmee de cirkel rond is. Door dit mechanisme bevestigt en versterkt het algoritme vooroordelen. Hoewel de uitspraken van het PP algoritme niet representatief zijn, krijgen wijken waarover meer data bekend zijn extra aandacht van de politie, die er bij voorbaat van uitgaat dat 'er wel weer iets zal gebeuren'. Het hebben van te veel agenten in een wijk is niet alleen inefficiënt, maar heeft ook een averechts effect. Het heeft negatieve effecten op de mentale gezondheid van de inwoners van die wijk. Voornamelijk mannen ervaren meer gevoel van nervositeit, waardeloosheid en ernstige psychologische angst (Sewell et al., 2016). Ook in het onderzoek van Van de Veer et al. (2012) komt naar voren dat meer politie op straat direct zorgt voor een onveilig gevoel bij de bewoners. Dit kan te verklaren zijn doordat het zien van politie automatisch leidt tot mentale associaties met bijvoorbeeld misdaad of geweld (Van de Veer et al., 2012). Een andere verklaring is dat het zien van politie het gevoel geeft dat het daadwerkelijk nodig is dat er veel politie in de wijk is en dat er dus veel misdaad is. Wijken die sterk onder toezicht worden gehouden krijgen dus een slechte reputatie, dat ervoor zorgt dat wijken impopulair worden, met mogelijk leegstand en verloedering tot gevolg.

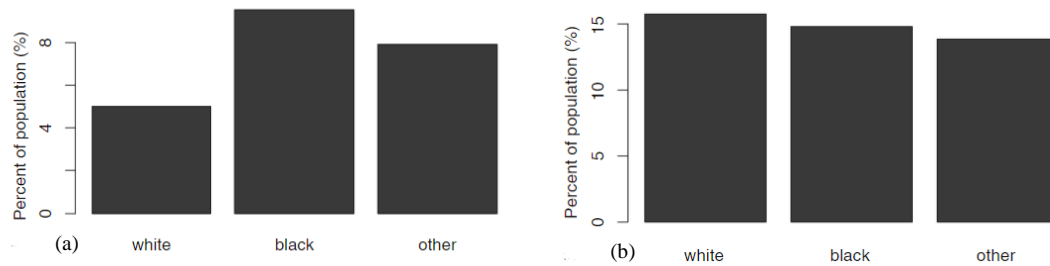
Deze oververtegenwoordiging van bepaalde wijken leidt automatisch tot ondervertegenwoordiging van andere wijken. Dit heeft als gevolg dat er minder vaak wordt opgetreden en de criminaliteit stijgt in de ondervertegenwoordigde wijken. Het onderzoek van Lum & Isaac (2016) geeft hier inzicht in. Zij hebben de aantallen drugsarrestaties (figuur 1(a)), de geschatte hoeveelheid drugsgebruikers van de National Survey on Drug Use and Health uit 2011 (figuur 1(b)) en de door het location-based algoritme PredPol gegenereerde hotspots (figuur 1(c)) met elkaar vergeleken (Lum & Isaac, 2016, pp.17-18).



Figuur 1: (a) Het aantal drugsarrestaties. (b) Geschatte hoeveelheid drugsgebruikers. (c) Aantal dagen in 2011 waarop PredPol een bepaald gebied heeft gemarkeerd als hotspot. (Lum & Isaac, 2016, pp.17-18).

Het in bezit hebben of gebruiken van drugs is in de Verenigde Staten verboden. Het bezitten en gebruiken van drugs kan leiden tot een celstraf van hoogstens een jaar en/of een boete van maximaal 1.000\$ (Government Publishing Office, 1984). Figuur 1(a) en figuur 1(c) laten zien dat er een groot verband is tussen de door PredPol geïdentificeerde hotspots en het aantal arrestaties, terwijl er een grotere spreiding van arrestaties door de stad heen een grotere spreiding in het aantal arrestaties zou rechtvaardigen.

Feedback loops bij person-based PP algoritmes werken op dezelfde manier als location-based PP algoritmes. Arrestaties of aanrakingen met de politie bevestigen de data van eerdere arrestaties, waardoor de nadruk enkel op een bepaald daderprofiel wordt gelegd. In het onderzoek van Lum en Isaac (2016) wordt het aantal drugsarrestaties (figuur 2(a)) en het aantal daadwerkelijke druggebruikers (figuur 2(b)) in Oakland met elkaar vergeleken. Deze laatste informatie is gebaseerd op de gegevens uit 2011 van National Survey on Drug Use and Health. Zij hebben indeling gemaakt op basis van ras, dat in het onderstaande figuur te zien is.



