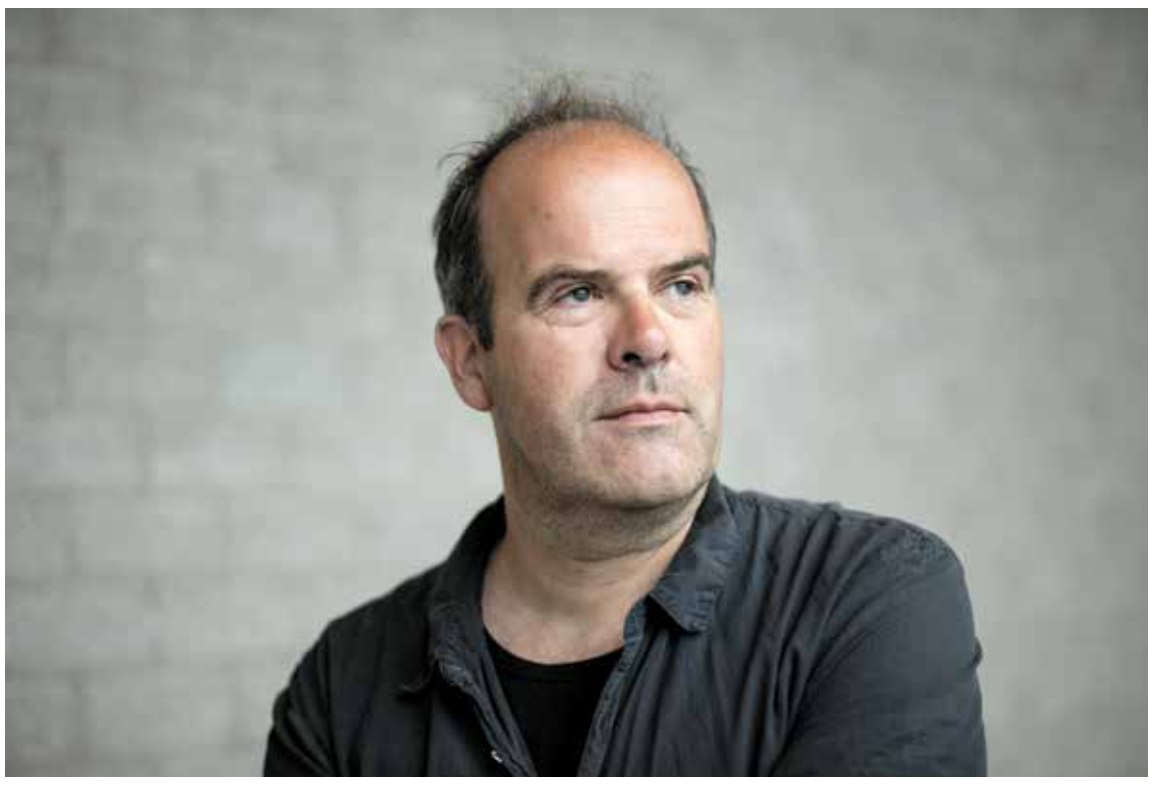
Hoogleraar Digitale Surveillance waarschuwt voor zelflerende algoritmes

‘De strafrechtketen kan niet meer zonder AI. Democratische controle is hard nodig’

Boudewijn Chorus

Justitiële inrichtingen zitten vol met toezichthoudende technologie. Opsporingsinstanties zoals de politie en het Openbaar Ministerie maken in toenemende mate gebruik van technologische hulpmiddelen om te monitoren, af te luisteren en telefonische chats of online berichten te ontsleutelen. Veel van die hulpmiddelen, zo niet alle, maken gebruik van Artificial Intelligence, AI, technieken die met algoritmes werken. Met de groei van al die digitale surveillance wordt de vraag steeds relevanter hoe betrouwbaar de ingezette middelen zijn. Wie controleert de ontwikkeling van die technologie? Is het wel mogelijk om de wildgroei in de toepassing ervan nog te beheersen? De Bonjo stapte met die vragen naar Marc Schuilenburg, hoogleraar Digitale Surveillance aan de Erasmus Universiteit in Rotterdam. Spoiler: zijn antwoorden nemen onze zorgen niet zomaar weg.



