eLaw Working Paper Series

No 2021/00 i - ELAW– 2021

M~¡uruoučxq,uz¡qxxusqz¡uq,uz,tq¡,_¡~mr~qot¡

Qqz,{£q~§uot¡,£mz,mo¡¢qxq,{z¡¤uwwqxuzsqz
Custers, B.H.M.

Universiteit Leiden
eLaw

Discover the world at Leiden University

B.H.M. Custers[1]

Artikelen

Artificiële intelligentie in het strafrecht

Een overzicht van actuele ontwikkelingen

Computerrecht 2021/157

Steeds vaker maken criminelen gebruik van ontwikkelingen in artificiële intelligentie (AI), maar ook bij politie en justitie groeit de interesse naar de mogelijkheden van de inzet van AI. In deze bijdrage worden ontwikkelingen onderzocht in zowel materieel als formeel strafrecht en worden bijbehorende rechtsvragen geïdentificeerd. Bij materieel strafrecht komt AI-technologie (A/B optimalisatie, *deepfake* technologie, *big data analytics)* aan bod die bijdraagt aan bestaande en nieuwe vormen van criminaliteit. Ook wordt ingegaan op de rol van AI in straffen en justitiële interventies, waaronder instrumenten voor risicotaxatie en evidence-based sanctioning. Bij formeel strafrecht wordt AI als opsporingstechnologie *(predictive policing, cyber agent technology)* onderzocht en komt de rol van AI bij bewijs (data-analyse na inbeslagname, waardering van bewijs, scenariovorming) aan bod. Ter afsluiting worden focusgebieden voor verder juridisch onderzoek voorgesteld.

1. Inleiding

Artificiële intelligentie (AI) is het nieuwe modewoord. In Nederland en andere landen worden grote onderzoeksbudgetten vrijgemaakt voor dit onderwerp.[2] De verwachting is dat AI veranderingen zal brengen in veel maatschappelijke sectoren. Ook binnen het juridische domein zal AI veranderingen brengen, enerzijds omdat de ontwikkelingen op het terrein van AI mogelijk om nieuwe, andere of verdere regulering vragen en anderzijds omdat AI steeds meer toepassingen biedt voor juridisch onderzoek en de rechtspraktijk.[3]

Inmiddels is er voor de Nederlandse situatie het een en ander gepubliceerd over de rol van AI[4] in het staats- en bestuursrecht[5] en in het privaatrecht.[6] Deze bijdrage, gebaseerd op literatuuronderzoek, beoogt dit aan te vullen door een overzicht te geven van AI-ontwikkelingen in het strafrecht.[7] Bij het materieel strafrecht wordt ingegaan op het gebruik van AI door criminelen en het gebruik van AI bij het opleggen van sancties en maatregelen (paragraaf 2). Bij het formeel strafrecht wordt ingegaan op het gebruik van AI in de opsporing en vervolging en de rol van AI in strafrechtelijk bewijs (paragraaf 3). Bij beide onderdelen wordt geïnventariseerd welke nieuwe (typen) rechtsvragen deze ontwikkelingen oproepen.
Deze bijdrage sluit af met conclusies en identificeert onderwerpen voor verder juridisch onderzoek (paragraaf 4).

In deze bijdrage is geen beschrijving opgenomen van wat AI is of wat hier wordt verstaan onder AI.[8] Daarover bestaat ook geen eenduidigheid in de literatuur.[9] Om zo veel mogelijk vrij te blijven van de discussie wat wel en niet telt als AI, is ervoor gekozen in deze bijdrage alleen vormen van AI-technologie te bespreken die duidelijk zelflerend en autonoom zijn. Met andere woorden, het gaat om technologie die zich naar verloop van tijd anders kan gedragen en die zelfstandig, dus zonder menselijke tussenkomst, kan handelen.[10] Verder worden in deze bijdrage alleen bestaande voorbeelden beschreven en geen toekomstige, hypothetische voorbeelden. Het doel hiervan is een actueel overzicht te geven van wat er momenteel gebeurt op het gebied van AI en het strafrecht in Nederland. Een dergelijk overzicht kan bijdragen aan de discussie over de regulering van AI, in het bijzonder in het strafrecht. Deze discussie is zeer actueel geworden nu de EU in april 2021 een con-

1 Prof. mr. dr. ir. B.H.M. (Bart) Custers is hoogleraar Law & Data Science bij eLaw, het centrum voor recht en digitale technologie van de Universiteit Leiden.

2 Zie in Nederland bijvoorbeeld https://nlaic.com en https://www.universiteitleiden.nl/en/sails. Zie ook M. Rosemain & M. Rose, 'France to spend $1.8 billion on AI to compete with US, China', *Reuters*, 29 March 2018; J. Harper, 'Federal AI Spending to Top $6 Billion', *National Defense Magazine* 10 February 2021.

3 B.H.M. Custers & F. Leeuw, 'Legal big data', *NJB* 2017/1854, p. 2449-2456.

4 Zie bijvoorbeeld het themanummer over AI van dit tijdschrift: *Computerrecht* 2020, afl. 1.

5 C.J. Wolswinkel, 'AR meets AI: een bestuursrechtelijk perspectief op een nieuwe generatie besluitvorming', *Computerrecht* 2020/4, p. 22-29; R. Passchier, 'Artificiële intelligentie en de rechtsstaat', Den Haag: Boom Uitgevers 2021; B.H.M. Custers, 'Nieuwe digitale (grond)rechten', *NJB* 2019/2775, p. 3288-3295.

6 P.H. Blok, 'Big data en het recht: een overzicht van het juridisch kader voor big data-toepassingen in de private sector', Den Haag: Sdu 2017.

7 Zie ook B.W. Schermer & J.J. Oerlemans, 'AI, Strafrecht en het recht op een eerlijk proces', *Computerrecht* 2020/3, p. 14-21.

8 Zie daarvoor bijv. het themanummer over AI van dit tijdschrift: *Computerrecht* 2020, afl. 1.

9 M.U. Scherer, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies', *Harvard Journal of Law & Technology*, 2015, Vol. 29, No. 2, Spring 2016, p. 353-400; R. Calo, 'Artificial Intelligence Policy: A Primer and Roadmap' (2017): https://ssrn.com/abstract=3015350. Meestal wordt daarbij een onderscheid gemaakt tussen algemene en specifieke AI. Een brede of smalle interpretatie is onder meer relevant voor het toepassingsbereik van de AVG, zie M. Hintze, 'Automated Individual Decisions to Disclose Personal Data: Why GDPR Article 22 Should Not Apply' (2020): http://dx.doi.org/10.2139/ssrn.3630026.

10 Het zelflerende aspect kan onder menselijke supervisie staan, waarbij mensen de AI-technologie trainen, maar het ook ongesuperviseerd zijn, waarbij de AI-technologie zichzelf traint op basis van beschikbare data. Voor meer details, zie bijv. K. Warwick, '*Artificial Intelligence: The Basics*', London: Routledge 2012.

ceptwetsvoorstel heeft gepubliceerd om (het gebruik van) AI stevig te reguleren binnen de EU.[11]

2. AI en materieel strafrecht

De ontwikkelingen op het gebied van AI bieden verschillende mogelijkheden voor criminelen.[12] Hier volgen verschillende voorbeelden van (vormen van) criminaliteit die mogelijk worden gemaakt door AI en vormen van criminaliteit die door de inzet van AI een hoge vlucht hebben genomen.