Figuur 2: (a) Percentage van de bevolking van drugsarrestaties door de politie. (b) Percentage van de bevolking van het daadwerkelijke drugsgebruik. (Lum & Isaac, 2016, p.18).

In figuur 2 is te zien hoe de politie bijna twee keer zoveel drugsarrestaties heeft verricht op zwarte mensen als op witte, terwijl er meer witte mensen zijn die drugs gebruiken (Lum & Isaac, 2016). Er is geen aanleiding om te veronderstellen dat er bij witte druggebruikers minder sprake is van criminaliteit dan bij de gebruikers met een donkere huidskleur. Een rijke witte jongen van 25 kan ook een drugsdealer zijn, maar past minder in het profiel en wordt daardoor eerder over het hoofd gezien (Ferguson, 2017). Maar het algoritme slaat deze eenzijdige data op en vormt de basis voor toekomstige besluiten van de politie. De eenzijdige data bevestigen en versterken dus ook de vooroordelen van de politie van meer criminaliteit bij de zwarte bevolking.

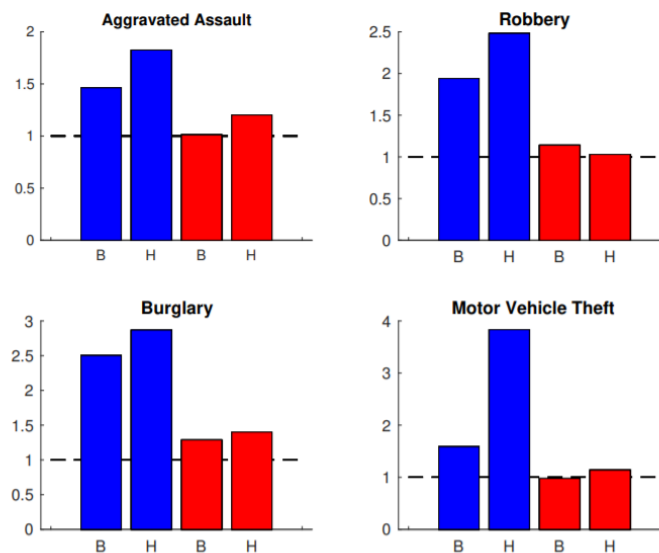
Een ander gevolg van feedback loops bij PP algoritmes is dat niet alle soorten misdaad worden meegenomen. Zo zijn de meeste PP algoritmes gericht op (wapen)geweld en inbraken. De data voor dergelijke misdaden neemt toe, terwijl er nauwelijks data over bijvoorbeeld witteboordencriminaliteit door het algoritme gebruikt kunnen worden. De feedback loops zorgen ervoor dat dit mechanisme dit versterkt. Bijkomend probleem is dat het lastiger is om onderzoek te doen naar witteboordencriminaliteit, omdat het profiel van de daders hiervan minder makkelijk te schetsen is. Het is echter wel belangrijk dat ook witteboordencriminaliteit wordt aangepakt door PP algoritmes, omdat dit een zeer vaak voorkomende en moreel verwerpelijke vorm van criminaliteit is. Uit onderzoek blijkt dat 49% van de bedrijven en 25% van de huishoudens slachtoffer zijn geweest van witteboordencriminaliteit (Richardson et al., 2019, p.41). Person-based PP algoritmes hebben eenzelfde blinde vlek. Zoals we eerder al zagen worden vaak dezelfde 'types' als input genomen, waardoor er veel potentiële misdadigers niet in beeld komen.

Samenvattend heeft het probleem van feedback loops bij PP algoritmes grote morele gevolgen. Door de eenzijdige focus op bepaalde wijken, blijven andere wijken buiten beeld, terwijl daar geen objectieve redenen voor zijn. Dit kan zorgen voor een slechtere mentale gezondheid, discriminatie en de verloedering van bepaalde wijken, terwijl de criminaliteit in andere wijken vrij spel heeft en het onnodig lang kan duren voordat dit duidelijk wordt. De uitkomsten van PP algoritmes veroorzaken en bevestigen vooroordelen, hetgeen gevolgen kan hebben voor de manier van optreden door de politie. Dit raakt aan het fundamentele principe dat alle mensen recht hebben op gelijke behandeling. In sectie 5 zal ik verder ingaan op mogelijke oplossingen voor het probleem van feedback loops.

## Sectie 5: Eisen voor een veilig gebruik van 'predictive policing' algoritmes

Nu we de problemen van de PP algoritmes in kaart hebben gebracht en weten wat er beter kan, kunnen we kijken aan welke eisen PP algoritmes zouden moeten voldoen om verantwoord ingezet te kunnen worden in de praktijk. We beschrijven hier drie belangrijke eisen: het voorkomen van bevooroordeelde data door feedbackloops, transparantie van het algoritme voor zowel gebruiker als dader en goed geregelde privacy.