Hoogleraar Digitale Surveillance Marc Schuilenburg

U bent sinds 2021 hoogleraar Digitale Surveillance aan de Erasmus Universiteit in Rotterdam. Wat mogen we verstaan onder de term Digitale Surveillance?

“Surveillance komt uit het Latijn en Frans en betekent ‘waken over’ en ‘toezicht houden’. Dat kijken en bekeken worden is zo oud als de mensheid. Aanvankelijk was surveillance lange tijd fysiek, van mens naar mens en gebeurde met het oog. Dat gebeurt nog steeds, bijvoorbeeld als je je kinderen laat zwemmen in een zwembad, dan vindt surveillance plaats door de badmeester.
Op een gegeven moment wordt surveillance steeds meer technologisch, vooral met de opkomst van de camera. Dat kun je de tweede fase van surveillance noemen. Die camera is dan het fysieke element geworden. Die is technologisch en staat op één plek.
In de jaren negentig zie je dat camera’s in een netwerk worden opgenomen. Dat is de derde grote fase van surveillance. Een beroemd voorbeeld is CCTV in Londen (Closed Circuit TV, besloten tv-netwerk), waar je per dag ruim 400 keer wordt gefilmd door verschillende camera’s, als je door Londen loopt.
Door dit ‘vernetwerken’ – het aan elkaar koppelen van camera’s – kun je over een grotere afstand gevolgd worden, zodat er traject-controle ontstaat en je een of meerdere personen de hele binnenstad door kunt volgen.
In de afgelopen tien jaar zie je dat daar nog een slag overheen komt. Over de digitalisering van de camera en de vernetwerking heen komen nu AI (Artificial Intellicence – kunstmatige intelligentie) en algoritmes. En dat betekent dat de camera opeens wordt uitgerust met gezichtsherkenningstechnologieën, zoals biometrie (het vaststellen van meetbare eigenschappen, zoals gezichtskenmerken). Of de camera kan connecties maken met enorme databases waar foto’s in staan van veroordeelden of van mensen die ontsnapt zijn, een belastingschuld hebben of die voetbalhooligans zijn.”

Focust u in uw onderwijs en onderzoek momenteel vooral op die vierde actuele fase in de ontwikkeling van surveillance?

“Ja, mijn leerstoel Digitale Surveillance richt zich met name op hoe AI en algoritmes die nieuwe slag over het ‘kijken en bekeken worden’ heen leggen. Daarmee komen totaal nieuwe vragen op met betrekking tot publieke waarden, zoals transparantie van een algoritme, betrouwbaarheid van de overheid, de output, kun je die nog wel begrijpen en vertrouwen of gaat het systeem zelf aan de haal, – denk bijvoorbeeld aan het OxRec algoritme, dat op een bepaalde manier voorspelt wat het recidiverisico is van een verdachte of van een gedetineerde. Naar dit soort kwesties wordt nauwelijks empirisch onderzoek gedaan. En dat terwijl we momenteel wel een enorme AI-mateloosheid zien. Alsof AI en algoritmes alle problemen gaan oplossen waar we voor staan. Daarom doe ik – samen met mijn promovendi – onderzoek in Nederland binnen de veiligheidspraktijk. Hoe worden daar AI en algoritmes ingezet en hoe verandert hierdoor bijvoorbeeld het voorkomen en opsporen van criminaliteit?”

We zien momenteel een enorme AI-mateloosheid

Kunt u een voorbeeld noemen over hoe AI wordt ingezet bij veiligheid? Gaat het dan bijvoorbeeld om het vergelijken van actuele camerabeelden van personen met eerder genomen beelden?

“Met gezichtsherkenningstechnologie worden camerabeelden gekoppeld aan big data, grote datasets. Om een match te kunnen maken tussen een beeld van jou en pakweg 1 miljoen mensen die in zo’n dataset zitten, heb je algoritmes nodig die enorm snel het beeld van jou kunnen linken aan beelden in die datasets die vergelijkbaar zijn met jou of daar mee overeen komen.
Dat is dus een flinke stap verder dan de vernetwerkte camera’s van CCTV. Want nu worden die camera’s gekoppeld aan zowel biometrie als aan datasets.
Het hele gebied van AI en algoritmes in de veiligheidspraktijk is natuurlijk veel breder dan dat. De politie voorspelt er ook criminaliteit mee. Het wordt gebruikt in de EncroChat-zaak met de miljoenen onderschepte crypto-communicatie data om terug te kunnen kijken. Daar heb je AI voor nodig, omdat je zelf niet meer in staat bent – of het natuurlijk veel te veel tijd zou kosten – om door al die in beslag genomen telefoongesprekken en chats te gaan.”