2.1 *A/B optimalisatie*

Op veel websites, bijvoorbeeld online winkels, websites voor hotelboekingen en nieuwssites, wordt gebruikgemaakt van zogeheten *A/B-testing*[13] (ook wel A/B optimalisatie genoemd). Bij A/B testing krijgen sommige gebruikers scherm A aangeboden en andere gebruikers scherm B. Scherm A en B hebben slechts één verschil, soms heel subtiel. Het verschil kan bijvoorbeeld zijn een lichtgele of een lichtblauwe achtergrondkleur, een logo in zwarte letters of donkerblauwe letters, of al dan niet een lijntje onder de kopteksten. Bij zowel variant A als B wordt gemonitord hoe lang mensen op de website blijven, reclame aanklikken of iets bestellen. Als blijkt dat variant A betere resultaten oplevert dan variant B, wordt die laatste verworpen en verdergegaan met variant A. Door dit herhaaldelijk en op grote aantallen gebruikers toe te passen, wordt een optimaal resultaat bereikt, namelijk de voor gebruikers meest verleidelijke manier van informatie aanbieden. In feite worden alle gebruikers tegelijkertijd als proefkonijn gebruikt om uit te vinden wat het beste werkt.

Uiteraard vindt A/B testing niet handmatig plaats. In de meeste gevallen zijn het algoritmen (doorgaans gebaseerd op technologieën als *data mining* en *machine learning*) die bepaalde verbanden blootleggen. Zelflerende software kan ook zelf variaties creëren in de lay-out van een website of de tekst van een bericht. Vervolgens wordt via algoritmische besluitvorming de informatie op een bepaalde manier aangeboden. Het is belangrijk te benadrukken dat voor A/B testing geen persoonsgegevens nodig zijn. Het kan worden toegepast op anonieme bezoekers van een website en is geen vorm van personalisatie. Het gaat om algemene voorkeuren, niet om persoonlijke voorkeuren.

Bedrijven kunnen met A/B testing bewerkstellingen dat mensen langer op hun website blijven en zelfs meer bestellingen plaatsen. Ook criminelen hebben dit ontdekt, wanneer ze aan de slag gaan met *phishing*[14] (waarbij ze proberen bankgegevens aan slachtoffers te ontfutselen), *ransomware*[15] (waarbij ze proberen computers of bestanden op slot te zetten en losgeld opeisen) of Whatsapp-fraude[16] (waarbij ze bijvoorbeeld proberen slachtoffers te overtuigen geld over te maken voor een vriend in nood). Telkens is de uitdaging voor een crimineel hetzelfde: overtuig slachtoffers te klikken op een link of bijlage die kwaadaardige software installeert of om direct geld over te maken. Een crimineel is dus telkens op zoek naar het meest overtuigende scherm. Met behulp van A/B testing en (heel veel) proberen kan dit worden geoptimaliseerd. De spam die wordt gebruikt bij al deze vormen van cybercrime, is niet slechts een schot met hagel (zoals de term *phishing* ook suggereert) – criminelen kijken ook wie wanneer toehappen en kunnen daarmee hun aanpak verbeteren (zoals weergegeven in de term *spearphishing*).[17]

Deze ontwikkelingen zorgen ervoor dat schermen die we te zien krijgen steeds echter lijken. Het wordt steeds moeilijker om echt en nep, bijvoorbeeld berichten van je bank of werkgever, van elkaar te onderscheiden. Bij Whatsapp-fraude worden bijvoorbeeld foto's van vrienden of familieleden ingezet om berichten geloofwaardiger te maken. Het is niet verbazingwekkend dat nietsvermoedende slachtoffers in steeds grotere aantallen hier in trappen. Europol ziet de omvang van deze vormen van cybercrime jaar in jaar uit groeien.[18]

2.2 *Deepfake technologie*

In het verlengde hiervan verdient een andere AI-technologie aandacht. *Deepfake* technologie biedt de mogelijkheid bestaande afbeeldingen en bewegende videobeelden te combineren en over elkaar heen te zetten. Ook is er de mogelijkheid om compleet nieuwe beelden te genereren, bijvoorbeeld beelden van niet-bestaande mensen.[19] Deze technologie kost inmiddels weinig en er is geen technische kennis vereist. *Deepfake* technologie kan iemand beter of slechter laten voorkomen of zelfs compleet anders (figuur 1). In alle gevallen kan dit misleidend zijn.

11 Verordening van het Europees Parlement en de Raad tot vaststelling van geharmoniseerde regels betreffende artificiële intelligentie, COM(2021) 206 final, 2021/0106(COD), Brussel, 21 april 2021.

12 M. Caldwell, J.T.A. Andrews, T. Tanay & L.D. Griffin, 'AI-enabled future crime', *Crime Science* 2020, Vol. 9, No. 14, p. 13.

13 R. Kohavi & S. Thomke, 'The Surprising Power of Online Experiments', *Harvard Business Review*, September 2017, p. 74-82.

14 D. Lacey, P. Salmon & P. Glancy, 'Taking the Bait: A Systems Analysis of Phishing Attacks', *Procedia Manufacturing* 2015, Vol. 3, p. 1109-1116.

15 J.J. Oerlemans, B.H.M. Custers, R.L.D. Pool & R. Cornelisse, 'Cybercrime en witwassen; bitcoins, online dienstverleners en andere witwasmethoden bij banking malware en ransomware', Meppel: Boom Criminologie 2016.

16 J. Rooyakkers & M. Weulen Kranenbarg, 'Vissen met een nieuwe hengel: een onderzoek naar betaalverzoekfraude', *Justitiële Verkenningen* 2020, afl. 2, p. 19-43; M.S. van 't Hoff-de Goede & E.R. Leukfeldt, 'Whatsapp-fraude in Nederland' (2019): https://www.dehaagsehogeschool.nl/docs/default-source/default-document-library/w2101-0071-infographic-what%27s-app-fraude-digi.pdf?sfvrsn=3ada4f88_0.

17 W. Jingguo, T. Herath, C. Rui, A. Vishwanath & H.R. Rao, 'Phishing susceptibility: An investigation into the processing of a targeted spear phishing email', *IEEE Trans. Prof. Commun.*, 2012, Vol. 55, No. 4, p. 345-362.

18 Europol, 'The Internet Organised Crime Threat Assessment (IOCTA) 2020', The Hague: Europol 2020.

19 Zie www.thispersondoesnotexist.com.

Figuur 1. *Deepfake* technologie legt videobeelden van gezichten over elkaar heen, waardoor de identiteit van personen onherkenbaar wordt[20]

Als iemand met *deepfake* technologie gunstig of juist ongunstig wordt geportretteerd, kan dat uiteraard de beeldvorming over die persoon enorm beïnvloeden. Dat kan bijvoorbeeld verkiezingen bedreigen, wanneer personen woorden in de mond worden gelegd die verschillen van hun eigenlijke standpunten.[21] Misleidende boodschappen kunnen ook worden ingezet om mensen aan te zetten tot crimineel gedrag of het plegen van terroristische daden.

Een andere vorm van misleiding met behulp van *deepfakes* is de mogelijkheid pornografische beelden te maken van bekende mensen (figuur 2).[22] Daarbij kleedt de technologie mensen als het ware uit, door beelden van beroemdheden en pornografische beelden over elkaar te leggen. Actrices als Emma Watson, Natalie Portman en Gal Gadot zijn hiervan slachtoffer geworden.[23] Ook onbekende mensen zijn steeds vaker slachtoffer. Dergelijke beelden kunnen levens van slachtoffers danig ruïneren, zeker wanneer ze wijdverspreid raken op het internet.[24]

Figuur 2. *Deepfake* technologie voor het maken van pornografie van bekende mensen[25]

Een heel andere vorm van *deepfakes* is het creëren van (beelden van) nieuwe of andere personen. De huidige technology biedt de mogelijkheid om levensechte beelden te genereren van bestaande of niet-bestaande personen. In het eerste geval kunnen overleden mensen weer tot leven worden gewekt en in actuele beelden worden geïncorporeerd. Dat dit ver voorbij entertainmenttoepassingen[26] kan gaan, bleek in 2019 toen de president van Gabon in *deepfake* video's tot het publiek sprak.[27] Hij was na een medische ingreep in het buitenland maandenlang niet meer gezien. De video leidde tot allerlei speculaties en, korte tijd daarna, tot een couppoging. Dit voorbeeld betreft vooral nationale veiligheid, maar criminelen kunnen deepfake technologie ook gebruiken om toegangscontroles met gezichtsherkenning te omzeilen, digitaal bewijs te vervalsen en valse betaalverzoeken te genereren.[28] Beelden van niet-bestaande personen kunnen op langere termijn nog veel meer twijfel zaaien. Personen die iedereen alleen maar van het scherm kent, kunnen mogelijk helemaal niet echt bestaan. Ingezet als acteurs is het risico misschien vooral werkeloosheid voor menselijke acteurs, maar ingezet als politici wordt onnavolgbaar wie werkelijk aan de macht is.

