Dit is het 'author accepted manuscript' van een gepubliceerd boekhoofdstuk. Gelieve hier als volgt naar te refereren:
van Beek, H., Henseler, H., Mieremet, A., & Snaphaan, T. (2024). Digital Forensics: Van beeldmateriaal naar bewijs. In M. Walrave, C. Van de Heyning, J. Janssen & E. Kolthoff (Eds.), *Misbruik van beelden* (Cahiers Politiestudies, nr. 70, pp. 129–150). Gompel&Svacina.

Cahiers Politiestudies nr. 70 'Politie en misbruik van beelden'

Digital Forensics: Van beeldmateriaal naar bewijs

Harm van Beek[1], Hans Henseler[2], Arjan Mieremet[3] & Thom Snaphaan[4]

Abstract

In deze bijdrage wordt ingegaan op beeldmateriaal als bron van bewijs in het kader van strafrechtelijke onderzoeken. In het bijzonder zal de rol van *digital forensics* oftewel digitaal-forensisch onderzoek uiteengezet worden. Digitaal-forensisch onderzoek richt zich op het veiligstellen, toegankelijk maken, inzichtelijk maken en duiden in de specifieke context van een strafzaak van verschillende vormen van data afkomstig van digitale bronnen. Nadat ruimere ontwikkelingen zijn omschreven die het digitaal-forensisch onderzoek onmisbaar maken in de context van strafrechtelijke onderzoeken, zal uiteengezet worden wat exact onder digitaal-forensisch onderzoek wordt verstaan. Vervolgens wordt toegelicht hoe het concept van 'Digital Forensics as a Service' in Nederland wordt ingevuld. Ook wordt ingegaan op het toenemend gebruik van artificiële intelligentie in het digitaal-forensisch onderzoek. In de discussie wordt tot slot op deze en andere actuele ontwikkelingen gereflecteerd, en worden aanbevelingen gedaan over hoe hier in de praktijk mee omgegaan kan worden.

1 Inleiding

Als gevolg van de toenemende digitalisering van onze samenleving, worden (of in ieder geval kunnen) steeds meer aspecten van ons dagelijks leven digitaal vastgelegd (worden). Dit brengt kansen en risico's met zich mee. Op het vlak van criminaliteit zien we bijvoorbeeld een sterke toename van cybercriminaliteit, terwijl andere (fysieke) vormen van criminaliteit al jaar na jaar afnemen (Centraal Bureau voor de Statistiek, 2022). Aan de andere kant levert deze ontwikkeling voor het politiewerk bijvoorbeeld wel meer aanknopingspunten bij de beeldvorming over criminaliteit, zowel op fenomeenniveau als in een specifiek opsporingsonderzoek. Deze bijdrage gaat in op de laatstgenoemde ontwikkeling, en – in het

[1] Senior digitaal-forensisch onderzoeker, Nederlands Forensisch Instituut, Ministerie van Justitie & Veiligheid, Nederland.
[2] Senior digitaal-forensisch adviseur, Nederlands Forensisch Instituut, Ministerie van Justitie & Veiligheid, Nederland en lector Digital Forensics & E-Discovery, Hogeschool Leiden, Nederland.
[3] Senior forensisch deskundige beeldonderzoek, Nederlands Forensisch Instituut, Ministerie van Justitie & Veiligheid, Nederland.
[4] Senior onderzoeker, Centre of Expertise Veiligheid & Veerkracht (V&V), Avans Hogeschool, Nederland, en postdoctoraal onderzoeker en docent, Institute for International Research on Criminal Policy (IRCP), Universiteit Gent, België. Zijn werk aan de Universiteit Gent wordt gedeeltelijk gefinancierd door de Europese Unie in het kader van het BIGDATPOLproject (ERC, BIGDATPOL, 101088156).

licht van het centrale onderwerp van dit themanummer – in het bijzonder in relatie tot beeldmateriaal.

Hoofdzakelijk op basis van de praktijkervaring van de auteurs wordt invulling gegeven aan deze bijdrage. De bijdrage heeft daarom ook in het bijzonder betrekking op digitaal-forensisch onderzoek bij het Nederlands Forensisch Instituut (NFI). Er wordt een overzicht geboden van het digitaal-forensisch onderzoek dat bijdraagt om van ruw beeldmateriaal naar bewijs te komen, dat gebruikt kan worden in bijvoorbeeld strafzaken. Aangezien het in het bestek van deze bijdrage aan de ruimte ontbreekt om volledig in de diepte te gaan op alle onderdelen, zullen op gepaste wijze ook suggesties om verder te lezen worden opgenomen. De onderzoeksvragen die centraal staan in deze bijdrage zijn:

(1) Welke digitaal-forensische methoden worden tegenwoordig toegepast in de Nederlandse praktijk om van ruwe beelddata naar bewijs te komen?
(2) Welke ontwikkelingen en uitdagingen kunnen geïdentificeerd worden op het vlak van deze digitaal-forensische methoden in de Nederlandse praktijk?

In de volgende paragraaf zal allereerst worden ingegaan op de ruimere ontwikkelingen die bekendstaan als de vierde industriële revolutie en het dientengevolge aantredende 'big-datatijdperk'. In de paragraaf daarna zetten we uiteen wat we onder beeldmateriaal verstaan en hoe dit kadert in het bredere palet van databronnen en (digitaal) bewijs. Daarna gaan we in op wat forensisch onderzoek gerelateerd aan beeldmateriaal precies inhoudt. Hier beschrijven we achtereenvolgens wat forensisch beeldonderzoek en biometrie is, en wat forensische data-analyse van die gegevens inhoudt. Hierna verbreden we de scope opnieuw, door toe te lichten hoe digitaal-forensische onderzoek op grote schaal kan worden ondersteund door een centraal data-analyseplatform ('Digital Forensics as a Service' genoemd, kortweg DFaaS) en hoe dit in Nederland in de praktijk vorm krijgt. Ook zullen we hierbij aandacht hebben voor de rol van (de grootschalige toepassing van) artificiële intelligentie in het onderzoek naar beeldmateriaal. Nadat we hebben beschreven hoe de huidige praktijk eruitziet, zullen we in de discussie ingaan op relevante ontwikkelingen die op ons afkomen en welke kansen en uitdagingen dit met zich meebrengt.

2 De vierde industriële revolutie en het big-datatijdperk

Door relatief recente (technologische) ontwikkelingen die elkaar in rap tempo opvolgen, is er vandaag de dag geen opsporingsonderzoek zonder digitale component. We bevinden ons in de zogeheten vierde industriële revolutie, waarin technologische doorbraken op het vlak van robotica, artificiële intelligentie, nanotechnologie, *quantum computing* en het *Internet of Things* centraal staan (Schwab, 2017). In de eerste twee industriële revoluties stonden analoge technologische ontwikkelingen centraal. De eerste industriële revolutie (van de 18de tot de 19de eeuw) wordt gekenmerkt door de opkomst van de ijzer- en textielindustrie en de opkomst van de stoommachine. De tweede industriële revolutie (1870-1914) wordt gekenmerkt door de ontwikkeling van staal, olie en elektriciteit. Deze technologische vooruitgang vindt zijn toepassing in bijvoorbeeld de telefoon, de gloeilamp en de interne verbrandingsmotor. De derde industriële revolutie (1980-heden) wordt ook wel de digitale revolutie genoemd. In dit tijdperk staan toepassingen als de computer, het internet en informatie- en communicatietechnologie

(ICT) centraal. Deze toepassingen hebben op hun beurt geleid tot vele nieuwe toepassingen of vernieuwingen in bestaande toepassingen in alle sectoren, van wetenschappelijk onderzoek tot industrie. Men kan stellen dat de eerste drie industriële revoluties de vierde industriële revolutie mogelijk hebben gemaakt. Deze vierde revolutie bouwt vooral voort op de digitale revolutie, met nieuwe manieren om technologie te verbinden, te implementeren en te gebruiken. Individuen vormen een belangrijke groep binnen de markt voor deze technologieën en genereren een ongekende datastroom. Deze datastromen worden zowel bewust als onbewust gegenereerd, zowel door de apparaten zelf als door de interacties die individuen via de apparaten met elkaar hebben (Henseler, 2017).

In dit verband spreekt men ook wel van het big-datatijdperk (*big data era*; Arbia, 2021; Snaphaan & Hardyns, 2021a). Dit werkt door op alle facetten van de maatschappij, zo ook op het forensisch werk (forensische big-data-analyse; Nederlands Forensisch Instituut [NFI], z.d.-a) en politiewerk (*big data policing*; zie Snaphaan et al., 2023). Wat big data precies omvatten, wordt op verschillende manieren ingevuld. Sommige auteurs definiëren big data als concept (zie bijv. Favaretto et al., 2020), andere auteurs beargumenteren dat dit eigenlijk ondoenbaar is, omdat de achterliggende databronnen zo divers zijn. Big data karakteriseren op basis van enkele veelvoorkomende kenmerken is dan wellicht vruchtbaarder. Het gaat hierbij om de 3 V's (De Mauro et al., 2015; Laney, 2001): grote hoeveelheden data (*Volume*), uit zeer diverse bronnen en in diverse formaten (*Variety*), die in of near-real time gemaakt en verwerkt worden (*Velocity*). Het big-datatijdperk betekent dat (haast) alles om ons heen direct of indirect verbonden is met een databron en dat bijna alles in ons leven digitaal wordt vastgelegd (Sarker, 2021) en deze vastlegging is de basis voor het digitaal-forensisch onderzoek. In de volgende paragraaf zullen we in het bijzonder ingaan op één type (big) data: beeldmateriaal.