Elk algoritme dient transparant te zijn

Uw nieuwe boek komt deze week uit: Making Surveillance Public: Why You Should Be More Woke About AI and Algorithms (Maak surveillance publiek: Waarom men meer woke zou moeten zijn over AI en algoritmes). Voor ons is de vraag natuurlijk interessant of u daarin ook aandacht besteedt aan het digitale toezicht in justitiele inrichtingen.

“Het boek gaat over de hele strafketen, hoe deze – van politie tot rechters, het gevangeniswezen, en de reclassering – verandert door de opkomst van AI en algoritmes. Het gaat dan vaak over simpele algoritmes van het niveau Excel-sheet ‘als A - dan B’ die nog heel veel gebruikt worden, zoals het OxRec-algoritme. Maar je hebt ook ‘robotrechters’, algoritmes die digitaal vonnissen uitspugen. En vergeet niet de politie. De politie maakt gebruik van drie soorten algoritmes. Algoritmes die criminaliteit voorspellen, in het geval van ‘predictive policing’. Algoritmes die in het hier en nu

kijken, zoals gezichtsherkenningstechnologie. En algoritmes die terugkijken in enorme in beslag genomen databestanden. Daar is de EncroChat-zaak een actueel voorbeeld van."
Autoriteit Persoonsgegevens
"In de praktijk is het onderscheid dat ik maak tussen terugkijken, hier en nu kijken en vooruitkijken niet zo scherp te maken. Het terugkijken kan immers weer leiden tot nieuwe opsporingsonderzoeken. Dat gaat allemaal gepaard met juridische kwesties die nu niet goed zijn geregeld in het huidige wetboek. Zo is het verwerken van gegevens geregeld in de Wet Politiegegevens. Met die wet is de AP de belangrijke autoriteit, de Autoriteit Persoonsgegevens.
Maar het verzamelen van gegevens is juist geregeld in Wetboek van Strafvordering met de rechter en de officier van justitie als belangrijkste personen. Dus je ziet verschillende regimes waarbij het steeds lastiger wordt om het verzamelen en verwerken van elkaar te scheiden. Vaak doen die tools namelijk alles in één.
Ik zie dan ook steeds meer dat AI-tools en algoritmes bij de genoemde partijen in de strafrechtsketen – en trouwens ook op het departement – worden gebruikt om zowel de veiligheid te vergroten als om efficiëntiewinsten te behalen. Dat is de dominante benadering, die noem ik technisch-economisch.
Veiligheid en efficiëntie zijn beide publieke waarden. Maar je hebt

De criminologie heeft digitale controle verwaarloosd

ook andere publieke waarden. Bijvoorbeeld publieke waarden als betrouwbaarheid, transparantie, non-discriminatie, privacy. De stelling van het boek is dat die sets van publieke waarden het onderspit delven ten opzichte van de sturende waarden, veiligheid en efficiëntie. Vandaar ook de titel van het boek.
Het boek is een pleidooi om al bij het ontwerp van nieuwe technologie in de veiligheidspraktijk, dus ook in de strafketen, te proberen alle publieke waarden te verdisconteren. Want als je dat niet doet, als die technologie er eenmaal is, dan gaat die alle kanten op en dan wordt het er niet meer beter van. Daarom moet je al in de ontwerpfase andere vormen van kennis en andere personen betrekken, dan alleen de technische kennis van dataprofessionals. Daar is het boek mede een pleidooi voor."