Een andere, zeer controversiëlere vorm van *deepfakes* is het genereren van virtuele kinderpornografie. Hoewel er het argument is dat aan virtuele kinderporno geen kindermisbruik ten grondslag ligt, is het tegenargument dat virtuele kinderporno kan aanzetten tot kindermisbruik. Het is om die reden dat virtuele kinderpornografie in veel landen strafbaar[29] is, via implementatie van het Cybercrimeverdrag[30] en Richtlijn 2011/92/EU ter bestrijding van seksueel misbruik en seksuele uitbuiting van kinderen en kinderpornografie.

2.3 *Big data analytics*

Naast de meer visuele toepassingen van AI-technologie, zijn er ook vormen van AI die veel meer gericht zijn op andere soorten data. Net als in andere sectoren het geval is, kunnen ook criminelen gebruikmaken van *profiling*,[31] een techniek waarbij wordt gelet op eigenschappen en voorkeuren die bepaalde personen hebben. Voor criminelen wordt het daarmee niet alleen makkelijker slachtoffers te overtuigen, zoals hierboven beschreven, maar ook te selecteren welke individuen en groepen makkelijke of vermogende slachtoffers zijn.

20 Bron: Facebook.
21 Zie https://www.youtube.com/watch?v=T76bK2t2r8g.
22 Milena Popova, 'Reading out of context: pornographic deepfakes, celebrity and intimacy', *Porn Studies*, 2020, 7:4, 367-381, DOI: 10.1080/23268743.2019.1675090.
23 D. Lee, 'Deepfakes porn has seriouse consequences', *BBC News*, 3 February 2018.
24 D. Scott, 'Deepfake nearly ruined my life', *Elle*, 6 February 2020.
25 Bron: https://www.ethicsforge.cc/deepfake-the-age-of-disinformation/.
26 Zoals de schilder Dali die met deepfake technology weer tot leven wordt gewekt in het Dali Museum, zie D. Lee, 'Deepfake Salvador Dali takes selfies with museum visitos: it's surreal, all right', *The Verge*, 10 May 2019.
27 S. Cahlan, 'How misinformation helped spark an attempted coup in Gabon', *The Washington Post*, 13 February 2020.
28 M. Caldwell, J.T.A. Andrews, T. Tanay & L.D. Griffin, 'AI-enabled future crime', *Crime Science* 2020, Vol. 9, No. 14, p. 13.
29 Art. 240b Sr. Hier heeft de wetgever bewust de frase 'of schijnbaar is betrokken' ingevoegd, al is dit volgens sommige auteurs ongelukkig (want te veel ruimte voor interpretatie) gekozen, zie B.J. Koops & J.J. Oerlemans, '*Strafrecht en ICT*', Den Haag: Sdu 2019.
30 Convention on Cybercrime, Budapest 23 November 2001.
31 B.H.M. Custers, 'Data Dilemmas in the Information Society', in: B.H.M. Custers, T. Calders, B. Schermer & T. Zarsky (red.), '*Discrimination and Privacy in the Information Society,*Heidelberg: Springer 2013.

In tegenstelling tot A/B testing moet voor *profiling* doorgaans worden onthouden om welke gebruiker het gaat, vandaar dat cookies en andere trackers in deze context een belangrijke rol spelen. Via de voorkeuren die een gebruiker expliciet aangeeft dan wel impliciet laat blijken uit lees- en klikgedrag, kunnen criminelen potentiële slachtoffers selecteren. Ook geldezels voor het witwassen van criminele opbrengsten kunnen zo worden gerekruteerd.[32]

Vormen van cybercrime die hier verder op inspelen, zijn CEO-fraude en Whatsapp-fraude. Beide kennen verschillende varianten, maar bij CEO-fraude is de kern dat uit naam van een (financieel) directeur een betaalopdracht wordt gestuurd naar de financiële afdeling.[33] Bij Whatsapp-fraude is de kern dat de crimineel zich voordoet als een vriend of familielid in nood en dringend geld nodig heeft.[34] Voor deze vormen van criminaliteit moeten cybercriminelen eerst persoonlijk gegevens verzamelen over het slachtoffer en over de persoon als wie ze zich willen voordoen.

2.4 *Evidence-based sanctioning*[35]

Eén van de belangrijke doelen van het opleggen van sancties en maatregelen is specifieke preventie: voorkomen dat een delinquent recidiveert. Op basis van grote hoeveelheden gegevens kan met behulp van geautomatiseerde analyses kwantitatief-empirisch worden bepaald welke interventies het beste resultaat opleveren in termen van recidivereductie. Dit vraagstuk kan op dezelfde wijze worden gemodelleerd als een arts die een patiënt behandelt: op basis van de ziekte of aandoening (maar in toenemende mate ook de karakteristieken van de patiënt)[36] stelt deze de medicatie, therapie of behandeling vast. Naar analogie kunnen ook rechters justitiële interventies 'toedienen' naar de omstandigheden van een casus, waarbij ze kijken welke 'behandeling' het beste werkt vanuit oogpunt van voorkomen van recidive. Specifieke karakteristieken die kunnen worden meegewogen, zijn dan de aard van het delict en de omstandigheden waaronder het begaan is, maar ook persoonskenmerken van dader en slachtoffer. Mogelijke 'behandelingen' zijn dan de strafmodaliteit (gevangenisstraf, taakstraf of boete), voorwaardelijkheid en bijzondere voorwaarden, of de geschiktheid van bepaalde trainings- en opleidingsprogramma's, bijvoorbeeld programma's gericht op het verbeteren van cognitieve of sociale vaardigheden, omgaan met agressief gedrag of verslavingsproblematiek.

Deze *evidence-based* benadering kan zowel worden toegepast op groepsniveau (wat werkt het beste voor een bepaalde categorie) als op individueel niveau (wat werkt het beste in een concreet geval). Voor beide niveaus bestaan reeds toepassingen in Nederland en het buitenland. In Nederland publiceert de overheid data over recidive op groepsniveau via het systeem REPRIS.[37] Deze cijfers komen van de recidivemonitor van het Ministerie van Justitie. Op basis van deze en andere onderzoeksresultaten beoordeelt de Erkenningscommissie Justitiële Interventies de kwaliteit en effectiviteit van de interventies.[38] In de Verenigde Staten publiceert het National Institute of Justice vergelijkbaar onderzoek via de website Crime Solutions.[39] Voor elke interventie wordt aangegeven in hoeverre deze wel of niet effectief is. Veel onderzoek is nog klassiek-empirisch, maar in toenemende mate worden analyses geautomatiseerd om steeds grotere hoeveelheden gegevens te kunnen incorporeren in de beoordelingen.