Ten eerste moeten de algoritmes vrij zijn van feedback loops. Hiervoor zagen we het probleem van de feedback loops en de raciale bias die zij tot gevolg hebben. In hun onderzoek hebben Mohler et al. (2018) geprobeerd om dit probleem op te lossen. Dit hebben ze gedaan door middel van een functie met 'penalized likelihood-methods' en 'demographic parity'. 'Demographic parity' zorgt voor een eerlijke verhouding binnen een statistische berekening. Ik zal dit nader uitleggen door middel van een voorbeeld. Stel er is een bepaalde opleiding die dezelfde hoeveelheid mensen uit de stad als mensen van het platteland wil aannemen. Het toelatingsproces verloopt door middel van een test, waarbij er in totaal honderd mensen mogen worden toegelaten. Als we aannemen dat mensen uit steden gemiddeld slimmer zijn, zorgt dit voor het probleem dat er meer mensen uit de stad worden toegelaten. Om dit effect te voorkomen wordt de vereiste score voor mensen van het platteland tegenover de mensen uit de stad aangepast, zodat er uit beide bevolkingsgroepen vijftig man kan worden toegelaten (Machine Learning Glossary: Fairness, z.d.). In het geval van PP algoritmes hebben Mohler et al. (2018) hun berekeningen zodanig aangepast dat er in verhouding van elke bevolkingsgroep evenveel misdaden worden meegenomen. In 2019 waren er ongeveer 328 miljoen Amerikaanse inwoners, waarvan er 60 miljoen Latinos en 268 miljoen niet-latino's waren (United States Census Bureau, 2019). Stel dat latino's, de kleinere bevolkingsgroep, meer misdaden hebben gepleegd, zouden volgens de 'penalized likelihood-methods' van Mohler et al. (2018) deze misdaden minder zwaar wegen dan die van de niet-latino's. Door deze methode toe te passen worden de feedback loops gecorrigeerd en wordt de raciale bias geneutraliseerd, dat te zien is in figuur 3.



Figuur 3: Patrouilles per Latino of zwart persoon vergeleken met witte mensen (gestippelde lijn) per soorten misdaad. Rood is met 'penalized likelihood-methods' en blauw is zonder (Mohler et al., 2018, p.4)

De gestippelde lijn, die te zien is in alle vier de grafieken, geeft het aantal patrouilles weer voor een wit persoon. Dit is vervolgens vergeleken met de zwarte en Latino populatie. In het rood is de functie van Mohler et al. (2018) gebruikt met 'penalized likelihood-methods', en bij de blauwe is deze functie niet gebruikt. Het is aan de blauwe staven duidelijk te zien dat er zonder 'penalized

likelihood-methods' voor deze misdaden meer politiepatrouilles worden gehouden bij zwarte mensen of latino's dan bij witte mensen. Bijvoorbeeld bij "Motor Vehicle Theft", dus het stelen van motors, worden latino's zelfs vier keer zo vaak gecontroleerd dan witte mensen. Bij de functie die gebruik maakt van 'penalized likelihood-methods' is dit verschil veel kleiner, namelijk ongeveer 0.1 keer. Het gebruiken van functies met 'penalized likelihood-methods' zou dus een goede manier zijn om raciale bias bij PP algoritmes uit te sluiten.

Het nadeel is echter dat deze toepassing van Mohler et al. (2018) minder nauwkeurig is dan de bekende methodes waarbij de uitkomsten alleen gebaseerd is op data uit de wijk en er dus wel feedback loops optreden. Het location-based PP algoritme PredPol markeert bijvoorbeeld locaties als hotspots van 150 vierkante meter, terwijl de toepassing van Mohler et al. (2018) 300 vierkante meter hotspots genereert. Dit is een vier keer zo groot gebied, en dus is hun manier minder nauwkeurig dan PredPol.

Een tweede eis waar PP algoritmes aan moeten voldoen is dat ze transparant moeten zijn. Inspecteur-generaal Mark E. Smith (2018) heeft de raad van politiecommissarissen in Los Angeles geadviseerd over de PP algoritmes die toentertijd in gebruik waren, waarbij hij voornamelijk veel aanbevelingen had over het in sectie 3 besproken person-based PP algoritme LASER. Zijn aanbevelingen gingen voornamelijk over dat er meer sprake moest zijn van transparantie. Zo viel hem op dat er duidelijkere communicatie moest zijn naar de personen die als een hoog risico worden gezien. Daarbij noemde Smith ook dat het moet worden vermeld wanneer de personen met een hoog risico weer van de lijst af zijn en niet meer als 'chronic offenders' worden gezien. Dit kan voor hen een stimulans zijn om hun leven te beteren en op het rechte pad te blijven, met als gevolg dat de criminaliteit ook zal doen dalen (Smith, 2019, p.32). Transparantie is ook vanuit wettelijk en ethisch perspectief van groot belang.

