

Moderne digitale technologieën in de strijd tegen kindermishandeling en huiselijk geweld?

- Een beschouwing op potentieel, voorwaarden en ethische overwegingen

Drs. Eef Dooren-Sulsters - Lectoraat Orthopedagogisch Handelen en Diversiteit; Fontys Pedagogiek.

Inleiding

Kindermishandeling en huiselijk geweld vormen mondiaal een hardnekkig, ogenschijnlijk onoplosbaar probleem met verstrekkende gevolgen voor het slachtoffer, diens directe omgeving maar ook de samenleving als geheel. De WHO schat dat er ca. 1 miljard kinderen jaarlijks slachtoffers zijn (WorldHealthOrganization, 2006). Kindermishandeling komt overal ter wereld voor en ondanks pogingen om (huiselijk) geweld tegen kinderen tegen te gaan, blijft het aantal slachtoffers hoog. Met de opkomst van moderne digitale technologieën – denk hierbij aan toepassingen van big data, artificial intelligence, machine learning, het gebruik van algoritmes – worden er voorzichtige pogingen gedaan om te onderzoeken of er wellicht oplossingsrichtingen zijn om beter zicht te krijgen op en wellicht zelfs verminderen van kindermishandeling en huiselijk geweld. Tegelijkertijd roept het gebruik van technologieën ook de nodige vraagstukken op. Kan het gebruik van artificial intelligence een bijdrage leveren aan de kennis over of zelfs voorspelling van dit sociale vraagstuk? Kunnen fundamentele mensenrechten zoals het recht op eerbiediging van het privéleven wel worden beschermd bij het gebruik van beschikbare big data (Eynikel, 2017)? Elke nieuwe (technologische) toepassing roept de vragen op van sociale en morele aard, over de wenselijkheid van het gebruik (Verbeek, 2014). Dit artikel wil inzicht verschaffen in de vraag welke potentie het gebruik van moderne digitale technologieën heeft als het gaat om het verminderen of voorkomen van kindermishandeling en het onderzoek hiernaar. Daarnaast zal worden stilgestaan bij de vraag welke voorwaarden en ethische overwegingen in ogenschouw dienen te worden genomen bij het gebruik ervan. Aangezien de sociale en/of welzijnssector en de digitale vooralsnog behoorlijk gescheiden werelden zijn, is echter het zaak om eerst kort stil te staan bij de vraag om welke moderne digitale technologieën het gaat. In de literatuur worden termen als big data, artificial intelligence, machine learning en predictive analytics veelvuldig gebruikt en soms wat door elkaar gebruikt, maar wat betekenen deze?

Moderne digitale technologieën – big data, algoritmes, artificial intelligence, machine learning en predictive analytics

De term big data verwijst naar een zeer grote hoeveelheden digitale en gevarieerde data die op diverse plaatsen zijn opgeslagen. De term verwijst daarnaast ook naar de technieken die gebruikt worden om deze data te beheren, op te slaan, te verwerken en te analyseren (Hardyns & Snaphaan, 2020). Door de grotere rekenkracht van hedendaagse computers is het mogelijk om zeer grote hoeveelheden, zeer gevarieerde data met elkaar te verbinden en te verwerken. Er wordt bij het gebruik van big data vaak gesproken over vier V's: volume, variety, velocity en veracity (Blauw, 2021). Niet alleen de kwantiteit van data staat centraal (volume). Ook het type data dat wordt gebruikt is zeer gevarieerd (variety). Ter illustratie, chat- of tekstbestanden kunnen met videofragmenten en locatiegegevens met elkaar in verband worden gebracht, terwijl ze zeer verschillende informatie bevatten. Om de data met elkaar in verband te kunnen brengen moeten deze data wel relatief snel kunnen worden opgeslagen en opgehaald om te kunnen worden verwerkt (velocity). Tot slot is het bij big data van belang dat de gebruikte data betrouwbaar zijn en dat er betrouwbare resultaten worden verkregen (veracity). Het verwerken en analyseren van big data gebeurt