2.5 *Instrumenten voor risicotaxatie*

Ook op individueel niveau heeft deze aanpak meerwaarde, met name op het terrein van risicotaxatie. Instrumenten voor risicotaxatie worden veel gebruikt in het strafrecht, bijvoorbeeld bij het opleggen van tbs, bij voorwaardelijke invrijheidsstelling en proefverloven. In Nederland maakt de reclassering gebruik van het systeem RISC. Onderdeel daarvan is OXREC, een actuarieel risicotaxatie-instrument waarmee statistische risico's worden voorspeld.[40] In toenemende mate lijken deze modellen een rol te spelen bij het werk van Reclassering Nederland en rechters. In enkele delen van de Verenigde Staten wordt het systeem COMPAS gebruikt voor de inschatting van recidiverisico's.[41] In rechtbanken aldaar wordt sterk geleund op deze modellen.

Het gebruik van dergelijke modellen kan voordelen bieden: inschattingen kunnen verder worden gestructureerd en geobjectiveerd. Subjectieve beoordelaars kunnen inschattingsfouten maken of door voorkeuren en vooroordelen worden beïnvloed. Als de systemen zelflerend zijn, kunnen ze bovendien nieuwe trends en ontwikkelingen

32 B.H.M. Custers, J.J. Oerlemans & R. Pool, 'Laundering the Profits of Ransomware: Money Laundering Methods for Vouchers and Cryptocurrencies', *European Journal of Crime, Criminal Law and Criminal Justice,* 28 (2020), p. 121-152.

33 Europol 'Internet Organised Crime Threat Assessment (IOCTA) 2020', Den Haag: Europol 2020.

34 J. Rooyakkers & M. Weulen Kranenbarg, 'Vissen met een nieuwe hengel: een onderzoek naar betaalverzoekfraude', *Justitiële Verkenningen* 2020, afl. 2, p. 19-43.

35 Het sanctierecht wordt in deze bijdrage als onderdeel van het materieel strafrecht beschouwd, omdat in het algemene materieel strafrecht mogelijke straffen en maatregelen zijn vastgelegd en in het bijzonder materieel strafrecht (maximum) straffen zijn gesteld op elk delict. Echter, zeer regelmatig wordt het sanctierecht (of penitentiair recht) als afzonderlijk gebied in het strafrecht beschouwd. Cf. J. de Hullu, *Materieel strafrecht,* Deventer: Wolters Kluwer 2018.

36 Zogeheten *personalized medicine*.

37 https://data.overheid.nl/dataset/repris.

38 https://www.justitieleinterventies.nl.

39 https://crimesolutions.ojp.gov.

40 https://oxrisk.com/oxrec-nl-2-backup/. Zie ook G. van Dijck, 'Algoritmische risicotaxatie van recidive: over de Oxford Risk of Recidivism tool (OXREC), ongelijke behandeling en discriminatie in strafzaken', *NJB* 2020/1558; M. de Vries, J. Bijlsma, A.R. Mackor, F. Bex & G. Meynen, 'AI-risicotaxatie: Nieuwe kansen en risico's voor statistische voorspellingen van recidive', *Strafblad* 2021, afl. 2, p. 58-66.

41 https://assets.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf.

herkennen en meenemen. En uiteraard kan er sneller en goedkoper worden gewerkt. Maar er is ook kritiek op deze werkwijze, omdat instrumenten soms niet beter lijken te presteren en er risico's aan zijn verbonden, waaronder bias die kan resulteren in discriminatie.[42] Het Amerikaanse systeem COMPAS leek bijvoorbeeld systematisch Afro-Amerikanen hogere recidiverisico's toe te kennen.[43] Daarmee lijkt het systeem discriminerend te zijn. Verdachten hebben via de rechter geprobeerd inzicht te krijgen in dit risicotaxatiesysteem, maar zulke verzoeken zijn bij herhaling afgewezen, onder meer op grond van intellectuele-eigendomsrechten van de makers en gebruikers. COMPAS blijft daarmee ondoorzichtig. Een veelgehoord verweer is dat dergelijke modellen geen etnische gegevens gebruiken en derhalve niet discriminerend zijn.[44] Echter, gegevens als etniciteit kunnen doorgaans eenvoudig worden voorspeld en worden door zelflerende technologie zodoende vaak gereconstrueerd zonder dat dit voor gebruikers zichtbaar is.[45] Voorzichtigheid is dus geboden.[46]

2.6 *Rechtsvragen*

Uit bovenstaande zijn ruwweg drie categorieën rechtsvragen te destilleren voor het materieel strafrecht. De eerste categorie betreft vragen over de interpretatie van geldend recht. Het gaat dan met name om de vraag of bepaalde handelingen gedekt worden door bepaalde strafbepalingen. Dit heeft betrekking op de voorbeelden genoemd bij A/B testing en *deepfake* technologie. Zo kan men zich afvragen wanneer een bepaalde technologie kwalificeert als geautomatiseerd werk (art. 80sexies Sr) of technisch hulpmiddel (bijv. art. 126s Sr). Dit zou kunnen blijken uit de wetsgeschiedenis of jurisprudentie, maar regelmatig zijn er te weinig aanknopingspunten voor interpretatie van nieuwe technologie. De wetgever kan niet altijd alle technologische ontwikkelingen vooraf voorzien en de hoeveelheid jurisprudentie is vaak niet omvangrijk.

De tweede categorie betreft vragen welke handelingen strafwaardig zijn en, als er (nog) geen strafbepalingen voor bestaan, welke nieuwe strafbaarstellingen in het leven zouden moeten worden geroepen. Zo is de vraag welke vormen van *deepfake* technologie wellicht strafbaar gesteld moeten worden.[47] Wellicht zijn er ook redenen om A/B testing voor criminele doeleinden strafbaar te stellen. Als er nieuwe strafbaarstellingen moeten komen voor bepaalde gedragingen, is tevens de vraag welke strafmaat daarbij past. Verder onderzoek op het terrein van *evidence-based* sanctioneren kan daaraan bijdragen.

De derde categorie heeft betrekking op vragen over het gebruik van gegevens. Dit betreft vragen in hoeverre het verzamelen en analyseren van gegevens, bijvoorbeeld bij genoemde risicotaxaties, proportioneel is in verhouding tot andere te beschermen belangen, waaronder intellectueel eigendom, privacy en het recht op non-discriminatie (niet alleen van verdachten, maar ook van niet-verdachten in controlegroepen). Bij het gebruik van gegevens voor criminologisch onderzoek is doorgaans duidelijker wat wel en niet is toegestaan. Bijvoorbeeld art. 89 Algemene verordening gegevensbescherming (AVG) biedt de mogelijkheid om persoonsgegevens onder bepaalde voorwaarden te gebruiken voor wetenschappelijk onderzoek. Voor strafrechtelijk onderzoek bieden de Wet politiegegevens en de Wet justitiële en strafvorderlijke gegevens wettelijke kaders.

3. AI en formeel strafrecht

Ook politie en justitie kunnen op uiteenlopende manieren artificiële intelligentie inzetten. In de opsporing en vervolging kan AI het werk ondersteunen of zelfs (deels) vervangen. Hieronder worden van beide ontwikkelingen voorbeelden gegeven.

3.1 *Predictive policing*

Met behulp van grote hoeveelheden data en geavanceerde analysemethoden kunnen trends en ontwikkelingen in criminaliteit worden ontdekt. Deze toepassingen kunnen ook worden ingezet om voorspellingen omtrent criminaliteit te doen, zoals op welke locaties criminaliteit zal plaatsvinden, wie mogelijk dader of slachtoffer van criminaliteit zal zijn en hoe bijvoorbeeld criminele netwerken of criminele carrières zich zullen ontwikkelen. Dit wordt aangeduid met de term *predictive policing*.[48]

Een typisch voorbeeld zijn zogenoemde *crime heat maps* (figuur 3), waarin misdaadcijfers gevisualiseerd worden op plattegronden. Op zulke plattegronden is goed herkenbaar in welke wijken veel criminaliteit heerst (*'hot spots'*). Met behulp van AI kunnen naast statische kaarten met momentopnames ook dynamische, *real-time* kaarten worden gegenereerd. En naast terugkijken kan ook vooruit

42 G. van Dijck, 'Algoritmische risicotaxatie van recidive: over de Oxford Risk of Recidivism tool (OXREC)', ongelijke behandeling en discriminatie in strafzaken, *NJB* 2020/1558.