Publieke waarden borgen

"De titel van het boek heeft drie betekenissen. Met ***Making surveillance public*** wil ik laten zien welke vormen van surveillance we vaak niet meer als zodanig herkennen. Van de Apple Watch en de Tesla, die ook alles opneemt, tot de Amazon deurbel Ring en de Fitbit (een activiteiten-tracker die fysieke prestaties meet) tot de locatie-trackers (waar is mijn kind nu?), dat is allemaal surveillance, dat gebeurt allemaal in de veiligheidspraktijk.
De tweede betekenis is dat het debat over publieke waarden beter moet worden gevoerd. Dus niet alleen de publieke waarden veiligheid en efficiëntie, maar ook non-discriminatie, privacy, accountability en transparantie.
En de derde betekenis is dat je al bij het ontwerp van nieuwe technologie een breed publiek verzamelt. Niet alleen een publiek van technici, maar denk ook ethici, juristen, minderheden, gedetineerden, om andere lagen van de bevolking en andere vormen van kennis in de ontwerpfase te betrekken. Want technische kennis alleen is onvoldoende is om die publieke waarden goed te kunnen borgen."

In een online promotietekst van Gary Marx voor uw nieuwe boek stelt hij dat de criminologie het onderwerp digitale controle van mensen heeft verwaarloosd. Geldt dit naar uw mening ook voor de criminologie hier te lande?

"Gary Marx is een grote naam in de criminologie. En wat hij zegt over digitale controle klopt, er is heel weinig aandacht in de criminologie voor surveillance. Ook in Nederland. Onderzoek naar het gebruik van technologie komt nauwelijks voor. Criminologen richten zich vooral op personen.
Maar nu zie je dat er zelfs een vorm van AI-criminaliteit gaat ontstaan die de technologie in de hand werkt. Je kunt aan een chatbot (een geautomatiseerde gesprekspartner) als *ChatGPT* vragen: 'Schrijf een code om op onderwerp x of y fraude te plegen'. En die schrijft dat gewoon!
Ik bedoel daarmee te zeggen, dat de technologie zelf een actor wordt. Algoritmes gaan zelf nadenken en zelf verbanden leggen. Zo'n algoritme is wel door iemand gemaakt, maar het bijzondere is, dat dat algoritme vervolgens zelf eigen connecties gaat leggen. Zo'n zelflerend algoritme onttrekt zich na creatie ervan vaak aan het zicht en vermogen van dataprofessionals. Dus je weet niet meer wat onder de motorkap gebeurt.
Maar ook de aandacht die er wel is voor digitalisering, beperkt zich eigenlijk tot vormen van cybercriminaliteit, terwijl AI-criminaliteit de nieuwe grote dreiging wordt. De criminologie onderkent de impact die AI en algoritmes gaan hebben in de volledige strafrechtketen onvoldoende. Intussen zijn we daar al zover dat de systemen zonder AI functioneren niet meer functioneren. De hele samenleving functioneert trouwens niet meer zonder AI."

Dat er geen controle meer kan plaatsvinden op wat zelflerende algoritmes kunnen, klinkt behoorlijk zorgwekkend.

"Daarom houd ik ook een pleidooi om niet zo snel mee te gaan in technologische innovaties. Vanochtend nog heb ik hier aan tafel de directeur-generaal van een ministerie gehad bij wiens enthousiasme voor technologische ontwikkelingen ik toch wat mitsen en maren moest plaatsen. Zo zijn er verschillende onderzoeken geweest naar voorspellende algoritmes. Onder andere bij de politie. Voorspellen waar en door wie criminaliteit wordt begaan en wie slachtoffer van criminaliteit zou worden. Daarbij blijkt dat data die voor de voorspelling gebruikt worden 'vuil' kunnen zijn. Vuile data zijn in dit verband bijvoorbeeld data die komen uit onrechtmatig optreden en/of die gewoon onjuist zijn. Als je je systeem voedt met vuile data, zijn de uitkomsten ook vuil. Garbage in, garbage out wordt dat genoemd. Dat is een van de bekende nadelen van het werken met AI en algoritmes. Als die besmet zijn met vuile data, zijn de uitkomsten ook besmet.

Bij zulke zorgen lijkt me de vraag gerechtvaardigd of u een groot publiek hebt voor uw bevindingen. Zijn er naar uw idee voldoende mensen geïnteresseerd in uw wetenschappelijke publicaties?

"Ik ben een tijdje ook columnist geweest van NRC Handelsblad. Ik kan wel zeggen, dat je als columnist meer impact hebt dan als wetenschapper."

Maakt dat u ontevreden over wat u in uw huidige functie doet?