3 Beeldmateriaal als datatype en bewijs

Data zijn het resultaat van het abstraheren van verschijnselen in de 'fysieke wereld' door deze weer te geven in categorieën, maten of andere representatieve vormen (cijfers, karakters, symbolen, afbeeldingen, geluiden, elektromagnetische golven, bits, et cetera; Kitchin, 2014). Rosenberg (2013) stelt dat er een belangrijk onderscheid gemaakt dient te worden tussen feiten, bewijs en data: een 'datum' (i.e., het enkelvoud van data) kan een feit zijn, net zoals een feit een bewijs kan zijn. Het bestaan van een datum staat los van enige overweging over de waarheid. Wanneer bewezen wordt dat een feit onjuist is, houdt het op een feit te zijn. Niettemin blijven valse data wel data. Dit geldt evenzeer voor beeldmateriaal: beelddata staan niet noodzakelijk gelijk aan bewijs.

Beeldmateriaal is een vorm van kwalitatieve, ongestructureerde data. Het gaat om kwalitatieve data, omdat het niet-numerieke data betreffen. Andere vormen van kwalitatieve data zijn tekstdata en audiodata (Palys & Atchison, 2013). Deze kwalitatieve data kunnen gekwantificeerd worden, maar deze omzetting vergt abstractie en reductie, waardoor mogelijk wordt ingeboet op de rijkheid van de data (Kitchin, 2014). Het zijn tevens ongestructureerde data omdat ze veelal geen enkele vorm van structurering en ook geen vooraf gedefinieerd formaat hebben. De hoeveelheid ongestructureerde data neemt alleen maar toe en het gros van

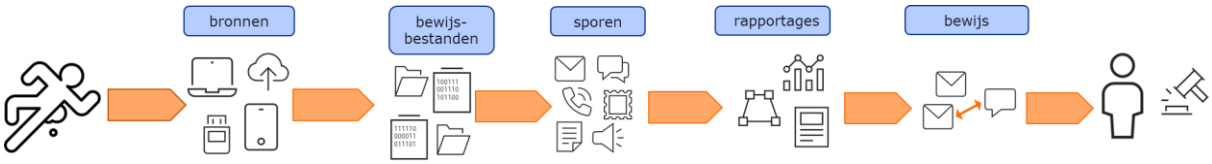
de bestaande data (naar schatting tussen de 80% en 95% van de totale hoeveelheid data, ook binnen bedrijven) is ongestructureerd (Gandomi & Haider, 2015; Schneider, 2016).

Beeldmateriaal betreft in feite een verzameling van verschillende soorten data, want het betreft zowel statische beelden (foto's) als dynamische beelden (video's). De inhoud van zulke beelden kan doorslaggevend bewijs leveren is strafzaken. Denk bijvoorbeeld aan camerabeelden van een dader of een delict. De computationele verwerking en analyse van beeldmateriaal staat ook wel bekend als *computer vision* (Marr, 1982; Szeliski, 2022), refererend aan de nabootsing van het menselijk zintuiglijk vermogen om te zien. Computer vision is een belangrijk interdisciplinair domein gerelateerd aan artificiële intelligentie. Tegenwoordig werken veel computer vision toepassingen op basis van zogeheten convolutionele neurale netwerken (ook wel CNN's of ConvNets) (Krizhevsky et al., 2012; LeCun et al., 2015). CNN's vormen een specifiek type van neurale netwerken, waarvan de eigenschappen bijzonder geschikt zijn voor de analyse van beeldmateriaal (zie Snaphaan & Hardyns, 2020). Naast de inhoud van de beelden, kan beeldmateriaal ook veel andere informatie bevatten. Dit gaat verder dan enkel wat er op het beeld met het menselijk oog waarneembaar is. De opgaven omtrent beeldmateriaal liggen met andere woorden niet uitsluitend op het op grote schaal verwerken van de data, maar ook op het ontginnen van *alle* informatie uit specifieke beelden. Hierbij kan gedacht worden aan enerzijds informatie die niet direct observeerbaar is, zoals bij QR-codes of steganografie (NFI, 2022), maar ook ruispatronen in beelden bevatten informatie (Julliand et al., 2016). Daarnaast bevatten beelddata metadata (i.e., data over de data), zoals het bestandtype, de datum dat het bestand gecreëerd is en de resolutie. In bepaalde gevallen zijn deze metadata gestandaardiseerd, zoals het *Exchangeable image file format* (kortweg Exif); een standaard voor metadataspecificatie van onder andere digitale camera's.

Data, dus ook digitale beelden, zijn overal. Vaak zijn ze vastgelegd op gegevensdragers (zoals harde schijven en geheugenchips) in digitale apparaten bij verdachten of slachtoffers (zoals computers en smartphones) of bij een serviceprovider (in geval van cloud-data). Data kunnen ook vluchtig zijn, wat betekent dat deze zijn vastgelegd in geheugen dat wordt gewist zodra de voeding wordt onderbroken. Tot slot kunnen data ook stromen, wat betekent dat deze 'onderweg' zijn tussen apparaten. Zulke data kunnen voor tactisch opsporingsonderzoek en/of bewijsdoeleinden (na een juridische toets) afgetapt worden. Zulk beeldmateriaal vormt namelijk een link tussen de fysieke en digitale wereld. Afhankelijk van het type geheugen, kunnen data veiliggesteld worden voor digitaal-forensisch onderzoek. Hierop zal in de volgende paragraaf nader worden ingegaan.

4 Digitaal-forensisch onderzoek

Digitaal-forensisch onderzoek richt zich op het veiligstellen en toegankelijk maken van de verschillende vormen van data afkomstig van digitale bronnen, op het inzichtelijk maken van die data en op het duiden van de data in de specifieke context van een strafzaak. In Figuur 1 is het proces van gebeurtenis, die gecapteerd wordt in data, naar het gebruik van bewijs opgenomen.



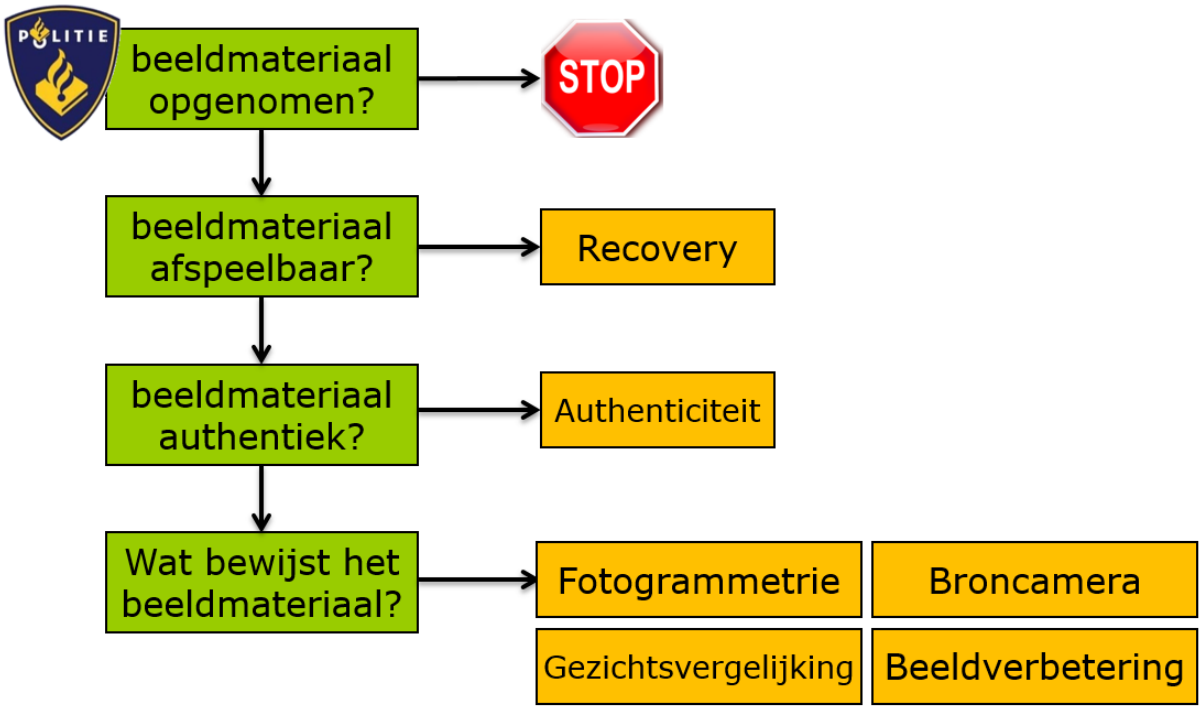
Figuur 1 Het proces van een gebeurtenis (gecapteerd in data) naar het gebruik van bewijs.

Voor digitaal-forensisch onderzoek begint dit met het vastleggen van veiliggestelde data uit bronnen in bewijsbestanden. Om de integriteit van de data in deze bewijsbestanden te garanderen, worden bij het bestand zogenaamde bestandskenmerken berekend (NFI, 2018), ook wel *digests* of *secure hashes* genoemd. Deze bestandskenmerken zijn reproduceerbaar en worden vaak in de bewijsbestanden opgenomen. In deze bijdrage gaan we niet nader in op het veiligstellen en toegankelijk maken van beelddata, omdat dit een vakgebied op zich is (zie Casino et al., 2022). Over het algemeen worden zulke data in eerste instantie gebruikt voor tactisch onderzoek. Een afbeelding kan bijvoorbeeld gemaakt zijn op een specifieke locatie of informatie bevatten die de verdachte koppelt aan een delict.

Bij ieder forensisch onderzoek is het van belang dat het onderzoek reproduceerbaar is en dat alle resultaten en conclusies onderbouwd en herleidbaar zijn (Robertson et al., 2016). Dit gebeurt door het vastleggen van alle handelingen (*chain of custody*). Daarnaast moet voor ieder (digitaal) bewijsstuk duidelijk zijn hoe het te herleiden is naar de bron (*chain of evidence*). In de volgende twee paragrafen bespreken we forensisch beeldonderzoek en forensische data-analyse. Het grote verschil tussen deze twee is dat in forensisch beeldonderzoek het onderzoek gericht is op de inhoud van de beeldbestanden terwijl bij forensische data-analyse de focus ligt op de (digitale) context waarin de beeldbestanden aangetroffen worden.