43 J. Angwin, J. Larson, S. Mattu & L. Kirchner, 'Machine Bias', *ProPublica*, 23 May 2016.

44 M. Maas, E. Legters & S. Fazel, Professional en risicotaxatie-instrument hand in hand: hoe de reclassering risico's inschat, *NJB* 2020/1814, p. 2055-2059.

45 Cf. F. Kamiran, T. Calders & M. Pechenizkiy, 'Techniques for discrimination-free predictive models', in: Custers, et al. (eds.), '*Discrimination and Privacy in the Information Society*', Heidelberg: Springer 2013.

46 J. Bijlsma, F.J. Bex & G. Meynen, 'Artificiële intelligentie en risicotaxatie: drie kernvragen voor strafrechtjuristen', *NJB* 2019/2778, p. 3313-3319.

47 Momenteel voorwerp van onderzoek binnen het ministerie van justitie, zie: https://www.wodc.nl/onderzoek-in-uitvoering/welk-onderzoek-doen-we/3137---het-tegengaan-van-deepfakes.

48 A.G. Ferguson, 'Predictive Policing Theory', in: T. Rice Lave & E.J. Miller (eds.), *'The Cambridge Handbook of Policing in the United States*', Cambridge University Press 2019; M. Schuilenburg, 'Predictive policing: de opkomst van gedachtenpolitie?', *Ars Aequi*, december 2016, p. 931-936.

worden geblikt, door voorspelmodellen in de kaarten te verwerken. Hiermee worden de kaarten bruikbaar voor surveillanceplanning en andere politiestrategieën.[49]

Figuur 3. Zogeheten *crime heat maps* geven weer in welke wijken criminaliteit hoger is. Met behulp van AI kunnen ook real-time en prospectieve kaarten worden gegenereerd[50]

Predictive policing kan op locatie zijn gericht, maar ook op personen. Met gebruik van hierboven beschreven *profiling* strategieën kunnen voorspellingen worden gedaan wie mogelijk een misdrijf gaat plegen. Dat kan relevant zijn voor recidive, maar natuurlijk ook voor mensen die voor de eerste keer de fout in gaan. Op basis van bepaalde persoonskenmerken en situationele kenmerken kan worden voorspeld wie een verhoogd risico heeft te vervallen in criminaliteit.[51] Ook hier kan AI-gerelateerde technologie nieuwe, onverwachte verbanden blootleggen en veel meer prospectief te werk gaan, onder meer door socialemediadata hierin te betrekken. *Real-time* overzichten stellen de politie in staat ook ter plekke in te grijpen, wanneer de pakkans het grootst is. Hoewel deze aanpak verschillende voordelen biedt wat betreft doeltreffendheid en doelmatigheid van politiewerk, is voorzichtigheid geboden: er kan sprake zijn van waterbedeffecten (verplaatsing van criminaliteit),[52] discriminatie[53] en tunnelvisie ten gevolge van beperkte betrouwbaarheid van de risicoprofielen.[54]

3.2 *Cyber agent technology*

De criminaliteitscijfers nemen al jaren af in westerse landen, maar dat lijkt niet van toepassing op cybercrime. Daar lijkt juist sprake van een toename.[55] Dat is niet heel wonderlijk, want veelal zijn de pakkansen lager en de opbrengsten hoger in vergelijking met offline criminaliteit. Ook traditionele vormen van criminaliteit, zoals georganiseerde drugscriminaliteit, gaan steeds meer online, via online marktplaatsen op het *darkweb* (het niet door zoekmachines geïndexeerde deel van het internet dat alleen toegankelijk is met speciale software). Eén van de eerste illegale marktplaatsen was *Silk Road*, opgericht in 2011 en neergehaald door de FBI in 2013, waar vooral drugs en wapens werden verhandeld, maar ook diensten van huurmoordenaars werden aangeboden. Na het neerhalen van Silk Road volgden nieuwe varianten, waaronder Silk Road 2.0 (2014), Evolution (2015), AlphaBay (2015), Hansa (2017), Outlaw (2017), Digital Shadows (2018), Dream Market (2019), DeepDotWeb (2019) en Darkmarket (2021).[56]

Het is voor opsporingsautoriteiten ingewikkeld en tijdrovend om de activiteiten op deze online marktplaatsen te volgen. Toegang tot deze marktplaatsen vereist bijvoorbeeld het zorgvuldig opbouwen van een reputatie. Voor opsporingsdiensten is het in zulke context regelmatig nodig bijzondere opsporingsbevoegdheden in te kunnen zetten, zoals pseudokoop, stelselmatig inwinnen van informatie, opnemen van vertrouwelijke communicatie, en infiltratie. Zodra deze bevoegdheden worden ingezet, mag uiteraard nergens sprake zijn van uitlokking.[57]

Vanwege het intensieve en precaire karakter van opsporingsactiviteiten op online ondergrondse marktplaatsen, wordt ook AI ingezet. Het gaat dan om *cyber agent techno-*

49 D. Weisburd C.W. Telep, 'Hot Spots Policing', *Journal of Contemporary Criminal Justice*, 2014, 30(2), p. 200-220.

50 Bron: https://spotcrime.wordpress.com/2009/07/20/houston-crime-map-new-data-and-shooting-heat-map/.

51 E.R. Kleemans & C.J. de Poot, 'Criminal Careers in Organized Crime and Social Opportunity Structure', *European Journal of Criminology* 2008, Vol. 5, Nr. 1, p. 69-98.

52 D. Weisburd, L.A. Wyckoff, J. Ready, J.E. Eck, J.C. Hinkle & F. Gajewski, 'Does Crime Just Move Around the Corner? a Controlled Study of Spatial Displacement and Diffusion of Crime Control Benefits', *Criminology* 2006, 44 (3), p. 549-592.

53 S. Barocas & A.D. Selbst, 'Big Data's Disparate Impact'. 104 *California Law Review* 2016, 671.

54 B.H.M. Custers, '*Effects of Unreliable Group Profiling by Means of Data Mining*', in: G. Grieser, Y. Tanaka & A. Yamamoto (eds.), '*Lecture Notes in Artificial Intelligence, Proceedings of the 6th International Conference on Discovery Science (DS 2003)*', Sapporo, Japan. Berlin, Heidelberg, New York: Springer-Verlag 2003, Vol. 2843, p. 290-295.

55 Europol, 'The Internet Organised Crime Threat Assessment (IOCTA) 2020', The Hague: Europol 2020: https://www.europol.europa.eu/activities-services/main-reports/internet-organised-crime-threat-assessment-iocta-2020.

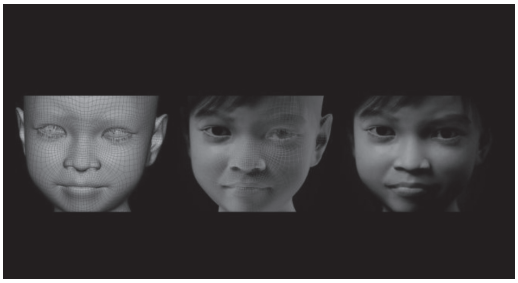
56 See, for instance, Europol Press Release, 'Darkmarket: world's largest illegal dark web marketplace taken down', 12 January 2021: https://www.europol.europa.eu/newsroom/news/darkmarket-worlds-largest-illegal-dark-web-marketplace-taken-down.