Vanuit ethisch oogpunt is het belangrijk dat de gebruikers goed weten hoe het algoritme werkt en welke data er wordt gebruikt. Het moet zeer helder zijn dat algoritmes correct werken en dat zij onbetwistbare resultaten geven, aangezien zij zwaarwegende uitspraken doen. PP algoritmes baseren hun uitspraken op kansberekeningen, en in sectie 4 hebben we gezien dat de data die ze gebruiken ook niet altijd gebalanceerd zijn. Ook kunnen mensen zwaar benadeeld worden bij een foutieve uitspraak van een PP algoritme. Het is daarom belangrijk dat de mensen die de PP algoritmes gebruiken weten hoe deze werken, zodat ze volledige verantwoordelijkheid kunnen nemen voor de gevolgen van de uitspraken (Walmsley, 2020).

De Algemene verordening gegevensbescherming (AVG) stelt in de gehele Europese Unie de regels op voor de verwerking van persoonsgegevens (Autoriteit Persoonsgegevens, z.d.). Door de AVG is het wettelijk vastgesteld dat wanneer iemand is betrokken bij een besluit door een geautomatiseerd besluit op basis van zijn of haar persoonlijke informatie, dat diegene het recht heeft op een uitleg van hoe tot die beslissing is gekomen (Walmsley, 2020). Dit betekent dus dat de gebruikers van de PP algoritmes moeten begrijpen hoe deze werken, en dit kunnen communiceren naar anderen toe. Door transparantie is er ook meer controle op de werking van algoritmes. Zo kan er gekeken worden naar hoe de uitkomsten tot stand zijn gekomen en dus ook of deze gebaseerd zijn op racistische vooroordelen. Zo was er bij een ander algoritme, namelijk Google Translate, sprake van seksisme bij de vertalingen van talen waarbij de enkelvoudsvorm van de derde persoon genderneutraal is, zoals bij Fins of Turks: zinnen met het woord 'nurse' of 'cook' werd vertaald met 'zij', en 'doctor' en 'engineer' steeds met 'hij'. Deze vooroordelen werden uiteindelijk opgelost en uit het algoritme gehaald dankzij menselijke interventie. Omdat er dus binnen de werking van Google Translate transparantie aanwezig is, was het duidelijk waar het probleem vandaan kwam en kon het opgelost worden (Walmsley, 2020). Daarnaast is transparantie een belangrijke voorwaarde om het vertrouwen van het grote publiek in het algoritme te verkrijgen: waarom zou je de uitkomst van een algoritme vertrouwen als je niet begrijpt hoe het tot een bepaalde uitkomst komt (Chiao, 2019, p.127)?

Een eis die samenhangt met transparantie is dat ook de privacy gewaarborgd moet zijn. De wetten over privacy zijn per werelddeel verschillend. In Europa wordt de duidelijke en strikte privacywetgeving dus vastgesteld door de AVG (zie hierboven). Iedereen heeft recht op privacy en heeft de controle over zijn eigen data, ongeacht de bron of het gebruik van de data. De gegevens van een persoon mogen dus alleen gebruikt worden voor het specifieke doel waarvoor de gegevens zijn verzameld. In Amerika, waar de meeste PP algoritmes worden gebruikt, is de privacy voor verschillende markten door aparte wetten geregeld. Een voorbeeld hiervan is de Federal Communications Commission (FCC) die bepaalt welke data serviceproviders mogen verkopen of doorverkopen (Gryffroy, 2017). Daarnaast zijn er ook nog aparte wetten per staat. Er is steeds meer een spanning tussen privacy en veiligheid. In Amerika is sinds 11 september 2001 met het aannemen van de Patriot Act deze spanning tussen individuele privacy en terrorismebestrijding alleen maar sterker geworden en zijn de rechten van de mens ingeperkt. In Amerika is vooral veel te behalen op het gebied van transparantie over de werking van PP algoritmes, maar ook over welke persoonsgegevens er worden gebruikt en waarom. Ook moet het helder zijn welke gegevens de commerciële partij die het algoritme levert behoudt en wat het er mee doet. PredPol is een dergelijk zelfstandig bedrijf dat het location-based PP algoritme ontwikkelt dat bij verschillende politieafdelingen wordt gebruikt. Het moet duidelijk zijn wat dit bedrijf doet met de gegevens en wie de eigenaar van de data is (Gstrein et al., 2019). In een in hoog tempo veranderende wereld, met nieuwe ontwikkelingen op internet, de globalisering en de toename van steeds nieuwe vormen van terrorisme en criminaliteit, probeert de overheid snel maatregelen te nemen om de veiligheid van de burgers te garanderen. Het is van groot belang dat daarbij de ethische aspecten niet uit het oog verloren worden en de rechten van de individuele burger niet in gevaar komen. In een artikel in *The Verge* (Stroud, 2014) wordt pijnlijk duidelijk waartoe een grote inbreuk op privacy van een persoon kan leiden. Het beschrijft de casus van een man die door het algoritme werd geïdentificeerd als een risico, ook al had hij geen strafblad of ooit iets onwettigs gedaan.