4.1 Forensisch beeldonderzoek

Forensisch beeldonderzoek is het vakgebied dat zich richt op de beantwoording van forensische vragen over digitaal en analoog beeldmateriaal. Dit vakgebied wordt in CSI-achtige series neergezet als het vakgebied waarin antwoorden op onderzoeksvragen verkregen worden door goed te kijken en even in te zoomen op het beeld en deze dan te verscherpen. In werkelijkheid is inzoomen en verscherpen meestal niet het antwoord op de onderzoeksvraag. Een overzicht van het forensisch beeldonderzoek is weergegeven in Figuur 2. Verschillende onderzoeksrichtingen binnen het vakgebied beeldonderzoek en hun onderlinge relatie zijn hierin weergegeven door het stellen van vier vragen; (1) Is er beeldmateriaal beschikbaar? (2) Is het beeldmateriaal afspeelbaar? (3) Is het beeldmateriaal authentiek? En tot slot (4) wat bewijst het beeldmateriaal?



Figuur 2 Een overzicht van de verschillende onderzoeksrichtingen binnen forensisch beeldonderzoek.

De eerste vraag is een vraag die typisch door de politie (bijv. door de forensische opsporing) gesteld wordt en als het goed is kort na het incident al beantwoord wordt. Dit houdt bijvoorbeeld in dat uitgezocht wordt of er bewakingscamera's aanwezig zijn of dat er personen aanwezig waren die opnamen maakten met hun smartphone. Logischerwijs, indien dit niet het geval is, zal er ook geen forensisch beeldonderzoek verricht hoeven te worden. Forensisch beeldonderzoek is wel nodig als er bijvoorbeeld vragen zijn over de inhoud van het beeldmateriaal omdat dit betwist wordt. Dit wordt onder andere uitgevoerd door het Nederlands Forensisch Instituut (NFI, z.d.-b).

Indien camera's opnamen gemaakt hebben, dan is logischerwijs de volgende vraag of deze opnamen ook afspeelbaar zijn (of te bekijken in het geval van foto's). In de praktijk komt het soms voor dat eigenaren van bewakingssystemen er belang bij hebben als bepaalde opnamen niet beschikbaar zijn voor het politieonderzoek en dus worden opnamen nog wel eens gewist. De onderzoeksrichting 'Recovery' houdt zich bezig met het herstellen en weer afspeelbaar maken van gewist beeldmateriaal. Aan een één-op-één-kopie van het opnamesysteem wordt dan onderzoek gedaan waarbij de datastructuur op bit- en byte-niveau geanalyseerd wordt om zo te doorgronden hoe het opnamesysteem de data wegschrijft. Vervolgens worden scripts geschreven waarmee het opslagmedium geanalyseerd wordt en de relevante delen worden herkend en worden weggeschreven naar nieuwe bestanden. Deze bestanden kunnen vervolgens afgespeeld worden en kunnen van groot belang zijn voor het beantwoorden van de vraag "wat is er gebeurd?".

Als beeldmateriaal afspeelbaar is, is de tweede vraag of het authentiek is. Meestal is deze vraag niet van belang omdat er geen enkele reden is om te geloven dat beeldmateriaal afkomstig van een bewakingssysteem gemanipuleerd zou zijn, maar dit is anders voor beeldmateriaal afkomstig van internet. De onderzoeksrichting 'Authenticiteit' houdt zich bezig met het

vaststellen of er aanpassingen hebben plaatsgevonden die de interpretatie van het beeldmateriaal beïnvloeden. Hierbij moet duidelijk onderscheid gemaakt worden tussen de integriteit en authenticiteit van beeldmateriaal. Bij integriteit gaat het erom of de data nog exact hetzelfde zijn als bij de oorspronkelijke opname. Bij authenticiteit gaat het erom of de beeldinhoud zodanig is aangepast dat het beeld niet langer weergeeft wat de werkelijke situatie was. Wanneer bijvoorbeeld een foto opnieuw wordt opgeslagen en dus opnieuw gecomprimeerd wordt, dan is de integriteit niet langer behouden maar de authenticiteit wel, want de beeldinhoud is niet zodanig aangepast dat je het anders interpreteert (tenzij er zodanig sterke compressie heeft plaatsgevonden dat het beeld substantieel is gewijzigd). Alhoewel authenticiteitsonderzoek al lang bestaat, is met de opkomst van neurale netwerken, grotere rekenkracht en grote databanken het maken van bijvoorbeeld deepfakes en volledig gesynthetiseerde beelden relatief gemakkelijk geworden.

In de praktijk geldt dat bij het overgrote deel van alle beeldonderzoeksvragen er geen afspeelproblemen zijn of twijfel over de authenticiteit. Veelal gaan de vragen over de inhoud van de beelden en wat dit bewijst. Hierbij dient men stil te staan bij twee belangrijke aspecten. Ten eerste dat voorzichtigheid geboden is bij de interpretatie van de beelden. Hoe duidelijk iets ook mag lijken, beelden kunnen verkeerd geïnterpreteerd worden. Er zijn hiervoor vele redenen. Hetzelfde object of dezelfde persoon kan in verschillende beelden totaal anders lijken door een andere aanzichthoek of belichting. Ook kunnen compressie-artefacten details verkeerd weergegeven of laten verdwijnen. Om een forensisch verantwoorde uitspraak te kunnen doen moeten daarom vaak referentie-opnamen gemaakt worden met het betreffende camerasysteem. Ten tweede bevat een beeld niet oneindig veel informatie. Een camera legt alleen vast wat het ziet en geen informatie van wat er achter een object staat of wat buiten het bereik van de camera gebeurt. Een beeld is uiteindelijk ook maar een 2D-weergave van een 3D-werkelijkheid. Bovendien wordt een groot deel van de informatie die wel waargenomen wordt ook nog ‘weggegooid’. Bedenk dat huidige camera’s typisch 25 beelden per seconde opnemen en dat elk beeld typisch tientallen miljoenen beeldpunten (pixels) bevat. Dan is het niet vreemd dat er ontzettend veel opslagruimte nodig is om alles te kunnen bewaren. Om dit te voorkomen, worden de beelden eerst gecomprimeerd. Bekende compressieformaten voor foto’s en afbeeldingen zijn JPG en PNG. Voor video’s is H.264, of diens opvolger H.265, een zeer gangbaar videocoderingsformaat. Deze compressieformaten bewaren een benadering (oftewel een vereenvoudigde weergave) van het oorspronkelijke beeld. Dit bespaart vaak al gauw een factor 10 tot 1000 aan opslagruimte. Dat je dit niet ziet, geeft aan hoe goed de hedendaagse compressieformaten zijn. Het is deze vereenvoudigde weergave waarop de onderzoeksvragen betrekking hebben. Dit is ook waar het in CSI-achtige series fout gaat. Daarin is het mogelijk om zonder problemen 100x zo niet 1000x in te zoomen op een beeld om vervolgens wederom een scherp beeld te verkrijgen. In werkelijkheid is deze informatie er dus niet (meer).

Ondanks dat we dus voorzichtig moeten zijn met de interpretatie van beelden en we er bewust van moeten zijn dat we niet oneindig veel informatie hebben, kan forensisch beeldonderzoek veel onderzoeksvragen beantwoorden. Deze vragen kunnen typisch in vier onderzoeksrichtingen verdeeld worden: fotogrammetrie, broncamera-onderzoek, gezichtsvergelijking en beeldverbetering. Deze onderzoeksrichtingen worden onderstaand achtereenvolgens toegelicht.

Fotogrammetrie is de wetenschap van het meten in beelden en kan worden toegepast om vragen over de lengte van daders, afmetingen en snelheden van objecten te beantwoorden. Omdat beelden zelf geen afstandsinformatie bevatten is het noodzakelijk om op de locatie de onderlinge afstanden van kenmerkende punten in de beelden vast te leggen, bijvoorbeeld door een 3D-laserscan. Op basis hiervan kan gemeten worden welke lengte of afstand (en dus snelheid) er is afgelegd. Dit is echter niet voldoende, want op dat moment is nog niet bekend hoe nauwkeurig de meting is. Daarom moeten op de locatie ter aanvulling referentie-opnamen gemaakt worden van personen van bekende lengte of ritten gemaakt worden met auto's met bekende snelheden. Ook deze beelden zullen geanalyseerd worden en op basis daarvan kan berekend worden hoe nauwkeurig de lengte of snelheid bepaald kan worden op basis van de beelden. Door het maken van referentieopnamen is het dus mogelijk niet alleen een beste schatter maar ook een nauwkeurigheid te rapporteren. In het bovenstaande is steeds aangenomen dat op de locatie nieuwe referentieopnamen gemaakt kunnen worden. Dit is natuurlijk niet altijd het geval. Camerasystemen worden verwijderd, locaties worden verbouwd, en dan is het niet meer mogelijk om referentieopnamen te maken waarvan de beelden technisch vergelijkbaar zijn met de te onderzoeken beelden. In zulke gevallen kan dus geen gevalideerd onderzoek worden uitgevoerd.

Het broncamera-onderzoek is gebaseerd op de aanwezigheid van imperfecties van de sensor van de camera. Tijdens de productie van deze sensor zijn er zeer kleine variaties tussen de pixels ontstaan waardoor de ene pixel nu eenmaal beter is in het omzetten van een lichtsignaal in een elektronisch signaal dan de andere pixel. Deze niet-uniformiteit is vervolgens aanwezig in elke foto die met die sensor gemaakt wordt, maar is voor elke sensor weer anders. Populair gezegd, elke sensor heeft een eigen vingerafdruk. Deze vingerafdruk staat bekend als het PRNU-patroon (Photo Response Non-Uniformity) en maakt het mogelijk om een camera-identificerend onderzoek of een gemeenschappelijke-bron onderzoek uit te voeren (Julliand et al., 2016). Bij een camera-identificerend onderzoek wordt onderzocht in welke mate het PRNU-patroon in de foto overeenkomt met het PRNU-patroon van een camera (bv. van verdachte). Dit type onderzoek kan extreem hoge bewijswaarden opleveren. In de praktijk wordt deze onderzoeksrichting vaak ingezet om aan te tonen dat verdachte de producent is van beeldmateriaal van seksueel (kinder)misbruik. Voor dit type onderzoek is het echter wel noodzakelijk dat de vermoedelijke camera beschikbaar is. Dit is echter niet altijd zo. Camera's kunnen verkocht zijn of niet teruggevonden zijn. In dat geval is een gemeenschappelijke-bron onderzoek nog een optie. Daarbij wordt een grote collectie foto's van verdachte onderzocht. Van elke foto wordt het PRNU-patroon zo goed mogelijk achterhaald en vervolgens worden deze PRNU-patronen geclusterd op basis van de mate van overeenkomst. Op deze wijze worden foto's gegroepeerd op basis van hun mogelijke bron.