"Laat ik het zo zeggen: het verbaast mij, hoe justitie in de strafrechtketen allerlei vormen van AI invoert zonder te beschikken over empirisch wetenschappelijk bewijs of het ook werkt. Mijn boek is ook een oproep aan justitie om de band met de wetenschap meer aan te halen. Omdat het zonder empirisch bewijs invoeren van dit soort systemen – waarvan we weten hoe ingewikkeld ze zijn en waarvan de uitkomsten zo diffuus zijn – kan leiden tot veel problemen."

Een van de actuele zorgen op allerlei terreinen waar gebruik van digitale middelen wordt gemaakt, betreft de controle op algoritmes. In dat verband valt onder meer een term als algoracism op. Kunt u voorbeelden noemen van in ons land werkende systemen die van dergelijke algoritmes gebruik maken?

"Kijk naar de Toeslagenaffaire. De Belastingdienst heeft gebruik gemaakt van algoritmes die kennelijk gebaseerd waren op vooroordelen. Mensen uit achterstandswijken, mensen met dubbele nationaliteit, buitenlandse achternamen, etniciteit. Met zulke criteria vol vooronderstellingen werd het algoritme gevoed om 'fraudeurs' onder ontvangers van toeslagen op te sporen. Dit soort algoritmes versterken die vooronderstellingen alleen maar."

Datageweld

"Een ander voorbeeld is *Waze*. Dat is een app waarmee je je route bepaalt in het verkeer. Vergelijkbare apps eerder waren *Ghetto Tracker* waarvan de naam later is veranderd in Good Part of Town. Die apps worden zo gevoed dat bepaalde gebieden rood worden omcirkeld. Dus die raden dan een route aan met omwegen, zodat je niet door die rode gebieden gaat rijden. De vraag is waarom die gebieden rood kleuren.

Je weet niet meer wat er onder de motorkap gebeurt

Idealiter gaat het om criminaliteitscijfers. Dat is nog begrijpelijk en tot op bepaalde hoogte ook toetsbaar.
Dat wordt anders als het om subjectieve indrukken, gevoelens en 'ervaringen' van gebruikers gaat. Want wie toetst of die reëel zijn of dat ze gebaseerd zijn op aannames? Daar zie je dus dat stigmatiserende effecten en ook racistische vooroordelen eenvoudig via de backdoor van technologie binnenkomen. Zulke vormen van algoracisme, wat ik ook wel datageweld noem, worden eigenlijk nooit herkend als geweld. Omdat we al gauw denken dat het hierbij gaat om handige dingen. Als ik mijn studenten vraag of iemand van hen wel eens een algoritme heeft gezien, steekt ook niemand zijn vinger op. Terwijl er op elke telefoon minimaal zo'n 25 zitten. Niemand weet hoe zo'n ding eruitziet. Maar zo'n app kan dankzij de algoritmes waar mee gewerkt wordt een vorm van datageweld opleveren met heel ongewenste consequenties."

VERVOLG OP PAGINA 18

Wie is Marc Schuilenburg?

Marc Schuilenburg (1971) is hoogleraar Digitale Surveillance aan de Erasmus Universiteit Rotterdam. In 1996 studeerde hij af als jurist, in 1998 behaalde hij tevens een master in de filosofie. In 2012 promoveerde hij in de Sociale Wetenschappen. Voor zijn benoeming tot hoogleraar werkte hij onder meer zes jaar in diverse functies bij het Openbaar Ministerie. Van 2015 tot 2022 was hij online columnist van NRC-Handelsblad. Sinds 2021 is hij redacteur van het Tijdschrift voor de Politie. Hij schreef diverse boeken op zijn vakgebied.

VERVOLG VAN PAG.17

Er moet meer democratische controle op AI komen

Zijn dergelijke algoritmes niet per definitie discriminerend? Zodra je immers gaat onderscheiden, dan ben je al heel snel aan het discrimineren. Oftewel: kunnen zulke apps überhaupt 'schoon' werken?

"Helemaal schone data bestaan vrijwel niet. Je werkt al gauw met data waar je niet zomaar een voorspellende waarde aan kunt geven. Algoritme-ontwerpers moeten zich daarvan bewust zijn. Dat is bij het OxRec-instrument goed gebeurd. Dat werkt met een bepaald aantal 'als-dan'-variabelen. Elke variabele is wetenschappelijk onderbouwd. Er zijn bovendien tests geweest met controlegroepen in Noorse gevangenissen. Dan rolt er een min of meer betrouwbaar algoritme uit. Maar kun je vertekening helemaal uitbannen? Nee. Wel kun je het algoritme veel meer transparant maken, de keuzes die je hebt gemaakt en de gevolgen van die keuzes."