Concluderend kunnen we stellen dat een van de eisen waar een PP algoritme aan zou moeten voldoen is dat etnisch profileren niet zou mogen worden versterkt maar in tegendeel juist uitsluit. Dit zou kunnen door bij PP algoritmes gebruik te maken van 'penalized likelihood-methods'. Een andere eis is dat PP algoritmes transparant moeten zijn. Dit is ten eerste belangrijk omdat de uitspraken van de algoritmes onderbouwd kunnen worden, ten tweede is transparantie van belang zodat het algoritme aan kan worden gepast om problemen op te lossen, en tot slot ook zodat het duidelijk is waarvoor persoonlijke gegevens gebruikt worden en het recht op privacy niet in het geding komt.



## Sectie 6: Conclusie

In dit onderzoek is gezocht naar een antwoord op de vraag: 'Wat zijn de morele overwegingen om een algoritme, dat wordt gebruikt voor het voorspellen van criminaliteit, verantwoord in te kunnen zetten in de praktijk?'

Uit literatuuronderzoek is gebleken dat de resultaten van het inzetten van dergelijke algoritmes bemoedigend zijn: het blijkt dat er politieafdelingen zijn waar PP algoritmes worden gebruikt waar de criminaliteit is gedaald. In plaats van pas in te grijpen nadat de misdaad is begaan, wordt er met behulp van zowel location- als person-based PP algoritmes preventief opgetreden. Hiermee kan voorkomen worden dat er slachtoffers vallen en kunnen mensen met een hoog risico op het plegen van misdaden geholpen worden. Dit heeft als gevolg dat de algehele criminaliteit zal dalen. Zo hebben PP algoritmes in Kansas en Los Angeles al hun vruchten afgeworpen en is de criminaliteit fors gedaald (Predpol, z.d.; Ferguson, 2017, p.37).

Aan de andere kant is echter gebleken dat PP algoritmes nog niet voldoende zijn ontwikkeld en dat er te weinig aandacht is geweest voor de ethische aspecten die een rol spelen bij de inzet. Dit heeft bijvoorbeeld ertoe geleid dat er de nodige problemen en protesten zijn geweest toen het person-based PP algoritme LASER is ingezet. Er was te weinig aandacht geweest voor de ethische consequenties, waardoor de privacy in het geding kwam, en het algoritme het probleem van etnisch profileren versterkte. Er zijn schrijnende voorbeelden van personen die zonder strafblad toch werden opgepakt omdat 'het algoritme het zei'. Naar mijn mening is dit een ernstig probleem dat eerst opgelost moet worden voordat PP algoritmes daadwerkelijk op grote schaal gebruikt kunnen worden. Het schenden van het recht op privacy is een inbreuk op het fundament van de grondwet, hetgeen simpelweg onacceptabel is bij het introduceren van een toepassing. Ook de feedback loops, die ervoor zorgen dat de nadruk op bepaalde wijken of personen blijft liggen, vormen een ongewenst neveneffect bij het gebruiken van PP algoritmes. Feedback loops hebben tot gevolg dat bepaalde gebieden worden oververtegenwoordigd en andere juist ondervertegenwoordigd (Lum & Isaac, 2016). Politieagenten worden dus niet efficiënt ingezet, waardoor de criminaliteit niet zal dalen waardoor het doel van PP algoritmes niet bereikt wordt.

Er is in de literatuur al naar meerdere oplossingen voor deze problemen gezocht. Mohler et al. (2018) hebben het probleem van de feedback loops geprobeerd op te lossen door gebruik te maken van 'penalized likelihood-methods', wat de raciale bias neutraliseert. Dit gaat echter ten koste van de nauwkeurigheid van het algoritme. Ook inspecteur-generaal Smith (2018) heeft hier onderzoek naar gedaan. Hij benoemde vooral het belang van transparantie in PP algoritmes. Dit houdt in dat het voor verdachten en voor het publiek heel helder moet zijn op basis van welke informatie iemand verdacht wordt en hoe het algoritme tot de uitspraak is gekomen dat iemand of een wijk als een hoog risico wordt gezien. Dit is belangrijk om een dergelijk hulpmiddel te vertrouwen en om duidelijk te hebben welke aanpassingen gemaakt moeten worden als er problemen in het algoritme zitten, maar ook zodat er controle blijft over het nastreven van de privacywetgeving. Uit het onderzoek van Dietvorst et al. (2016) blijkt dat het kunnen aanpassen van algoritmes zorgt voor een groter vertrouwen. Vertrouwen in algoritmes is nodig voordat daadwerkelijk van de uitkomsten van PP algoritmes worden geaccepteerd. Wanneer dit vertrouwen er niet is zullen de PP algoritmes niet gebruikt kunnen worden.