Gezichtsvergelijking en objectvergelijking zijn gebaseerd op het vergelijken van vormkenmerken, zichtbaar in beeldopnamen van personen en objecten die onder zoveel mogelijk vergelijkbare omstandigheden zijn opgenomen. Alhoewel in de praktijk vaak over gezichtsvergelijking gesproken wordt, beperkt het forensisch vergelijkend-onderzoek zich niet alleen tot het vergelijken van gezichten. Ook andere biometrische kenmerken zoals handen en voeten worden regelmatig onderzocht. Daarbij moet niet vergeten worden dat ook andere uiterlijke kenmerken, zoals tatoeages, meegenomen kunnen worden. Vaak worden beelden van

pinautomaten gebruikt voor gezichtsvergelijkend onderzoek. Om ‘herkenning’ te voorkomen dekt de dader daarom zijn of haar gezicht vaak af met bijvoorbeeld een zonnebril of ten tijde van de coronapandemie met een mondkapje. Vaak zijn in dit type beelden ook de handen met eventuele tatoeages in beeld die dan zeer goed bruikbaar zijn voor vergelijkend onderzoek.

Als laatste onderzoeksrichting is er dan beeldverbetering. Zoals eerder aangegeven is het vaak niet mogelijk om een beeld eenvoudig scherper te maken om de simpele reden dat er niet meer informatie aanwezig is in de beelden. Dit wil niet zeggen dat beeldverbetering nooit mogelijk is. Beeldverbetering is een subjectief begrip; wat de een een verbetering vindt, is dat voor een ander niet. Beeldverbetering focust daarom op het vergemakkelijken van de interpretatie van beelden (en niet op het doen van de interpretatie zelf). Een illustratief voorbeeld van wat wel mogelijk is met beeldverbetering. Tegenwoordig wordt veel beeldmateriaal opgenomen met smartphones waarbij de camera continu in beweging is. Bij het bekijken van dit soort beelden is de kijker onbewust vooral bezig met het volgen van de camerabeweging waardoor de focus op het interpreteren minder is. Door beeldstabilisatie wordt het kijken naar deze beelden gemakkelijker en kan de kijker zich meer focussen op wat er echt te zien is in de beelden. Een ander voorbeeld is dat door de grote toename van het aantal bewakingscamera’s een incident in sommige gevallen waar te nemen is vanuit verschillende standpunten. De onderlinge relatie tussen de beelden in tijd is in zulke gevallen niet altijd duidelijk. Een meerluik van beelden die gesynchroniseerd zijn kan de onderlinge relatie tussen verschillende camerabeelden verduidelijken waardoor oorzaak en gevolg niet langer ter discussie staan.

Zoals uit bovenstaande blijkt is forensisch beeldonderzoek meer dan alleen maar het inzoomen en het verscherpen van beelden. Dit onderzoeksgebied vereist onder andere algemene kennis over het uitvoeren van goed forensisch onderzoek, het hebben van diepgaande digitale kennis op bit- en byte-niveau voor het herstellen van gewiste beelden, het hebben van diepgaande kennis over beeldvorming door camerasystemen, experimentele ervaring met het maken van referentieopnamen op locaties, kennis over beeldartefacten en nog vele andere aspecten. Alleen als al deze kennis aanwezig is kan een forensisch beeldonderzoek een betrouwbaar resultaat opleveren.

4.2 Digitaal-forensische data-analyse

Digitaal-forensische data-analyse definiëren we als het vakgebied dat zich richt op de beantwoording van forensische vragen over digitale sporen. Dat zijn vaak vragen over de herkomst van specifiek beeldmateriaal (zoals: hoe komt deze video op deze telefoon?) of vragen naar het gebruik ervan (zoals: is deze foto door de verdachte geopend? Of: heeft de verdachte deze afbeelding bewust gedownload?). Dergelijk onderzoek vindt typisch plaats na vragen van een tactisch rechercheur of analist van een opsporingsdienst die een digitaal expert vraagt voor duiding. Maar deze vragen zijn vaak ook afkomstig van een officier van justitie, advocaat-generaal, rechter(-commissaris), raadsheer(-commissaris) of advocaat van de verdachte. Bij de beantwoording van zulke vragen wordt doorgaans niet gekeken naar wat er op een beeld staat (dat is al relevant geacht door het opsporingsteam), maar naar de metadata van het te onderzoeken materiaal en de context waarbinnen het is aangetroffen.

Een groot deel van de forensische data-analyses wordt gedaan met behulp van (forensische) tools. Deze tools verwerken de data en maken de metadata inzichtelijk. Veel tools zijn toegespitst op één specifieke soort data of een groep vergelijkbare data (bijv. alleen afbeeldingen of alleen videomateriaal). Deze data zijn deels door mensen gemaakt (waaronder beeldmateriaal zoals foto's en video's, maar ook bijv. e-mails, chatberichten en documenten), deels zijn het sensor-registraties (bijv. locatiegegevens of getelde stappen, maar ook opnames van beveiligingscamera's) en deels technische registraties (bijv. logbestanden en systeeminstellingen). Deze gegevens kunnen ook in combinaties voorkomen, zoals een afbeelding waarin in de Exif-metadata locatiegegevens zijn vastgelegd. Al deze digitale artefacten noemen we *digitale sporen*.

Het kan zo zijn dat bestanden (waaronder beeldmateriaal) al dan niet opzettelijk zijn verwijderd. Met speciale technieken kunnen zulke verwijderde bestanden soms teruggehaald worden. Deze techniek heet 'carven' en maakt gebruik van de interne structuren in de data die het beeldmateriaal coderen (Van Bree & Schramp, 2019). Afhankelijk van hoe deze bestanden zijn verwijderd, kunnen ze wel of niet dienen als bewijsmateriaal in strafzaken (van den Hurk et al., 2023).

Beantwoording van forensische vragen over beeldmateriaal vindt in het algemeen plaats door analyse van sporen die met het beeldmateriaal in verband staan. Voor de herkomstbepaling van bijvoorbeeld een kindermisbruikafbeelding is het van belang wat de naam is van het bestand en waar deze afbeelding staat (op een USB-stick, op het bureaublad, in de map met downloads, of in de cameramap van de smartphone van de verdachte). Voor een goede duiding zijn niet alleen deze metadata van het spoor zelf, maar ook onderzoek aan verwante sporen essentieel. Zo laat het plaatsen van een USB-stick in een computer meestal sporen achter in logbestanden en configuratiebestanden in computers. Onderzoek aan logbestanden in combinatie met tactische informatie kan inzicht geven wie welke gebruikersnaam op een computer gebruikt. Ook om te achterhalen welke map gebruikt wordt voor het opslaan van downloads, worden instellingen van het apparaat geanalyseerd. Als de afbeelding op de smartphone staat, kan met het eerder genoemde broncamera-onderzoek worden onderzocht of de afbeelding gemaakt is met de camera in de smartphone.

Net zoals beelden, vormen tijdsregistraties een link tussen de fysieke en de digitale wereld. Tijdsregistraties zijn dan ook regelmatig relevant in een forensische data-analyse, bijvoorbeeld om vast te stellen of een afbeelding is gemaakt vóór of ná het delict. Dan vindt onderzoek plaats naar de geregistreerde tijden in de metadata van de afbeelding, maar ook bij sporen die op en rond hetzelfde moment zijn geregistreerd. Tevens wordt onderzoek gedaan naar de betrouwbaarheid van deze registraties, zoals naar welke klok is gebruikt bij het vaststellen van de tijd, of die klok goed liep en of de tijdzone en zomer-/wintertijd goed waren ingesteld op het moment van het registreren van de tijd. Al deze informatie is relevant voor het goed kunnen beantwoorden van vragen over de herkomst en het gebruik van beeldmateriaal. Dit is slechts een greep uit de veelheid aan mogelijke onderzoekstoepassingen. Casino et al. (2022) hebben recentelijk een goed overzicht gepresenteerd van alle trends, uitdagingen en opkomende technieken in het digitaal-forensisch domein.

In het algemeen is dan ook niet met honderd procent zekerheid vast te stellen wat er is gebeurd. Er zijn altijd scenario's die naar hetzelfde sporenbeeld leiden. Daarom is het wenselijk dat de sporen worden geanalyseerd in het licht van meerdere scenario's. De forensisch deskundige rapporteert dan bij welk van de gegeven scenario's de sporen het beste passen (Aitken & Taroni, 2004; Kokshoorn et al., 2020).