57 J.J. Oerlemans & R. van Wegberg, 'Opsporing en bestrijding van online drugsmarkten', *Strafblad* 2019, afl. 5, p. 25-31. Inzet van cyber agent technology, nepaccounts en avatars zal doorgaans vallen in Nederland onder 'werken onder dekmantel', bijvoorbeeld pseudokoop (art. 126i Sv), stelselmatige informatie-inwinning (art. 126j Sv) en infiltratie (art. 126h en 126w Sv). Uitlokking wordt bepaald aan de hand van het Tallon-criterium: verdachte mag niet tot andere strafbare feiten worden gebracht dan waarop diens opzet reeds tevoren was gericht, zie HR 4 december 1979, ECLI:NL:HR:1979:AB7429, *NJ* 1980/356, m.nt. Th. W. van Veen (*Tallon-arrest*).

logy, technologie die *cyber agents* (online actoren) ondersteunt. Deze technologie kan een bepaalde mate van autonomie bevatten en daarmee zelfstandig handelen.[58] Het zijn intelligente programma's die interacties met anderen kunnen aangaan en kunnen handelen zonder tussenkomst van mensen.[59] Door de inzet van deze technologie kunnen veel meer interacties worden aangegaan op de *darkweb forums*.

Eén van de meest concrete toepassingen is een *chatbot* (een geautomatiseerde gesprekspartner) met de naam Sweetie (figuur 4).[60] Vormgegeven als een 10-jarig meisje uit de Filipijnen, kon deze *chatbot* op internet conversaties aangaan met personen die seksuele interesse tonen voor kinderen, met als doel deze verdachte personen te identificeren en vervolgens te vervolgen of te waarschuwen.[61] De *chatbot* is getraind voor basale communicaties, met de woordenschat van een 10-jarige. Zodra er aanwijzingen zijn dat iemand verkeerde bedoelingen heeft, geeft het systeem een signaal richting opsporingsambtenaren. Die kunnen kiezen of ze de conversatie door de chatbot laten vervolgen of zelf overnemen (waarmee de *chatbot* verwordt tot een *avatar*).

Figuur 4. Sweetie 2.0 is cyber agent technology die kan bijdragen aan online opsporing



De AI-technologie is alleen bruikbaar als die zodanig geavanceerd is dat wordt voldaan aan de Turing-test,[62] waarbij mensen niet doorhebben dat ze met AI praten. Dat bleek geen probleem, ongeveer 20.000 mannen uit zo'n 71 landen zochten contact met Sweetie, in de veronderstelling dat zij een echt kind was.[63] De technologie mag bovendien geen illegale gedragingen uitlokken of zelf crimineel gedrag aanleren.[64] Dat criterium bleek veel lastiger: in veel jurisdicties (waaronder Nederland) is er twijfel of mogelijk sprake is van uitlokking.[65] Een ander punt was dat veroordelingen voor kindermisbruik in bepaalde jurisdicties onmogelijk waren omdat er geen echt sprake was van misbruik.[66] Zelfs in jurisdicties waarin alleen intenties reeds strafbaar zijn, bleek vervolging lastig, omdat geen sprake was van een echt kind.[67] Niettemin leidde de technologie in Australië,[68] België[69] en Engeland[70] tot veroordelingen.[71]

Bij het vergaren en beoordelen van strafrechtelijk bewijs kan AI op verschillende manieren een rol spelen. Hieronder komen achtereenvolgens aan bod het doorzoeken van grote hoeveelheden gegevens, het waarderen van bewijs en scenariovorming tijdens de waarheidsvinding.

3.3 *Doorzoeking na inbeslagname*

Onder bepaalde omstandigheden en voorwaarden (waaronder een rechterlijke machtiging) kan de politie tijdens een opsporingsonderzoek gegevensdragers in beslag nemen voor verdere doorzoeking. De politie kan de gegevensdragers, zoals mobiele telefoons, tablets, laptops en usb sticks, vervolgens door het NFI laten doorzoeken op bewijs. Naast beschadigingen of versleutelingen is een probleem bij de gegevensdragers die het NFI ontvangt dat het enorme hoeveelheden gegevens betreft. Slechts kleine brokjes informatie zijn relevant als bewijs, bijvoorbeeld om onderdelen in de bewijsvoering rond te krijgen. In fei-

58 B.W. Schermer, *Software agents, surveillance, and the right to privacy: a legislative framework for agent-enabled surveillance*, Leiden: Leiden University Press 2007.

59 H.S. Nwana, 'Software Agents: An Overview', *Knowledge Engineering Review,* 1996, 21 (3): 205-244; M. Luck, P. McBurney & C. Preist, 'A Manifesto for Agent Technology: Towards Next Generation Computing', *Autonomous Agents and Multi-Agent Systems* 2004, 9, 203-252.

60 https://www.terredeshommes.nl/programmas/sweetie-20-webcam-seks-met-kinderen-de-wereld-uit. Zie ook C, van der Wal, 'Sweetie 2.0: nieuw virtueel meisje gaat op pedojacht', *Algemeen Dagblad* 13 februari 2016. https://www.ad.nl/binnenland/sweetie-2-0-nieuw-virtueel-meisje-gaat-op-pedojacht~ad3739ca/.

61 Deze technologie is ook inzetbaar voor inlichtingen- en veiligheidsdiensten, zie B.H.M. Custers, *Cyber agent technology en de Wet op de Inlichtingen- en Veiligheidsdiensten (WIV)*, Leiden: Universiteit Leiden, 30 september 2017, 30 p.

62 A. Turing, 'Computing machinery and intelligence', *Mind* 1950, 59, p. 433-460.

63 http://www.dawn.com/news/1054244.

64 Zoals Microsoft's chatbot Tay, die al na een paar uur racistische taal ging gebruiken, zie: P. Mason, 'Racist hijacking of Microsoft's chatbot shows how the internet teems with hate', *The Guardian* 29 March 2016.

65 S. van der Hof, I. Georgieva, B.W. Schermer & B.J. Koops, *'Sweetie 2.0. Using artificial intelligence to fight webcam child sex tourism'*, The Hague: Asser Press/Springer 2019.

66 B.W. Schermer, I. Georgieva, S. van der Hof & B.J. Koops, 'Legal aspects of Sweetie 2.0', in: S. van der Hof, I. Georgieva, B.W. Schermer & B.J. Koops (eds.), *'Sweetie 2.0. Using artificial intelligence to fight webcam child sex tourism'*, Information technology & law series nr. 31 The Hague: Asser Press/Springer Press 2019, p. 1-94.

67 Bijvoorbeeld in België is inmiddels de wet aangepast, zodat cyberlokking van een chatbot/niet-bestaande minderjarige ook strafbaar is.

68 G. Urbas, 'Substantive and Procedural Legislation in Australia to Combat Webcam-Related Sexual Child Abuse', in: S. van der Hof, I. Georgieva, B.W. Schermer & B.J. Koops (eds.), *'Sweetie 2.0. Using artificial intelligence to fight webcam child sex tourism'*, Information technology & law series nr. 31 The Hague: Asser Press/Springer Press 2019.

69 S. Royer, C. Conings & G. Marlier, 'Substantive and Procedural Legislation in Belgium to Combat Webcam-Related Sexual Child Abuse', in: S. van der Hof, I. Georgieva, B.W. Schermer & B.J. Koops (eds.), *'Sweetie 2.0. Using artificial intelligence to fight webcam child sex tourism'*, Information technology & law series nr. 31 The Hague: Asser Press/Springer Press 2019.