Het doel om de criminaliteit te verminderen met behulp van PP algoritmes is naar mijn mening goed en nastrevenswaardig. Computers zijn op vele gebieden al verder ontwikkeld dan mensen, dus het is goed om de politie te voorzien van informatie die zij zelf niet kan vergaren. Toch zijn de PP algoritmes nog niet ver genoeg ontwikkeld om ook daadwerkelijk zonder ethische bezwaren toegepast te kunnen worden in de praktijk. De nadelen lijken zwaarder te wegen dan de voordelen, omdat er grote consequenties kunnen zijn voor de bevolking. Het is maar de vraag of de zware inbreuk op privacy van burgers opweegt tegen het doel van het realiseren van een plaatselijke daling van de criminaliteit. Het is een enorme uitdaging voor de kunstmatige intelligentie om op

alle fronten de juiste informatie samen te bundelen. Wetenschappers uit dit gebied hebben een verantwoordelijkheid in het schetsen van een realistisch beeld van wat kunstmatige intelligentie kan bijdragen, maar ook van de grote verantwoordelijkheid die bij de gebruikers ligt. Deskundigen uit de kunstmatige intelligentie moeten samenwerken met sociaal-wetenschappers en juristen om alle ongewenste effecten zo goed mogelijk in kaart te brengen en ervoor te zorgen dat deze worden opgeheven. Hierna zouden PP algoritmes een uitstekende toevoeging zijn: met een enorme hoeveelheid data maken zij het mogelijk om gerichter te patrouilleren en efficiënter politieagenten in te zetten, waardoor de criminaliteit zal dalen.

In dit onderzoek heeft de focus gelegen op de morele problemen rondom PP algoritmes, en wegens de beperkte omvang zijn de statistische aspecten en problematieken niet meegenomen. Dit betekent dat de accuraatheid en de betrouwbaarheid van de PP algoritmes niet is behandeld in het onderzoek. Ook de discussie van etnisch profileren is niet volledig behandeld, omdat de omvang van die discussie te groot is om uitvoerig te kunnen behandelen. Wel heeft dit onderzoek een goed beeld kunnen geven van het feit dat etnisch profileren bij de politie een grote kwestie is die invloed heeft op de morele overwegingen om PP algoritmes in te kunnen zetten. Een andere beperking is geweest dat ik geen eerstehands toegang tot de exacte data van de politie of de uitkomsten van de PP algoritmes heb kunnen gebruiken voor dit onderzoek. Mijn bevindingen zijn dus enkel gebaseerd op de gegevens uit andere literatuur.

Een ander vraagstuk dat niet in dit onderzoek wordt behandeld, maar wel zeer interessant is, gaat erover of de PP algoritmes wel voor de juiste doeleinden gebruikt worden. Person-based PP algoritmes doen voorspellingen over personen die een hogere kans hebben om een misdaad te plegen. Deze informatie wordt gebruikt bij politie-eenheden, maar zou deze informatie niet beter voor andere doeleinden kunnen worden gebruikt? In sectie 3 zagen we dat in Kansas bepaalde overheidsinstanties hulpprogramma's aanboden om te voorkomen dat deze mensen de fout in zouden gaan, wat uiteindelijk zorgde voor een daling in de criminaliteit (Ferguson, 2017, p.37). Dit zou effectiever zijn dan ze voorbarig te verwijderen uit hun omgeving of om met dreigende patrouilles een buurt of wijk angst aan te jagen. Het zou interessant zijn om in een vervolgonderzoek te kijken naar hoe deze person-based algoritmes zouden werken bij bijvoorbeeld sociale werkers of psychiatrische instellingen.

Daarnaast zouden dergelijke algoritmes ook op heel andere terreinen van criminaliteit een belangrijk hulpmiddel kunnen zijn, zoals bij de sterk in opkomst zijnde cybercriminaliteit: in hoeverre zijn er patronen te vinden bij hackers en cybercriminelen en zijn er algoritmes in staat om deze te ontdekken?