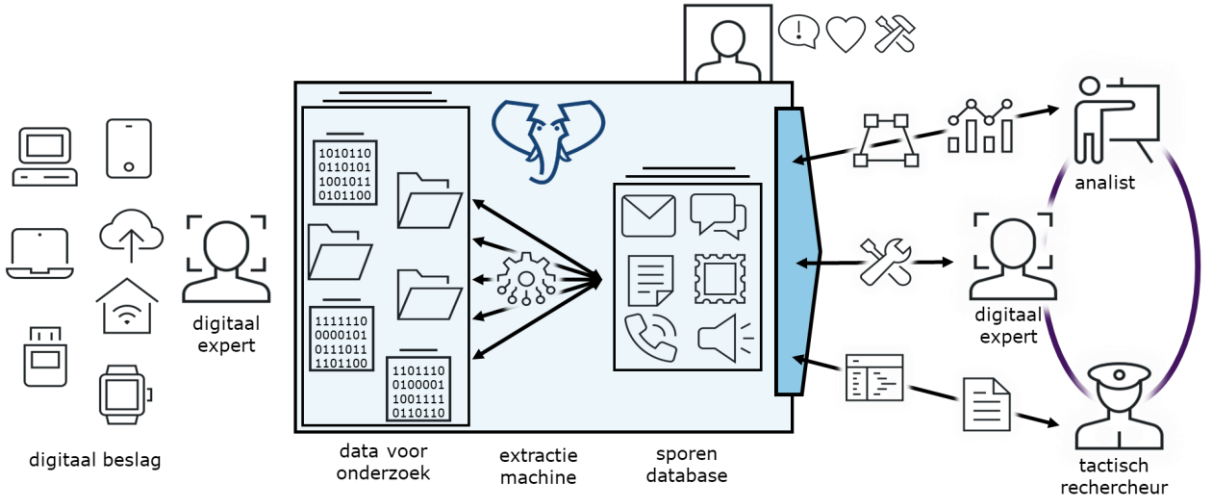
5 Een centraal platform voor digitaal-forensische data-analyse van beeldmateriaal

Zoals uit de eerdergenoemde voorbeelden blijkt, is voor een digitaal-forensisch onderzoek met betrekking tot beeldmateriaal een veel breder onderzoek nodig dan enkel gericht op de beelden zelf. In deze paragraaf leggen we eerst uit hoe een centraal platform kan helpen bij de forensische analyse van digitale sporen door digitaal experts in samenwerking met rechercheurs en analisten. Vervolgens leggen we uit hoe de grootschalige inzet van artificiële intelligentie de gebruikers kan ondersteunen bij hun onderzoek maar dat daar ook risico's aan verbonden zijn.

5.1 Digital Forensics as a Service

Het proces van het verwerken van data in bewijsbestanden, het identificeren van potentieel relevant beeldmateriaal en daarover forensisch rapporteren, gaat over veel stappen, waarbij meerdere personen in verschillende rollen betrokken zijn (Reedy, 2023). Om dit complex samenspel van deze betrokkenen goed te kunnen faciliteren, is het concept 'Digital Forensics as a Service' (kortweg DFaaS; Van Baar et al., 2014; Van Beek et al., 2015) geïntroduceerd. DFaaS is een centrale oplossing, waarbij alle betrokkenen in een onderzoek gebruikmaken van hetzelfde platform.

DFaaS volgde daarmee in 2014 het voorbeeld van de web-based review platformen die destijds al een geruime tijd door advocatuur en forensische accountants werden ingezet in grootschalige onderzoeken van, onder andere, toezichthouders. In dergelijke 'E-Discovery' onderzoeken (Henseler, 2010) ligt de nadruk meer op email- en documentenonderzoek terwijl in een digitaal-forensisch onderzoek veel meer soorten sporen onderzocht worden. Denk daarbij aan locatiesporen, gewiste bestanden, internetgeschiedenis, apps op mobiele telefoons, drones, auto's, et cetera. Makers van commerciële tools voor digitaal-forensisch onderzoek slaagden er destijds maar beperkt of helemaal niet in om hun traditionele tools (zoals Encase, Cellebrite Physical Analyser, Forensic Toolkit en Internet Evidence Finder) geschikt te maken voor een DFaaS-oplossing. Het lukte ze niet om grootschalige datasets te verwerken en ook niet om een web-based user interface te maken die geschikt is voor rechercheurs en analisten. Dit heeft ertoe geleid dat in Nederland meerdere overheidsdiensten zijn gaan samenwerken met het NFI aan een implementatie van het DFaaS-concept, genaamd Hansken (Hansken Community, z.d.). In Figuur 3 is Hansken, als DFaaS-implementatie, centraal weergegeven in het blauw.



Figuur 3 Hansken als Digital Forensics as a Service (DFaaS) implementatie.

Digitaal experts die data uit in beslag genomen apparaten hebben veiliggesteld, kunnen deze in het DFaaS-platform plaatsen. Op het platform kunnen verschillende tools specifieke verwerkingen van de data uitvoeren, zoals bestanden identificeren in de data (waaronder beeldmateriaal), Exif-informatie inzichtelijk maken, bestandskenmerken berekenen, bestandskenmerken matchen met bestaande databases (bijv. om afbeeldingen en video's met seksueel kindermisbruik te identificeren en labelen; Laureys, Snaphaan & Hardyns, 2021), objecten herkennen in beeldmateriaal, tekst in beelden herkennen, enzovoorts. De resultaten van al deze tools worden samengebracht in een gestructureerd sporenmodel en gepresenteerd aan tactisch rechercheurs die een strafzaak onderzoeken, aan analisten die zaaksoverstijgende fenomenen onderzoeken en aan digitaal experts, die hun collega's kunnen helpen bij geavanceerde zoekslagen en begrip van gevonden sporen.

Het sporenmodel (Van Beek et al., 2015) forceert dat tools hun resultaten op een uniforme manier beschikbaar stellen. Hierdoor kunnen rechercheurs eenvoudig zoeken naar bijvoorbeeld afbeeldingen, onafhankelijk van waar ze zijn aangetroffen. Zowel een foto ontvangen als bijlage aan een e-mail die is aangetroffen in een e-maildatabase op een computer als een screenshot die in een bestandsarchief (bijv. zip-bestand) in de downloadsmap in een mobiele telefoon staat, wordt op dezelfde manier gepresenteerd. Voor de eerdergenoemde *chain of evidence* en *chain of custody* is het natuurlijk wel van belang dat bij ieder beeldmateriaal expliciet is vastgelegd waar deze precies in welk bewijsbestand staat en met welke tools in welke volgorde de data verwerkt is.

Digital Forensics as a Service wordt al sinds 2010 operationeel ingezet en is inmiddels gebruikt in duizenden strafzaken. Twee belangrijke ontwikkelingen die betrekking hebben op het gebruik van het Digital Forensics as a Service platform Hansken zijn van meer juridische aard. Een centrale oplossing voor de verwerking van digitaal bewijsmateriaal maakt het namelijk mogelijk om verschillende juridische afspraken in het platform te borgen.

De eerste ontwikkeling die we zien, is de ondersteuning van het filteren van digitale sporen met verschoningsgerechtigde informatie. Dit betreft bijvoorbeeld communicatie tussen de verdachte en de advocaat die heeft plaatsgevonden met een in beslag genomen apparaat en dientengevolge

voorkomt in de data die is veiliggesteld uit dat apparaat. Over de manier waarop deze data gefilterd of zelfs vernietigd moeten worden, lopen verschillende rechtszaken (Doorenbos & Rosing, 2020; Gerechtshof 's-Hertogenbosch, 2023).

Een tweede belangrijke ontwikkeling die we zien als het gaat om digitaal bewijs in het algemeen en dus ook voor digitaal beeldmateriaal, is dat de advocatuur steeds vaker gebruikmaakt van het recht op inzage in het digitale politiedossier. Dit recht geldt ook voor al het digitaal materiaal dat in het opsporingsonderzoek gebruikt is (De Jonge & Janssen, 2021; Egberts, 2022). In Nederland is het inmiddels mogelijk om vanaf het advocatenkantoor met Hansken inzage te krijgen in de bewijsbestanden, waarbij de resultaten van de eerdergenoemde geavanceerde analyses ook voor de advocatuur beschikbaar zijn (Openbaar Ministerie, 2023).

Van Beek et al. (2020) geven aan dat deze nieuwe manier van werken veel mogelijkheden biedt op het gebied van samenwerking en kennisdeling, maar dat het voor grote organisaties lastig blijkt een grote verandering van het verwerken van digitaal bewijs door te voeren. Het hele proces verandert, wat onder andere invloed heeft op de manier van samenwerken, communicatie, verantwoordelijkheden, technische infrastructuur en financiering.

5.2 Grootschalige toepassing van AI

Inmiddels wordt artificiële intelligentie, of zoals eerder aangehaald AI, breed ingezet om beelden te analyseren. De belangrijkste winst is dat AI rechercheurs kan helpen om sneller door de grote hoeveelheden beeldmateriaal te gaan. Een gemiddeld opsporingsonderzoek bevat al snel een miljoen beelden (afbeeldingen en video's), die onmogelijk één voor één bekeken en beoordeeld kunnen worden. Dankzij de rekenkracht die DFaaS biedt en de uniformiteit van een centraal sporenmodel (zoals in Hansken) kan AI grootschalig worden toegepast zodat de berg aan beeldmateriaal veel effectiever en efficiënter doorzocht worden.

In de eerste plaats is het kunnen zoeken in en filteren van het beeldmateriaal een vereiste. Met getrainde AI-modellen, kunnen objecten in beelden herkend en gelabeld worden. Rechercheurs kunnen die labels vervolgens gebruiken om te filteren. Voorbeelden van modellen die in het (digitaal) forensische domein ontwikkeld en gebruikt worden, zijn modellen die specifiek getraind zijn met voorbeelden die betrekking hebben op crimineel handelen, bijvoorbeeld drugslabs, partijen drugs, wapens, cashgeld, identiteitspapieren, maar ook zeecontainers (NFI, 2021). Naast het zoeken op gelabelde beelden, is het ook wenselijk om te kunnen zoeken op beelden die lijken op een bestaand beeld. Een meer geavanceerde techniek die hiervoor ontwikkeld is, CLIP (OpenAI, 2021), is ook inzetbaar via Hansken.