70 A.A. Gillespie,' Substantive and Procedural Legislation in England and Wales to Combat Webcam-Related Sexual Child Abuse', in: S. van der Hof, I. Georgieva, B.W. Schermer & B.J. Koops (eds.), *'Sweetie 2.0. Using artificial intelligence to fight webcam child sex tourism*,. Information technology & law series nr. 31 The Hague: Asser Press/Springer Press 2019.

71 https://nl.wikipedia.org/wiki/Sweetie_%28virtueel_personage%29.

te is dit een speld-in-een-hooiberg probleem en AI kan hierbij helpen.[72]

Om dit probleem te tackelen, heeft het NFI Hansken[73] ontwikkeld. Dit systeem, een voorbeeld van *big data analytics*, kan grote hoeveelheden gegevens uit verschillende bronnen en in uiteenlopende formaten (tekst, video, audio, enz.) opslaan, indexeren en doorzoekbaar maken. Het labelen van gegevens gebeurt geautomatiseerd. De doorzoekbaarheid van de inbeslaggenomen gegevens verhoogt de effectiviteit van opsporingsinstanties, omdat relevante informatie minder vaak over het hoofd wordt gezien.[74] Daarnaast zorgt Hansken voor zeer snelle resultaten, hetgeen in de opsporing van groot belang kan zijn, aangezien de eerste 48 uur in een opsporingsonderzoek cruciaal zijn, zowel voor het opsporen van daders als het verzamelen van forensisch bewijsmateriaal.[75] Tegelijkertijd is er een risico op bias: soms vinden digitale forensische experts meer of minder bewijs op gegevensdragers afhankelijk van beschikbare contextuele informatie.[76]

3.4 *Waardering van bewijs*

Strafrechtelijk bewijs komt voor in verschillende soorten en maten. Technisch bewijs, zoals DNA-sporen, vingerafdrukken, ballistisch onderzoek of kras-, indruk- en vormsporen, gaat vrijwel altijd gepaard met bepaalde betrouwbaarheidsmarges. Die kunnen weer leiden tot foutpositieve of foutnegatieve testresultaten, bijvoorbeeld wanneer DNA-sporen op een plaats-delict gematcht worden met het DNA van een verdachte. Met behulp van AI is zogeheten *probabilistic genotyping* mogelijk om in te schatten of iemands DNA daadwerkelijk zit in gemengde DNA-sporen die zijn gevonden.[77]

In veel gevallen gaat het niet om de kans zelf (bijvoorbeeld een kans van 95%), maar de betrouwbaarheid ervan (bijvoorbeeld een foutmarge van 3%, waarmee de een kans eigenlijk ergens tussen de 92-98% ligt). Met behulp van zeer grote aantallen gegevens en zelflerende systemen kunnen betrouwbaarheden soms nauwkeuriger worden ingeschat (bijvoorbeeld door inachtneming van specifieke omstandigheden), waarmee foutmarges verkleinen. Zo kan de betrouwbaarheid van bewijsmateriaal scherper worden gekwantificeerd, met kleinere foutmarges, waardoor de betrouwbaarheid van forensisch bewijs uiteindelijk wordt vergroot.[78]

3.5 *Scenariovorming*

Uit de gedragspsychologie is bekend dat mensen tamelijk slecht zijn in het inschatten van kansen en risico's: vaak is een narratief overtuigender dan statistiek, omdat causale verbanden nu eenmaal evolutionair gezien sneller cognitief worden opgepakt.[79] Mensen blijken echter een stuk beter kansen in te kunnen schatten als ze scenario's naast elkaar krijgen gepresenteerd. AI kan bijdragen aan het construeren van verschillende scenario's die in de rechtbank tegen elkaar kunnen worden afgewogen.[80] Dit kan rechters helpen: in plaats van het afwegen van individuele scenario's kunnen zo ook scenario's met elkaar worden vergeleken. Daarbij wordt aan het beschikbare bewijsmateriaal per scenario in verschillende weging meegegeven. Die weging kan ook nul zijn, als bepaald bewijs in één van de scenario's wordt uitgesloten, bijvoorbeeld wanneer het onrechtmatig is verkregen. Door verschillende scenario's en hun bewijsrechtelijke onderbouwing verder te visualiseren (zie het voorbeeld in figuur 5), wordt bovendien inzichtelijk welke onderdelen van een scenario verdere onderbouwing behoeven. In figuur 5 zijn de witte blokken stappen in het narratief voor de waarheidsvinding. Na de eerste stap zijn er twee mogelijke verhaallijnen, elk onderbouwd met eigen bewijs. De rechter kan zo sneller zien welk bewijs bijdraagt aan welk onderdeel van de verhaallijn, maar bijvoorbeeld ook waar onderbouwing van een verhaallijn ontbreekt (in figuur 5 is elk element overigens onderbouwd met bewijs).

72 B. Hoelz, C. Ralha & R. Geeverghese, 'Artificial intelligence applied to computer forensics', *Proceedings of the ACM Symposium on Applied Computing. Honolulu*, 9-12 March 2009, p. 883-888.

73 https://www.forensischinstituut.nl/forensisch-onderzoek/hansken. Zie ook H.M.A. van Beek, E.J. van Eijk, R.B. van Baar, M. Ugen, J.N.C. Bodde & A.J. Siemelink, 'Digital Forensics As a Service: Game On', *Digital Investigation*, 2015, Vol. 15, p. 20-38.

74 N. Sunde & I. Dror, 'A Hierarchy of Expert Performance (HEP) applied to Digital Forensics: Reliability and Biasability in Digital Forensics Decision Making', *Forensic Science International: Digital Investigation* 2021, Vol. 37: https://doi.org/10.1016/j.fsidi.2021.301175.

75 B.W. Schermer & J.J. Oerlemans, 'AI, Strafrecht en het recht op een eerlijk proces', *Computerrecht* 2020/3, p. 14-21.

76 L. Geddes, 'Digital forensics experts prone to bias, study shows', *The Guardian* 31 May 2021: https://www.theguardian.com/science/2021/may/31/digital-forensics-experts-prone-to-bias-study-shows.

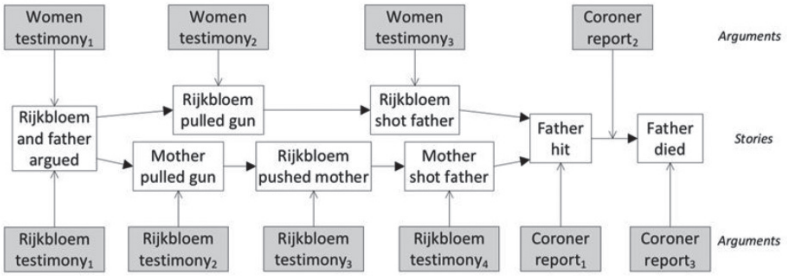
77 K. Kwong, 'The Algorithm Says You Did it: The Use of Black Box Algorithms to Analyse Complex DNA Evidence', *Harvard Journal of Law & Technology* 2017, Vol., 31, Nr. 1. P. 275-301.

78 M. Kwan, K.P. Chow, F. Law & P. Lai, 'Reasoning About Evidence Using Bayesian Networks', in: Ray I., Shenoi S. (eds.), '*Advances in Digital Forensics IV*', IFIP – The International Federation for Information Processing. Heidelberg: Springer 2008.

79 D. Kahnemann, '*Thinking, fast and slow*', New York: Penguin Books 2012.

80 F.J. Bex, B. Testerink & J. Peters, 'AI for Online Criminal Complaints: From Natural Dialogues to Structured Scenarios', *ECAI 2016 workshop on Artificial Intelligence for Justice (AI4J)*, Den Haag, augustus 2016, p. 22-29; M. Schraagen, B. Testerink, D. Odekerken & F. Bex, 'Argumentation-driven information extraction for online crime reports', *CKIM 2018 International Workshop on Legal Data Analysis and Mining (LeDAM 2018), CEUR Workshop Proceedings*.