## Bibliografie

- Amnesty International. (2020, 29 september). Etnisch profileren. Geraadpleegd op 6 januari 2021, van <https://www.amnesty.nl/wat-we-doen/themas/discriminatie/etnisch-profileren>
- Autoriteit Persoonsgegevens. (z.d.). Introductie AVG. Geraadpleegd op 26 januari 2021, van <https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/algemene-informatie-avg/algemene-informatie-avg>
- Baek, G., & Mooney, T. (2020, 24 februari). LAPD still all-in on data-driven policing after scrapping controversial LASER program — CBSN Originals documentary “Racial Profiling 2.0”. *CBS News*. Geraadpleegd op 16 december 2020, van <https://www.cbsnews.com/news/los-angeles-police-department-laser-data-driven-policing-racial-profiling-2-0-cbsn-originals-documentary/>
- Chiao, V. (2019). Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice. *International Journal of Law in Context*, 15(2), 126–139. <https://doi.org/10.1017/s1744552319000077>
- Dietvorst, B., Simmons, J. P., & Massey, C. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. *Journal of Experimental Psychology: General*, 144 (1), 114-126. <http://dx.doi.org/10.1037/xge0000033>
- Dietvorst, B., Simmons, J. P., & Massey, C. (2016). Overcoming Algorithm Aversion: People will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science*, 64 (3), 1155-1170. <http://dx.doi.org/10.1287/mnsc.2016.2643>
- Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., and Venkatasubramanian, S. (2017). Decision making with limited feedback: Error bounds for recidivism prediction and predictive policing. *In Proceedings of FAT/ML 2017*. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>
- Ferguson, A. (2017). The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement. *New York: NYU Press*. Geraadpleegd op 4 januari 2021, van <http://www.jstor.org/stable/j.ctt1pwtb27>
- Gryffroy, P. (2017, 28 april). Verschillen in privacywetgeving Europa en VS. *Timelex*. Geraadpleegd op 22 januari 2021, van <https://www.timelex.eu/nl/blog/verschillen-privacywetgeving-europa-vs>
- Google Developers. (z.d.). Machine Learning Glossary: Fairness |. Geraadpleegd op 23 januari 2021, van [https://developers.google.com/machine-learning/glossary/fairness#demographic\\_parity](https://developers.google.com/machine-learning/glossary/fairness#demographic_parity)
- Government Publishing Office. (1984, oktober). Title 21 - Food and Drugs. Geraadpleegd op 14 januari 2021, van <https://www.govinfo.gov/content/pkg/USCODE-2011-title21/pdf/USCODE-2011-title21-chap13-subchapI-partD-sec844.pdf>
- Gstrein, O. J., Bunnik, A., & Zwitter, A. (2019). Ethical, Legal and Social Challenges of Predictive Policing. *Católica Law Review*, 3(3), 77-98. Geraadpleegd op 22 januari 2021, van [https://www.rug.nl/research/portal/files/117795047/SSRN\\_id3447158.pdf](https://www.rug.nl/research/portal/files/117795047/SSRN_id3447158.pdf)
- Haskins, C. (2020, 22 april). LAPD Announces It Is Dropping PredPol Because Of The Coronavirus Pandemic. *BuzzFeed News*. Geraadpleegd op 13 januari 2021, van <https://www.buzzfeednews.com/article/carolinehaskins1/los-angeles-police-department-dumping-predpol-predictive>
- Koll, K. (2020). Reactions on Algorithms: a systematic Literature Review of Algorithm Aversion and Algorithm Appreciation. Geraadpleegd op 11 december 2020, van <https://diglib.uibk.ac.at/ulbtirolhs/content/titleinfo/5099158/full.pdf>
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. <https://doi-org.proxy.library.uu.nl/10.1016/j.obhdp.2018.12.005>
- Lum, K. and Isaac, W. (2016), To predict and serve?. *Significance*, 13: 14-19. <https://doi.org/10.1111/j.1740-9713.2016.00960.x>