Een andere ontwikkeling op het gebied is Scene Text recognition, een techniek die in afbeeldingen teksten herkent. Het gaat dan niet om tekst in gescande documenten die via optische karakterherkenning (OCR) omgezet kan worden in tekst, maar om het herkennen van teksten in een op beeld vastgelegde scene, zoals tekst op straatnaamborden, op nummerplaten, en op gebouwen, of tekst op een scherm waarvan een foto is gemaakt (Jaderberg et al., 2014; NFI, 2021). Hieraan gerelateerd, zijn ook de recente ontwikkelingen op het gebied van *large language models*, zoals ChatGPT (OpenAI, 2023), potentieel interessant om het werk van

rechercheurs te ondersteunen. Er zijn al verschillende onderzoeken gedaan die de potentie van deze technieken, waaronder het gebruik als copiloot bij het rechercheren in digitaal (beeld)materiaal met een oplossing als Hansken (Henseler & Van Beek, 2023).

Ook het detecteren van gezichten in beeldmateriaal is een waardevolle toepassing in het digitaal-forensisch domein. Hiervoor zijn verschillende AI-technieken beschikbaar (Kumar et al., 2019), die het tevens mogelijk maken om te zoeken naar ander beeldmateriaal met vergelijkbare gezichten. Voor onderzoek naar seksueel kindermisbruik, zijn AI toepassingen beschikbaar die het werk van de zedenrechercheurs verlichten (Laureys et al., 2021; Sanchez et al., 2019). Enkele voorbeelden hiervan zijn een AI-model om de leeftijd van een persoon op een afbeelding te bepalen (Anda et al., 2018) en een AI-model om lichaamsdelen te herkennen, zodat deze bij het tonen van seksueel getinte afbeeldingen afgedekt kunnen worden (Project VIC, z.d.). Daarnaast kan beeldherkenning ook ingezet worden om foto's van verschillende slachtoffers met elkaar in verband te brengen op basis van andere kenmerken in de foto, denk aan meubilair of herkenbare punten op de muur in de achtergrond (ook wel bekend als *Visual Place Recognition,* zie bijv. Zhang et al., 2021).

Het spreekt voor zich dat zulke toepassingen het werk van rechercheurs makkelijker maken. Dit is echter niet zonder risico's (Jerian, 2021). Digitaal-forensisch onderzoek (en dus ook het gebruik van een digitaal-forensisch platform) is zoals eerder benoemd, gestoeld op verschillende principes. Resultaten moeten accuraat en reproduceerbaar zijn, onderzoeksstappen moeten herhaald kunnen worden. Dit betekent dat de gebruikte algoritmes deterministisch en uitlegbaar moeten zijn en bij voorkeur gevalideerd. Ook moeten zulke algoritmes in principe niet afhankelijk zijn van externe data (die over de tijd heen veranderen). AI-modellen voldoen zelden aan deze criteria. Deze modellen zijn vaak getraind met externe data, die bias kunnen bevatten (Das & Schuilenburg, 2020). Ook zijn de AI-modellen vaak niet uitlegbaar. Waarom een AI-model een bepaald antwoord geeft is in die gevallen niet duidelijk.

Nayerifard et al. (2023) hebben een zeer uitgebreid literatuuronderzoek uitgevoerd naar studies op het gebied van AI in het digitaal-forensisch domein. Net als Solanke (2022) concluderen ook zij dat er steeds meer aandacht moet zijn voor uitlegbaarheid van de modellen en technieken. In digitaal-forensisch onderzoek is hier een belangrijke taak weggelegd voor de gebruiker. Dankzij AI-modellen kan de gebruiker sneller en efficiënter zoeken. Maar aan het gebruik van deze technieken kleven dus ook risico's waar gebruikers zich van bewust moeten zijn. In opleidingen voor rechercheurs en analisten moet uitgelegd worden welke technieken ze wel of niet in kunnen zetten, welke risico's er spelen en hoe ze bevindingen moeten rapporteren. Uiteindelijk zullen ook de lezers van die rapportages (de rechterlijke macht, officieren van justitie, advocatuur, et cetera) ook moeten leren hoe zij de rapportages moeten lezen. In de volgende, afsluitende paragraaf zal naast de beantwoording van de centrale onderzoeksvragen worden stilgestaan bij enkele ontwikkelingen en uitdagingen, en worden enkele aanbevelingen gedaan voor verder onderzoek en de praktijk.

6 Discussie en conclusie

Digitaal-forensisch onderzoek wordt ingezet om van beeldmateriaal tot bewijs te komen. Voor deze onderzoeken is het van belang dat forensische principes gewaarborgd zijn, wat onder andere betekent dat onderzoek reproduceerbaar is, resultaten en conclusies onderbouwd en herleidbaar zijn. Forensisch beeldonderzoek richt zich op de beantwoording van verschillende soorten vragen rondom beeldmateriaal. Enerzijds richt dit onderzoek zich op het technisch beschikbaar en (visueel) zichtbaar en afspeelbaar maken van het beeldmateriaal. Anderzijds richt het zich op vraagstukken over de authenticiteit van het materiaal en wat in het authentieke beeldmateriaal te zien is daadwerkelijk bewijst. Voor die laatste vraag worden verschillende forensische onderzoeksmethoden ingezet, afhankelijk van het beschikbare beeldmateriaal en de specifieke vraag. Deze vragen zijn gericht op het bepalen van afmetingen en snelheden (fotogrammetrie), op het bepalen van de gebruikte camera, op het vergelijken van gezichten en op het verbeteren van beelden. Al deze vraagstukken vergen naast digitale kennis op bit- en byte-niveau ook kennis van beeldvorming door camerasystemen. Forensische data-analyse richt zich op de beantwoording van aanvullende vragen over digitale sporen, zoals vragen over de herkomst of het gebruik. Om zulke vragen te beantwoorden, wordt niet gekeken naar wat er op een beeld staat, maar naar de metadata en de omgeving waarbinnen het is aangetroffen. Dit soort onderzoek vergt kennis over het gedrag van digitale systemen op alle niveaus. Omdat data tegenwoordig op vele plekken voorkomt, en op vele plekken wordt vastgelegd, is dit vakgebied continu in ontwikkeling. Naast kennis van een eenvoudige thuis-PC, is steeds meer kennis nodig over bijvoorbeeld de techniek, werking en het gedrag van smartphones, Internet of Things, (cloud)servertechnologie en -opslag, virtualisatie, en ga zo maar door.

Voordat een specifieke afbeelding of video voor duiding binnen een zaak in aanmerking komt, moet deze eerst geïdentificeerd worden. Bij het opvragen van camerabeelden is dat niet zo'n probleem, maar als in een zaak vele digitale apparaten in beslag genomen zijn, die allemaal potentieel beeldmateriaal kunnen bevatten, is dit een hele speurtocht. Om dat proces te faciliteren, is het concept 'Digital Forensics as a Service' (kortweg DFaaS) geïntroduceerd. Hierbij wordt op een centraal platform samengewerkt tussen alle betrokkenen in een onderzoek. In het bijzonder kunnen rechercheurs die bekend zijn met de zaak op zoek naar (beeld)materiaal dat kan dienen als bewijs. Dit vraagt om forensische tools die de verschillende datastructuren uit bovenstaande bronnen kunnen uniformeren, zodat ze op een overzichtelijke en eenduidige manier aan de gebruikers gepresenteerd kunnen worden.

In het vervolg van de discussie zullen we vier ontwikkelingen en uitdagingen toelichten die belangrijk zijn als het gaat om het digitaal-forensisch werk. Het gaat hierbij om (1) het toenemend gebruik van AI, (2) de uitdagingen die generatieve AI met zich meebrengt, (3) de eigenschap van specialisatie op specialisatie en (4) de vaststelling dat techniek lang niet de enige factor is.

Toenemend gebruik AI

Net als de algemene maatschappelijke ontwikkeling, wordt ook in digitaal-forensisch onderzoek steeds meer gebruikgemaakt van artificiële intelligentie, onder meer om rechercheurs te assisteren in het vinden van relevant (beeld)materiaal. Mede door de opkomst

van DFaaS, kunnen technieken voor bijvoorbeeld het herkennen van objecten of tekst op beelden, of het vinden van gelijkende gezichten breed ingezet worden. Voor het goed kunnen toepassen van zulke technieken, is goede duiding en begrip van AI nodig, inclusief de bijkomende complexiteit zoals vraagstukken rond bias, vals-positieven en -negatieven, betrouwbaarheid en uitlegbaarheid van de gebruikte techniek. Dit geldt zeker als digitaal-forensisch onderzoek voor de analyse van beeldmateriaal als bewijs, in de toekomst gebruik gaat maken van multi-modale large language modellen, ook wel aangeduid als *visual foundation models* (Wu et al., 2023).

Om de AI-modellen te trainen, is het noodzakelijk om de bias die de trainingsdata bevatten te minimaliseren en tegelijkertijd kennis te hebben over de resterende bias in de data. Hiervoor zijn raamwerken beschikbaar, zoals het Total Error raamwerk (Amaya et al., 2020; Snaphaan & Hardyns, 2021b), die tot doel hebben om inzicht te verschaffen in fouten en bias in de data en ook mogelijkheden biedt om de mate van bias te kwantificeren. In welke mate deze raamwerken toepasbaar zijn op (digitaal-)forensisch onderzoek, zal toekomstig onderzoek en de (digitaal-)forensische praktijk moeten uitwijzen. Het gebruik van data om de modellen te trainen die in eigen beheer worden verzameld, waardoor inzicht in de mogelijke biases en fouten in de totstandkoming ervan per definitie groter is, geniet hierbij de voorkeur.