geautomatiseerd door het gebruik van algoritmes. Algoritmes zijn een set met instructies of programma's die ervoor zorgen dat een computer weet wat er met data moet gebeuren (Lanier et al., 2020; O'Neil, 2016). Denk hierbij bijvoorbeeld aan het geven van een zoekopdracht in Google. Stel iemand wil weten wat "kindermishandeling" is. Door deze term in Google in te typen, verschijnen er websites die op een of andere manier gerelateerd zijn aan kindermishandeling, bijvoorbeeld omdat de term letterlijk op de website vermeld staat. Algoritmes kunnen echter ook zo geschreven zijn dat ze op basis van de data alleen patronen leren herkennen in de data, of door gebruik te maken van nieuwe data om zichzelf te verbeteren. Dit wordt aangeduid met machine learning en is een specifieke vorm van artificial intelligence. Bij deze technologie gaat het erom dat er geautomatiseerd wordt geleerd van ervaringen zodat er voorspellingen kunnen worden gedaan over een bepaalde kwestie en zo geautomatiseerde beslissingen worden genomen. Bij het eerdere voorbeeld blijvend: een algoritme kan ook zo zijn geschreven dat het op basis van websites die worden aangeklikt na invoeren van 'kindermishandeling' zichzelf traint wat de beste hit is, bijvoorbeeld omdat er het vaakst op wordt geklikt. Of door informatie uit de eerdere zoekgeschiedenis van de persoon te bepalen wat de meest relevante website zou zijn. Machine learning is in feite de computer die zichzelf op basis van data en algoritmes traint om tot steeds betere resultaten te komen zonder tussenkomst van een programmeur (O'Neil, 2016).

Globaal zijn er twee soorten machine learning: supervised en unsupervised. Unsupervised machine learning is het achterhalen van patronen zonder dat er een bepaald resultaat wordt verwacht. Het algoritme gaat zelf op zoek naar patronen en verbanden. Bij supervised machine learning wordt een deel van een dataset gebruikt waarbij de uitkomst (of het resultaat) wel bekend is. Op basis hiervan kan worden achterhaald welke patronen en/of variabelen onder dit resultaat liggen en kan het na verloop van tijd nieuwe uitkomsten gaan voorspellen. Het analyseren van big data kan op verschillende niveaus. Lanier et al. (2019) wijzen hier op een taxonomie die binnen de zogeheten business analytics wordt gebruikt:

- descriptief (fenomenen worden in kaart gebracht),
- diagnostisch (verklaren waarom iets gebeurt),
- voorspellend (voorspellen wat er in toekomst zal gebeuren)
- prescriptief (aangeven wat de beste handelwijze in een situatie is).

Met name dit voorspellende niveau, de predictive analytics, wordt in de literatuur rondom mogelijke toepassingen op het gebied van het (signaleren en mogelijk zelfs voorkomen kindermishandeling aangehaald. Dit zal in onderstaande paragraaf worden uitgewerkt. Het proberen te voorspellen van gebeurtenissen is op zich geen nieuw fenomeen, maar tegenwoordig kan dit door middel van enorm grotere hoeveelheden data, data die een enkeling alleen niet zou kunnen analyseren (Lanier et al., 2020; Leslie, Holmes, Hitrova, & Ott, 2020; O'Neil, 2016). Dat het gebruik van de genoemde technologieën op grote schaal plaatsvindt en dat het de nodige mogelijkheden biedt binnen diverse domeinen is in dit digitale tijdperk evident, maar zijn deze toepassingen ook mogelijk bruikbaar binnen het sociale domein?

De potentie van moderne digitale technologieën toegepast op vraagstukken rondom kindermishandelingen huiselijk geweld

Binnen het sociale domein wordt op grote schaal data gegenereerd (Coulton, Goerge, Putnam-Hornstein, & de Haan, 2015). Hierbij valt te denken aan gegevens die worden opgeslagen binnen instellingen als scholen, hulpverleningsinstanties maar ook aan medische

gegevens bij de huisarts of in het ziekenhuis. Scholen houden leerlingvolgsystemen bij, hulpverleningsinstanties verzamelen data over te bereiken doelstellingen, de geboden hulp, het resultaat hiervan, maar ook algemene gegevens over de achtergrond van cliënten. Daarnaast zijn er tal van andere domeinen waar mogelijk bruikbare data opgeslagen ligt – denk hierbij aan open databanken, posts op sociale media etc... (Coulton et al., 2015). Zijn deze data mogelijkwerwijs in te zetten om inzicht te krijgen in kindermishandeling en huiselijk geweld en mogelijk zelfs het verminderen ervan? De nieuwe digitale technologieën hebben bewezen sterk te zijn in het herkennen van patronen. Hierin blijken zij beter dan mensen omdat zij in staat zijn kwantitatief meer maar ook meer diverse data te analyseren en hieruit patronen te destilleren (Henman, 2020). Een kracht van het toepassen van digitale technologieën is dat het dus mogelijkwerwijs nieuwe kennis en inzicht kan opleveren ten aanzien van kindermishandeling en huiselijk geweld (Cash, 2019). Wellicht dat er in alle mogelijk onbekende beschermende of risicofactoren ten aanzien van kindermishandeling verborgen liggen? Zoals eerder aangegeven is de genoemde problematiek hardnekkig en mogelijk kunnen nieuwe inzichten tot nieuwe oplossingen leiden.