- Maxius.nl. (2020). *Art. 27 Sv - Artikel 27 Wetboek van Strafvordering :: Maxius.nl voorheen Lexius.nl. (c) 2020 Maxius.nl*. Geraadpleegd op 8 december 2020, van <https://maxius.nl/wetboek-van-strafvordering/artikel27>
- Miller, L. (2020, 22 april). LAPD to end controversial program that aimed to predict crime. *Los Angeles Times*. Geraadpleegd op 5 december 2020, van <https://www.latimes.com/california/story/2020-04-21/lapd-ends-predictive-policing-program>
- Mohler, G., Raje, R., Carter, J., Valasik, M., & Brantingham, J. (2018, October). A penalized likelihood method for balancing accuracy and fairness in predictive policing. In *2018 IEEE international conference on systems, man, and cybernetics (SMC)* (pp. 2454-2459). IEEE. <https://doi.org/10.1109/SMC.2018.00421>
- National Constitution Center. (2021). The 4th Amendment of the U.S. Constitution. Geraadpleegd op 8 januari 2021, van <https://constitutioncenter.org/interactive-constitution/amendment/amendment-iv#:~:text=The%20right%20of%20the%20people,and%20the%20persons%20or%20things>
- O'Donnell, R.M. (2019). Challenging Racist Predictive Policing Algorithms Under The Equal Protection Clause. *New York University School of Law. ISSN: 0028-7881*. Geraadpleegd op 20 november 2020, van <https://www.nyulawreview.org/wp-content/uploads/2019/06/NYULawReview-94-3-ODonnell.pdf>
- Perry, W.L., McInnis, B., Price, C.C., Smith, S.C., Hollywood, J.S. (2013). Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations. *The RAND Corporation*. Geraadpleegd op 3 december 2020, van [https://www.rand.org/content/dam/rand/pubs/research\\_reports/RR200/RR233/RAND\\_RR233.pdf](https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf)
- Poston, B., & Chang, C. (2019, 4 december). Analysis: LAPD searches blacks and Latinos more often. *Los Angeles Times*. Geraadpleegd op 15 januari 2021, van <https://www.latimes.com/local/lanow/la-me-lapd-searches-20190605-story.html>
- Predpol. (z.d.). Recent Examples of Crime Reduction. Geraadpleegd op 3 december 2020, van <https://www.predpol.com/results/>
- Richardson, R., Schultz, J., & Crawford, K. (2019). Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. *New York University Law Review Online, 40-42*. Geraadpleegd op 17 januari 2021, van [https://www.nyulawreview.org/wp-content/uploads/2019/04/NYULawReview-94-Richardson\\_et al-FIN.pdf](https://www.nyulawreview.org/wp-content/uploads/2019/04/NYULawReview-94-Richardson_et al-FIN.pdf)
- Selbst, A.D. (2017). Disparate Impact in Big Data Policing, *52 GA. L. REV. 109, 133-34*. Geraadpleegd op 3 december 2020, van <https://apo.org.au/sites/default/files/resource-files/2018-02/apo-nid134806.pdf>
- Sewell, A. A., Jefferson, K. A., & Lee, H. (2016). Living under surveillance: Gender, psychological distress, and stop-question-and-frisk policing in New York City. *Social Science & Medicine, 159, 1-13*. <https://doi.org/10.1016/j.socscimed.2016.04.024>
- Smith, M. P. (2019, 12 maart). Review of Selected Los Angeles Police Department Data-Driven Policing Strategies. Geraadpleegd op 8 januari 2021, van [http://www.lapdpolicecom.lacity.org/031219/BPC\\_19-0072.pdf](http://www.lapdpolicecom.lacity.org/031219/BPC_19-0072.pdf)
- Stop LAPD Spying Coalition. (2018, 8 mei). Before the Bullet Hits the Body – Dismantling Predictive Policing in Los Angeles. Geraadpleegd op 17 december 2020, van <https://stoplapdspying.org/wp-content/uploads/2018/05/Before-the-Bullet-Hits-the-Body-May-8-2018.pdf>
- Stroud, M. (2014, 19 februari). The minority report: Chicago's new police computer predicts crimes, but is it racist? *The Verge*. Geraadpleegd op 23 januari 2021, van <https://www.theverge.com/2014/2/19/5419854/the-minority-report-this-computer-predicts-crime-but-is-it-racist>

- United States Census Bureau. (2019). *Hispanic or Latino Origin by Specific Origin*. data.census.gov. Geraadpleegd op 23 januari 2021, van <https://data.census.gov/cedsci/table?q=B03001%3A%20HISPANIC%20OR%20LATINO%20ORIGIN%20BY%20SPECIFIC%20ORIGIN&tid=ACSDT1Y2019.B03001&hidePreview=true>
- Walmsley, J. (2020). Artificial intelligence and the value of transparency. *AI & SOCIETY*, 1-11. <https://doi.org/10.1007/s00146-020-01066-z>
- Warren, S. D., & Brandeis, L. D. (1890). The Right to Privacy. *Harvard Law Review*, 4(5), 193. <https://doi.org/10.2307/1321160>
- Washington, A. (February 4, 2019). How to Argue with an Algorithm: Lessons from the COMPAS ProPublica Debate. *In the Colorado Technology Law Journal. Volume 17 Issue 1*. Geraadpleegd op 22 november 2020, van [https://ctlj.colorado.edu/wp-content/uploads/2019/03/4-Washington\\_3.18.19.pdf](https://ctlj.colorado.edu/wp-content/uploads/2019/03/4-Washington_3.18.19.pdf)