Ook komt er nieuwe Europese wetgeving op ons af op dit vlak in de vorm van de zogeheten *AI Act*, oftewel de Europese wet op de artificiële intelligentie. Deze (momenteel voorgestelde) AI Act hanteert een risicogebaseerde benadering, waar AI-toepassingen in diverse risicoklassen worden ingedeeld. Eén van de hoogrisicotoepassingen wordt als volgt aangeduid in de voorgestelde versie: “AI-systemen die bedoeld zijn om door rechtshandhavingsautoriteiten te worden gebruikt voor de beoordeling van de betrouwbaarheid van bewijsmateriaal tijdens het onderzoek naar of de vervolging van strafbare feiten” (Europese Commissie, 2021, p. 6). Echter, de in deze bijdrage beschreven toepassingen van AI in digitaal-forensisch onderzoek worden niet gebruikt bij de beoordeling van de betrouwbaarheid van bewijsmateriaal maar worden gebruikt ter ondersteuning van rechercheurs en analisten bij het zoeken in grote hoeveelheden digitaal bewijs. De beoordeling van het gevonden bewijs is altijd nog een taak van de gebruiker.

Generatieve AI als uitdaging

Tot nu toe was digitaal-forensisch onderzoek bij het onderzoek naar beeldmateriaal als bewijs vooral gericht op de metadata van het materiaal om de authenticiteit en waar mogelijk de herkomst vast te stellen of de activiteit die tot een afbeelding heeft geleid. Door de opkomst van generatieve AI waarmee synthetische foto’s zijn te maken die nauwelijks van echt zijn te onderscheiden, denk aan deep fakes, zal digitaal-forensisch onderzoek zich steeds meer moeten richten op de herkomst van een afbeelding. Was het een afbeelding die gemaakt is van een natuurlijk object of is de afbeelding het resultaat van een generatieve AI techniek? Er zal nieuwe expertise ontwikkeld moeten worden om de ‘krassporen’ van zulke nieuwe technieken te herkennen en te duiden (Henseler, 2022). Merk op dat ook AI-systemen die bedoeld zijn om door rechtshandhavingsinstanties te worden gebruikt voor de opsporing van deep fakes in de voorgestelde AI Act ook vallen onder AI-systemen met een hoog risico (Europese Commissie, 2021). De systemen die hierbij gebruikt worden, zullen dus voldoende inzicht moeten geven

(c.q. uitleggen) aan de gebruiker op grond waarvan een afbeelding als deep fake wordt bestempeld.

Specialisatie op specialisatie
Forensische onderzoek vormt een vakgebied en specialisme op zichzelf. Dit type van onderzoek vraagt specifieke expertise, competenties en vaardigheden. Met meer en meer beschikbare *digitale* data, sporen en potentieel bewijs is digitaal-forensisch onderzoek logischerwijs een enorme subdivisie geworden binnen dit vakgebied, zo niet een vakgebied op zichzelf. Het past forensische uitgangspunten toe op digitale data, waarvoor specifieke expertise, competenties en vaardigheden nodig zijn. Personen werkzaam in het forensisch domein zijn immers niet allemaal *digital natives*. De voornoemde ontwikkelingen van het toenemend gebruik van AI in het digitaal-forensisch onderzoek boort een derde laag in deze stapeling van specialismen aan. Naast de digitale vaardigheden, dient men nu ook kennis te hebben van AI, *machine learning* en *deep learning* om opgelijnd te blijven met state-of-the-art technieken die worden toegepast in het forensisch domein.

Naast dat dit uitdagingen biedt voor het forensisch vakgebied op zichzelf, vergt het ook een blijvende inspanning om de brede groep van gebruikers van forensisch(e) onderzoek en analyses mee te nemen in (het proces en de uitkomsten van) het digitaal-forensisch onderzoek. Dit geldt niet alleen voor de direct betrokkenen bij het opsporingsonderzoek en forensisch onderzoek, maar ook de partijen die de uitkomsten van het onderzoek op een juiste manier moeten gebruiken, waaronder officieren van justitie, rechters en advocaten (Henseler & van Loenhout, 2018). Ook voor deze gebruikers kan generatieve AI (en dan met name de large language modellen zoals ChatGPT) hulp bieden in de vorm van een copiloot die uitleg kan geven over een rapport op een manier die de lezer begrijpt en eventueel zelfs in de vorm van een dialoog (Henseler, 2023).

De techniek is lang niet de enige factor
Hoewel met de opkomst van DFaaS een centraal digitaal-forensisch medium wordt geboden waarmee data en techniek toegankelijk worden gemaakt voor een brede groep gebruikers, is zo'n centraal digitaal-forensisch platform geen wondermiddel, zo leggen Van Loenhout en Van Beek (2022) uit. Digitaal specialisten en gespecialiseerde analysetools blijven absoluut noodzakelijk bij alle onderzoeken. Tegelijkertijd is het een continue zoektocht om het 'gat' tussen kennis van de zaak en het vermogen om data te analyseren te dichten (Roest, 2023). Rechercheurs kunnen dankzij een DFaaS-oplossing snel zelfstandig inzicht krijgen in het digitale bewijs zonder dat ze op een rapportage van de digitale experts hoeven te wachten. Op hun beurt kunnen de digitale experts zich daardoor concentreren op de belangrijkste en complexe forensische vraagstukken. Dit maakt de verwerking van grote hoeveelheden data beter beheersbaar, met forensisch bewijs als resultaat.

Het maakt ook nieuwe ontwikkelingen mogelijk die niet technisch van aard zijn. Zo blijft het een uitdaging om gegevens die onder het verschoningsrecht vallen te kunnen identificeren en ontoegankelijk maken of zelfs vernietigen. Het biedt ook kansen, waaronder het geven van toegang tot digitaal (beeld)materiaal aan andere partijen in het strafproces, zoals advocaten en rechter. Tevens blijkt het, zoals eerder aangehaald, voor (grote) organisaties lastig een

dergelijke verandering omtrent het verwerken van digitaal bewijs in de praktijk door te voeren (Van Beek et al., 2020).

Concluderend, om van beeldmateriaal naar bewijs te komen, is digitaal-forensisch onderzoek onontbeerlijk. We zijn aangekomen in het big-datatijdperk met een explosie van het gebruik van artificiële intelligentie. Dit maakt het bewaken van de forensische uitgangspunten belangrijker maar tegelijkertijd ook complexer dan ooit. Hoewel nieuwe ontwikkelingen het forensisch domein doen bewegen, moeten we vooral meebewegen en deze forensische uitgangspunten centraal blijven stellen, om (beeld)materiaal ook in de toekomst als bewijs te kunnen blijven gebruiken.

Referenties

Aitken, C., & Taroni, F. (2004). *Statistics and the evaluation of evidence for forensic scientists*. John Wiley & Sons.

Amaya, A., Biemer, P. P., & Kinyon, D. (2020). Total error in a big data world: Adapting the TSE framework to big data. *Journal of Survey Statistics and Methodology, 8,* 89-119. https://doi.org/10.1093/jssam/smz056

Anda, F., Lillis, D., Le-Khac, N. A., & Scanlon, M. (2018). Evaluating automated facial age estimation techniques for digital forensics. In *2018 IEEE Security and Privacy Workshops (SPW)* (pp. 129-139). IEEE.

Arbia, G. (2021). *Statistics, new empiricism and society in the era of big data*. Springer.

Casino, F., Dasaklis, T. K., Spathoulas, G. P., Anagnostopoulos, M., Ghosal, A., Borocz, I., Solanas, A., Conti, M., & Patsakis, C. (2022). Research trends, challenges, and emerging topics in digital forensics: A review of reviews. *IEEE Access, 10,* 25464–25493. https://doi.org/10.1109/ACCESS.2022.3154059

Centraal Bureau voor de Statistiek. (2022, 1 maart). *Minder traditionele criminaliteit, meer online criminaliteit.* https://www.cbs.nl/nl-nl/nieuws/2022/09/minder-traditionele-criminaliteit-meer-online-criminaliteit

Das, A., & Schuilenburg, M. (2020). 'Garbage in, garbage out'. Over predictive policing en vuile data. *Beleid en Maatschappij, 47*(3), 254-268.

De Jonge, D. N., & Janssen, S. L. J. (2021). Eindelijk toegang tot datasets. (Erg) langzaam maar zeker naar een nieuw normaal/ *Nederlands Juristenblad, 2021*(34), 2793–2799.

De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data. A consensual definition and a review of key research topics. *AIP Conference Proceedings, 1644*(1), 97–104.

Doorenbos, D. R., & Rosing, M. E. (2020) Recht doen aan het verschoningsrecht. *Tijdschrift voor Sanctierecht & Onderneming, 2020*(5-6), 217–224.

Egberts, M. M. (2022). De reikwijdte van het inzagerecht en 'equality of arms' in het licht van grote datasets, Hansken en toekomstige ontwikkelingen. *Tijdschrift voor Bijzonder Strafrecht & Handhaving, 2022*(2), 119–129. https://doi.org/10.5553/TBSenH/229567002022008002006

Europese Commissie. (2021, 21 april). *Bijlagen bij het Voorstel voor een verordening van het Europees Parlement en de Raad tot vaststelling van geharmoniseerde regels betreffende artificiële intelligentie (wet op de artificiële intelligentie) en tot wijziging van bepaalde wetgevingshandelingen van de Unie*. https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0005.02/DOC_2&format=PDF

Favaretto, M., De Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of big data? Researchers' understanding of the phenomenon of the decade. *PLoS ONE, 15*(2), Article e0228987. https://doi.org/10.1371/journal.pone.0228987

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods and analytics. *International Journal of Information Management, 35*(2), 137-144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gerechtshof 's-Hertogenbosch (2023). *Hoger beroep kort geding uitspraak in zaak met zaaknummer 200.310.320_01*. ECLI:NL:GHSHE:2023:1329

Hansken Community. (z.d.). *Hansken, the Open Digital Forensic platform*. https://hansken.org/

Henseler, H. (2010). E-*Discovery: Op zoek naar de digitale waarheid*. HVA Publicaties.

Henseler, H. (2017), *De (r)evolutie van digitaal bewijs.* Lectorale rede, Hogeschool Leiden. https://www.hsleiden.nl/binaries/content/assets/hsl/lectoraten/digital-forensics-en-e-discovery/publicaties/2017/lectorale-rede-hans-henseler-2017.pdf

Henseler, H. (2022). De opkomst van kunstmatige intelligentie in expertise en recht: De 'krassporen' van deeplearning. *Expertise en Recht*, *2022*(2), 31–34.