De Nederlandse Orde van Advocaten heeft grote zorgen geuit over nieuwe wetgeving die cameratoezicht mogelijk maakt bij contacten van advocaten met streng beveiligde gedetineerden in de EBI's en AIT's. Dat gaat vooral om aantasting van het beroepsgeheim van advocaten. Ook uit ander oogpunt zijn er opmerkingen te plaatsen bij deze ontwikkeling. Wat vindt u van dit soort beveiliging?

"Allereerst moet je differentiëren en zulke maatregelen alleen bij de zwaarste categorie gevangenen toepassen. Het gaat om visuele registratie, niet om afluisteren. Op basis van ervaringen in het buitenland, maar inmiddels ook in ons land, is het risico van in detentie voortgezette criminaliteit helaas reëel gebleken. Wij moeten dus niet naïef zijn en ervan uitgaan dat de kans groot is dat dit gedrag hier ook gaat toenemen. En als het uitdrukkelijk bij visuele registratie blijft en alleen bij de zwaarste categorieën wordt ingezet, denk ik dat de publieke waarde van veiligheid hierbij voorrang mag hebben boven een lichte inperking van het beroepsgeheim van advocaten."

Is het niet naïef om te veronderstellen dat als door de camera beeld wordt opgenomen dat daarbij geen geluid wordt opgenomen?

"We moeten zeker waakzaam blijven als zich verdere uitbreiding aandient. Mijn ervaring met technologie is – zoals die van de meeste mensen die ermee werken – dat het altijd wordt uitgebreid. Als het kan – en dat is in dit geval heel eenvoudig – om extra toepassingen erbij te schakelen, dan gebeurt het vroeg of laat ook. Tegelijk vind ik ook, dat als de samenleving volledig digitaliseert en kwaadwillende krachten steeds innovatiever te werk gaan, dat je dan je ogen niet moet sluiten voor de baten die nieuwe vormen van technologie kunnen hebben."

Tot slot nog een vraag over de toelaatbaarheid van het bewijs dat afkomstig is van PGP-telefoons. Wat vindt u van de bezwaren van veel mensen in de advocatuur, dat zij niet in staat worden gesteld om het bijvoorbeeld door Franse opsporingsautoriteiten ontsleutelde berichtenverkeer op dergelijke telefoons te controleren?

"Van oudsher – en dat zeg ik als jurist – had je eerst de verdenking (artikel 12 Wetboek van Strafvordering) en dan het nadere onderzoek via surveillance. De verdachte persoon kun je dan afluisteren, schaduwen enzovoort. Nu onze samenleving in toenemende mate gedigitaliseerd raakt, zie je dat die volgorde zich omdraait. We hebben nu eerst de surveillance en dan pas de verdenking. Wat de gevolgen van die verschuiving zijn, wordt nog onvoldoende onderkend. De politie gaat in hard tempo beschikken over immense datasets, zowel publieke als private datasets. Daarmee breekt voor de opsporingsinstanties een heel nieuw tijdperk aan. Het voorspellen van criminaliteit wordt minder belangrijk, het terugkijken en in het hier-en-nu kijken gaan een veel grotere plaats innemen dan voorheen. Dit betekent dat de democratische controle op het gebruik van AI en algoritmes moet worden versterkt. Nu kunnen de politie en de strafrechtsketen vaak autonoom beslissen welke nieuwe tools zij nodig hebben en welke zij willen inzetten, terwijl het eigenlijk aan de politiek zou moeten zijn om te beslissen of die tools wel gewenst zijn en zich voldoende verhouden tot belangrijke publieke waarden en grondrechten. Daarom moet de democratische controle op AI en algoritmes door de overheid worden versterkt, al bij het ontwerp ervan, en moet die controle ook door het parlement worden afgedwongen en niet worden overgelaten aan de uitvoerende partijen zelf."

Schuilenburgs recent verschenen boek 'Making Surveillance Public' wordt besproken op pagina 24.