Zoals benoemd is in de literatuur veel aandacht voor predictive analytics – soms verwoord als ‘predictive risk modelling’ (PRM) (de Haan & Connolly, 2014; Gillingham, 2016; Keddell, 2015). Predictive analytics, toegepast op vraagstukken rondom kindermishandeling en huiselijk geweld, gaat over het aangeven van hoe waarschijnlijk het zal zijn dat er in een gezin sprake van mishandeling is of dat dit in de toekomst zo zal zijn. Dit voorspellen of inschatten van het risico op onveiligheid binnen een gezin, is op zich niet iets wat alleen computers kunnen. Het inschatten van risico’s is een taak van hulpverleners opdat zij gezinnen kunnen bijstaan en adequate hulp te bieden. Kindermishandeling en huiselijk geweld vinden echter in de regel niet of nauwelijks plaats in de openbaarheid. Zowel slachtoffers als “daders” treden hier niet snel mee naar buiten (de Haan & Connolly, 2014). Hoe kan een hulpverlener inschatten op wie hij zich moet richten wanneer deze persoon of dit gezin zich niet heeft gemeld, en vooral ook, wanneer hij ook tal van andere werkzaamheden heeft? Of, wanneer hij meer meldingen heeft dan hij kan onderzoeken en dus keuzes moet maken (Russell, 2015)?

Om zicht te krijgen op de kansen die predictive analytics biedt, hoewel er ook zeker de nodige aandachtspunten liggen (Coulton et al., 2015; Lanier et al., 2020), eerst een voorbeeld van hoe dit voorspellen globaal werkt. Binnen de verzekeringswereld is het gebruikelijk om risico’s en de waarschijnlijk dat iets zal gebeuren in te schatten op basis van statistische analyse. Iemand die zijn fiets wil verzekeren tegen diefstal vult een aantal zaken in en op basis daarvan wordt ingeschat hoe groot de kans is dat de fiets zal worden gestolen. Verzekeringsmaatschappijen hebben enorme hoeveelheden data over fietsdiefstal. De kans dat een dure, nieuwe e-bike wordt gestolen, is groter dan een roestige oude damesfiets; in grote steden is de kans op diefstal groter dan op het platteland etc... Op basis hiervan wordt de premie voor de verzekering vastgesteld. Vergelijkbare principes kunnen ook worden toegepast als het gaat om kindermishandeling en huiselijk geweld. Op basis van beschikbare (big) data en analyse kan een risicoprofiel worden opgesteld en zo kan worden aangegeven bij welke gezinnen de kans op mishandeling groter of kleiner is. Het voordeel hierbij is dat de risico-inschatting bij iedereen op eenzelfde manier plaatsvindt en dat er geen zaken over het hoofd zouden kunnen worden gezien. Dit zou leiden tot objectievere en kwalitatief betere beslissingen leiden (Church & Fairchild, 2017) en tevens tot meer effectieve toekenning van hulp (Gillingham, 2016; Hunt et al., 2020; Leslie et al., 2020; Russell, 2015). Dit bieden van hulp hoeft overigens niet alleen reactief van aard zijn – dus wanneer er al sprake is van mishandeling, in welke vorm dan ook – maar kan ook preventief – hulp bieden

voordat er sprake is van onveiligheid (de Haan & Connolly, 2014; Gillingham, 2016; Lanier et al., 2020).

Zoals aangegeven lijkt de inzet van nieuwe digitale technologieën nieuwe kansen en mogelijkheden te bieden. Op een juiste manier toegepast kan het voordelen hebben als het gaat om het eerder in beeld hebben van kwetsbare gezinnen, het mogelijk meer inzicht krijgen in de prevalentie van kindermishandeling en huiselijk geweld, kan het tot kwalitatief betere en objectievere beslissingen leiden als het gaat over het bieden van gepaste ondersteuning. Daarnaast lijkt het gebruik te kunnen leiden tot meer efficiency en kostenbesparing, want analyses worden automatisch en dus snel gemaakt. Gezien de huidige stand van zaken rondom dit thema, dient er echter ook een aantal overwegingen in acht te worden genomen bij het gebruik van technologieën als big data en machine learning, waarbij ethische overwegingen niet in de laatste plaats een centrale rol innemen. Deze zullen onderstaand worden toegelicht. Hierbij zal allereerst de focus worden gelegd op de meer programmatische uitdagingen en vervolgens op de meer ethische aspecten, hoewel deze niet geheel los van elkaar te zien zijn.