Henseler, H. (2023). De opkomst van kunstmatige intelligentie in expertise en recht: de invloed van ChatGPT en grote taalmodellen. *Expertise en Recht, 2023*(5), 135-138.

Henseler, H., & Van Beek, H. (2023). ChatGPT as a copilot for investigating digital evidence. In J. G. Conrad, D. W. Linna Jr., J. R. Baron, H. Henseler, P. Bhattacharya, A. Nielsen, J. Vinjumur, J. Pickens & A. Jones (Eds.), *Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace* (pp. 58–69). https://ceur-ws.org/Vol-3423/

Henseler, H., & van Loenhout, S. (2018). Educating judges, prosecutors and lawyers in the use of digital forensic experts. *Digital Investigation, 24*, S76-S82. https://doi.org/10.1016/j.diin.2018.01.010

Jaderberg, M., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). *Synthetic data and artificial neural networks for natural scene text recognition*. https://doi.org/10.48550/arXiv.1406.2227

Jerian, M. (2021). *Can AI be used for forensics and investigations?* Amped Software Blog post. https://blog.ampedsoftware.com/2021/10/05/can-ai-be-used-for-forensics-and-investigations

Julliand, T., Nozick, V., & Talbot, H. (2016). Image noise and digital image forensics. In *Digital-Forensics and Watermarking: 14th International Workshop, IWDW 2015, Tokyo, Japan, October 7-10, 2015, Revised Selected Papers 14* (pp. 3–17). Springer.

Kitchin, R. (2014). *The data revolution. Big data, open data, data infrastructures & their consequences.* Sage.

Kokshoorn, B., De Koeijer, J. A., Aarts, B., Blankers, B. J., Matas Llonch, T., & Berger, C. E. H. (2020) Scenario's, hypothesen, aannamen en context-informatie; wat bedoelt de deskundige eigenlijk? *Expertise en Recht*, *2020*(3), 96–104.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Kumar, A., Kaur, A., & Kumar, M. (2019). Face detection techniques: A review. *Artificial Intelligence Review*, *52*, 927–948. https://doi.org/10.1007/s10462-018-9650-2

Laney, D. (2001). *3D data management: Controlling data volume, velocity and variety*. Retrieved from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Laureys, M., Snaphaan, T., & Hardyns, W. (2021). Online beeldmateriaal van seksueel kindermisbruik: Innovatieve technologieën ter identificatie van dader en slachtoffer. *Polinfo.be*. https://polinfo.kluwer.be

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444. https://doi.org/10.1038/nature14539

Marr, D. (1982). *Vision. A computational investigation into the human representation and processing of visual information.* W.H. Freeman.

Nayerifard, T., Amintoosi, H., Bafghi, A. G., & Dehghantanha, A. (2023). *Machine learning in digital forensics: A systematic literature review*. https://doi.org/10.48550/arXiv.2306.04965

Nederlands Forensisch Instituut. (2018). *Vakbijlage 'Forensisch gebruik van bestandskenmerken en bijbehorende hashalgoritmen'*. https://www.forensischinstituut.nl/over-het-nfi/vakbijlagenen-informatiebladen

Nederlands Forensisch Instituut. (2021). *Duizenden foto's sneller doorzoeken dankzij slim algoritme*. https://magazines.forensischinstituut.nl/atnfi/2021/35/duizenden-foto%E2%80%99s-sneller-doorzoeken-dankzij-slim-algoritme

Nederlands Forensisch Instituut. (2022). *'We moeten alert zijn op het herkennen van verborgen berichten van verdachten'*. https://www.forensischinstituut.nl/actueel/nieuws/2022/07/26/we-moeten-alert-zijn-op-het-herkennen-van-verborgen-berichten-van-verdachten

Nederlands Forensisch Instituut. (z.d.-a). *Forensische big data analyse*. https://www.forensischinstituut.nl/forensisch-onderzoek/forensische-big-data-analyse

Nederlands Forensisch Instituut. (z.d.-b). *Beeldonderzoek en biometrie (Forensisch).* https://www.forensischinstituut.nl/forensisch-onderzoek/beeldonderzoek-en-biometrie-forensisch

OpenAI. (2021). *CLIP: Connecting text and images.* https://openai.com/research/clip

OpenAI. (2023). *ChatGPT*. https://openai.com/chatgpt

Openbaar Ministerie. (2023). *Digitaal bewijsmateriaal via Hansken nu raadpleegbaar voor advocaten vanaf werkplek*. https://www.om.nl/actueel/nieuws/2023/03/20/digitaal-bewijsmateriaal-via-hansken-nu-raadpleegbaar-voor-advocaten-vanaf-werkplek

Palys, T., & Atchison, C. (2013). Text, image, audio, and video: Making sense of non-numeric data. In T. Palys & C. Atchison (Eds.), *Research decisions: Quantitative, qualitative and mixed-method approaches* (pp. 303-331). Nelson.

Project VIC. (z.d.). *VICSafer*. Project VIC international. https://www.projectvic.org/vicsafer

Reedy, P. (2023). Interpol review of digital evidence for 2019–2022. *Forensic Science International: Synergy*, 6, 100313.

Robertson, B., Vignaux, G. A., & Berger, C.E. (2016). *Interpreting evidence: Evaluating forensic science in the courtroom*. John Wiley & Sons.

Roest, D. (2023). Big data en politiewerk, een onoverbrugbare kloof? Hoe TROI de brug slaat van big data naar politiewerk. In T. Snaphaan, W. Hardyns, A. van Dijk, R. Spithoven & R. Van Brakel (Eds.), *Big data policing* (pp. 91–106). Gompel&Svacina.

Rosenberg, D. (2013). Data before the fact. In L. Gitelman (Ed.), *Raw data is an oxymoron*. MIT Press.

Sanchez, L., Grajeda, C., Baggili, I., & Hall, C. (2019). A practitioner survey exploring the value of forensic tools, AI, filtering, & safer presentation for investigating child sexual abuse material (CSAM). *Digital Investigation, 29*, S124-S142.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*, Article 160. https://doi.org/10.1007/s42979-021-00592-x

Schneider, C. (2016, 25 mei). *The biggest data challenges that you might not even know you have.* https://www.ibm.com/blogs/watson/2016/05/biggest-data-challenges-might-not-even-know/

Schwab, K. (2017). *The fourth industrial revolution*. Random House.

Snaphaan, T., & Hardyns, W. (2020). Het gebruik van beeldmateriaal in criminologisch onderzoek. Van turven met pen en papier tot geautomatiseerde detectie en classificatie van overlastfenomenen. In W. Hardyns & T. Snaphaan (Eds.), *Big data en innovatieve methoden voor criminologisch onderzoek* (pp. 305–332). Boom criminologie.

Snaphaan, T., & Hardyns, W. (2021a). Environmental criminology in the big data era. *European Journal of Criminology*, *18*(5), 713–734. http://doi.org/10.1177/1477370819877753

Snaphaan, T., & Hardyns, W. (2021b). Handvatten voor een kwaliteitsbeoordeling van big data: De introductie van het Total Error raamwerk. *Tijdschrift voor Veiligheid*, *20*(4), 63–88. https://doi.org/10.5553/TvV/.000033

Snaphaan, T., Hardyns, W., van Dijk, A., Spithoven, R., & Van Brakel, R. (Eds.). (2023). *Big data policing*. Gompel&Svacina.

Solanke, A. A. (2022). Explainable digital forensics AI: Towards mitigating distrust in AI-based digital forensics analysis using interpretable models. *Forensic Science International: Digital Investigation*, *42*, Article 301403.

Szeliski, R. (2022). *Computer vision: Algorithms and applications*. Springer.

Van Baar, R. B., Van Beek, H. M. A., & Van Eijk, E. J. (2014). Digital forensics as a service: A game changer. *Digital Investigation*, *11*, S54-S62.

Van Beek, H. M. A., Van Den Bos, J., Boztas, A., Van Eijk, E. J., Schramp, R., & Ugen, M. (2020). Digital forensics as a service: Stepping up the game. *Forensic Science International: Digital Investigation*, *35*, Article 301021.

Van Beek, H. M. A., Van Eijk, E. J., Van Baar, R. B., Ugen, M., Bodde, J. N. C., & Siemelink, A. J. (2015). *Digital forensics as a service: Game on*. Digital Investigation, 15, 20-38.

Van Bree, R. J. P., & Schramp, R. (2019) *Terug naar de bestanden. Technische toelichting over identificeren, verbergen en verwijderen van bestanden.* https://www.forensischinstituut.nl/over-het-nfi/publicaties/rapporten/2019/05/21/technische-toelichting-terug-naar-de-bestanden

Van den Hurk, J. W., Klaar, R. J. A., Mossink, J. J. (2023). *Art. 240b Sr, Juridische en digitaal-technische aspecten van strafvervolging wegens gedragingen met digitaal kinderpornografisch materiaal*. Kenniscentrum Cybercrime voor de Rechtspraak, versie 4.0 (2023). https://www.rechtspraak.nl/SiteCollectionDocuments/ebook-240b-sr.pdf

Van Loenhout, T. & van Beek, H. (2022). Tech Talk: Digital Forensics as a Service, Not a Silver Bullet, but a Crucial Tool. *Police Chief Magazine*, vol. 89, nr. 7, pp. 56-58. IAPC. https://www.policechiefmagazine.org/tech-talk-digital-forensics-as-a-service/

Wu, C., Yin, S., Qi, W., Wang, X., Tang, Z., & Duan, N. (2023). *Visual ChatGPT: Talking, Drawing and Editing with Visual Foundation Models* https://arxiv.org/pdf/2303.04671.pdf

Zhang, X., Wang, L., & Su, Y. (2021). Visual place recognition: A survey from deep learning perspective. *Pattern Recognition, 113,* Article 107760. https://doi.org/10.1016/j.patcog.2020.107760