Figuur 5. Met behulp van AI-technologie kunnen verschillende scenario's worden geconstrueerd door de weging van bewijs te variëren[81]



AI kan niet alleen bijdragen aan het (af)wegen van scenario's, maar ook zelfstandig nieuwe, wellicht onverwachte scenario's opstellen aan de hand van het beschikbare bewijs. Als in een rechtszaal de partijen tegenover elkaar staan met elk hun eigen scenario, kan de AI ook een derde of vierde scenario genereren, hetgeen mogelijke impasses kan doorbreken en de waarheidsvinding ten goede kan komen. Bij meerdere scenario's kan bovendien de overlap worden beschouwd: een stuk van een verhaallijn dat in elk scenario hetzelfde is, kan wellicht met meer zekerheid worden vastgesteld dan een stuk van een verhaallijn dat in elk scenario anders verloopt.

3.6 *Rechtsvragen*

De strafprocesrechtelijke voorbeelden in deze paragraaf geven aanleiding tot verschillende vragen. Onjuiste en onvolledige gegevens, de keuze van instrumenten voor data-analyse en de interpretatie van ontdekte profielen, kunnen leiden tot beperkte betrouwbaarheid van conclusies, waardoor vooroordelen en discriminatie in het proces van opsporing, vervolging en berechting kunnen sluipen. Dat roept ethische en rechtsfilosofische vragen op over rechtvaardigheid en, in het bijzonder, over een eerlijk proces. Omdat AI complex is en de werking ervan vaak lastig uit te leggen is, kan het voor verdachten moeilijk zijn zich hiertegen te verweren. Als beslissingen in het strafproces sterk(er) worden gebaseerd op AI-resultaten, kan dat potentieel leiden tot Kafkaëske situaties, waarin verdachten niet weten hoe het tot een beschuldiging is gekomen, op basis van welke gegevens, welke analyses en welke conclusies.

Transparantie lijkt een mogelijke oplossing en er wordt volop gewerkt aan *explainable AI* (XAI).[82] Daarbij zou inzichtelijk moeten zijn voor alle procespartijen hoe de AI tot conclusies komt. Echter, soms wil de politie bepaalde onderzoekstechnieken niet graag prijsgeven, omdat ze daarna niet meer kunnen worden ingezet. Hier rijst de vraag in hoeverre de politie de gebruikte technieken moet blootgeven, dan wel ze geheim mogen blijven. Met andere woorden, de vraag is dan hoe ver het recht op tegenspraak reikt.

Meer rechtspositivistische vraagstukken betreffen de reikwijdte van (bijzondere) opsporingsbevoegdheden. De voorbeelden van *predictive policing* en *cyber agent technology* roepen vragen op hoe ver bevoegdheden reiken in de nieuwe context van AI. Bij *cyber agent technology* komen daar ook vragen bij over hoe uitlokking wordt voorkomen en hoe wordt gegarandeerd dat zelflerende AI in een criminele context niet zelf crimineel gedrag gaat vertonen. Naast het duiden van de reikwijdte van bestaande bevoegdheden, is de vraag of deze bevoegdheden voldoende zijn voor de opsporing in deze veranderende context. Dat is overigens niet meteen een pleidooi voor meer bevoegdheden: wellicht kunnen aanpassingen in bestaande bevoegdheden al lacunes invullen.

Een ander punt is het reguleren van data-analyses in het strafrecht, online of nadat gegevensdragers in beslag zijn genomen. Het is opvallend dat enerzijds het verzamelen van gegevens zeer sterk is gereguleerd in het strafrecht (inclusief het gegevensbeschermingsrecht), terwijl data-analyse nauwelijks is gereguleerd.[83] Met andere woorden, zodra gegevens eenmaal zijn verzameld en bijeengebracht, is er veel vrijheid voor politie en justitie om uiteenlopende analyses daarop toe te passen. Regulering hiervan zou de rechtsbescherming van actoren in het strafproces (niet alleen verdachten) betere rechtsbescherming kunnen bieden, onder meer door meer inzicht en meer inspraak, en de rechtszekerheid kunnen vergroten.

81 Bron: F.J. Bex, 'An integrated theory of causal scenarios and evidential arguments', in *Proceedings of the 15th International Conference on Artificial Intelligence and Law (ICAIL 2015)*, New York: ACM Press, p. 13-22.

82 D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf & G.Z. Yang, 'XAI – Explainable artificial intelligence', *Science Robotics* 2019, Vol. 4, Nr. 37, DOI: 10.1126/scirobotics.aay7120.

83 B.H.M. Custers & L. Stevens, 'The Use of Data as Evidence in Dutch Criminal Courts', *European Journal of Crime, Criminal Law and Criminal Justice* 2021, Vol. 29, No, 1; B.W. Schermer, 'Het gebruik van Big Data voor opsporingsdoeleinden: tussen Strafvordering en Wet politiegegevens', *Tijdschrift voor Bijzonder Strafrecht en handhaving* 2017(4).

4. Conclusies

Het doel van deze bijdrage was een kort overzicht te bieden van verschillende AI-ontwikkelingen in het strafrecht. De voorbeelden laten zien dat AI in toenemende mate wordt gebruikt door criminelen, maar ook door politie en justitie. Gesteld kan worden dat dit kat-en-muisspel met de introductie van AI weer een nieuwe fase is ingegaan.[84] Om bij te blijven, zullen politie en justitie de komende tijd nog stevig moeten investeren in kennis en expertise.[85]

Vanuit het materieel strafrecht is verder onderzoek nodig naar de interpretatie en reikwijdte van bestaande strafbepalingen en naar de strafwaardigheid van bepaalde gedragingen die mogelijk worden gemaakt door deze nieuwe ontwikkelingen.[86] Vanuit het formeel strafrecht is verder onderzoek nodig naar de reikwijdte van bestaande opsporingsbevoegdheden, mogelijke aanpassingen daarin en de juiste balans tussen opsporingsbevoegdheden en grondrechten.[87] Het gebruik van AI kan allerlei voordelen bieden in de opsporing, maar alleen als vooroordelen, discriminatie en andere risico's worden vermeden. De regulering van data-analyses in opsporingsonderzoek, tot dusver nagenoeg afwezig, zou daaraan kunnen bijdragen.

84 Vergelijkbaar met andere technologieën die eerder in het veiligheidsdomein werden geïntroduceerd: B. Custers, B. Dorbeck-Jung, E. Faber, S. Iacob, B.J. Koops, R. Leenes, H. de Poot, A. Rip, W.B. Teeuw (ed.), A. Vedder (ed.) & J. Vudisa, *'Security Applications for Converging Technologies; impact on the constitutional state and the legal order'*, WODC rapport, Telematica Instituut, Enschede en Universiteit van Tilburg 2008.

85 Kennisontwikkeling vindt plaats bij onder meer het Kenniscentrum Cybercrime bij het Gerechtshof Den Haag (https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Gerechtshoven/Gerechtshof-Den-Haag/Over-het-gerechtshof/Organisatie/Paginas/Kenniscentrum-Cybercrime.aspx) en de High Tech Crime Unit van de politie (https://kombijde.politie.nl/vakgebieden/ict/cybercrime-aanpakken).

86 In samenhang met criminologisch onderzoek naar modi operandi, daderkarakterisitieken, enz. Zie A.M. Bossler & T. Berenblum, 'New directions in cybercrime research', *Journal of Crime and Justice* 2019, Vol. 42, No. 5, p. 495-499.

87 B.H.M. Custers, 'Het recht van de toekomst', oratie, Universiteit Leiden, 21 mei 2021.