Kwaliteit van de data(set)

Een van de belangrijkste uitdagingen in het gebruik van nieuwe digitale technologieën bij (het voorkomen van) kindermishandeling, kan gekoppeld worden aan het gezegde “garbage in, garbage out”. Om betekenisvolle patronen te kunnen herkennen en/of juiste voorspellingen te kunnen doen is het noodzakelijk om te beschikken over kwalitatief goede data. De praktijk wijst echter uit dat dit vooralsnog niet het geval is. Russel (2015) benoemt een viertal aspecten die hiermee samenhangen. Hij wijst erop dat er (kwantitatief) voldoende data beschikbaar moet zijn, dat er genoeg variabelen zijn, dat er niet teveel data mogen ontbreken en dat de data juist en consistent moet zijn ingevoerd. Binnen het sociale domein is dit nog niet altijd zo (Coulton et al., 2015; Leslie et al., 2020; Russell, 2015). De beschikbare data zullen sowieso eerst ‘gebruiksklaar’ moeten worden gemaakt voor zij überhaupt gebruikt kunnen worden (Coulton et al., 2015). Daarnaast is veel van de beschikbare data aanvankelijk bedoeld voor andere doeleinden en gegevens worden lang niet altijd systematisch en volledig geregistreerd (Leslie et al., 2020). Indien datasets onvolledig zijn – doordat bijvoorbeeld niet alle gegevens bekend of ingevoerd zijn – kan dit gevolgen hebben voor de representativiteit en zo voor de kwaliteit van de data op basis waarvan zal worden geanalyseerd. Een dataset waarin bijvoorbeeld inconsistent wordt aangegeven wat de gezinssamenstelling is, maakt dat er in de analyse een verband wordt gelegd tussen gezinssamenstelling en mogelijk voorkomen van kindermishandeling terwijl deze verbanden niet zouden worden gelegd bij een consistente invoer van gezinssamenstelling.

Zelfs al lijken datasets wel volledig, dan nog kan de representativiteit van de set problematisch zijn. Keddell (2015) wijst in dit geval op het grootschalige onderzoek rondom het ontwikkelen van een Predictive Risk Model (PRM, predictive analytics) in Nieuw-Zeeland op dat representativiteit hoe dan ook punt van aandacht is. In het betreffende onderzoek is de keuze gemaakt om een dataset te maken op basis van data vanuit uitkeringsinstanties in Nieuw Zeeland en vanuit kinderbeschermingsdiensten. Dit betekent dat data over mensen die geen uitkering krijgen en die niet bekend zijn bij (bepaalde) hulpverleningsinstanties niet zijn opgenomen in het model. Hoewel de gebruikte databases dus wel landelijk zijn, is het niet zo dat er data over alle burgers worden meegenomen in het model (Keddell, 2015). Dit roept vragen op over de representativiteit van de data. Het gebruik van dergelijke niet-representatieve of onvolledige datasets maakt dat bepaalde groepen mogelijk oververtegenwoordigd zijn en dit heeft invloed op de kwaliteit van de output van het model

(‘garbage out’). Ook kan dit onbedoeld negatieve en stigmatiserende consequenties hebben (de Haan & Connolly, 2014). Dit laatste zal nader worden toegelicht in de paragraaf ethische overwegingen.

Kwaliteit van algoritmes

Naast de kwaliteit van de data, is het tevens zaak dat de gebruikte algoritmes eveneens kwalitatief goed zijn. Hieronder kan worden verstaan dat het algoritme accuraat is, dat het de juiste patronen opspoort of bij predictive analytics, correcte voorspellingen doet. Met name de kwaliteit van de algoritmes bij pogingen te voorspellen welke kinderen risico lopen, is onderwerp van discussie (Gillingham, 2019b; Keddell, 2019; Leslie et al., 2020). Het voorspellen welke kinderen meer of minder risico lopen op mishandeling wordt hoe dan ook bemoeilijkt door het feit dat, hoewel een substantieel deel van de bevolking slachtoffer is, het merendeel dit niet is. Het voorspellen van “rare events” is lastig, ongeacht de gebruikte methode (Church & Fairchild, 2017; Lanier et al., 2020). Als het gaat om het gebruik van algoritmes, zijn er daarnaast nog twee belangrijke zaken om in ogenschouw te nemen, te weten het (onbedoeld) invoeren van vooroordelen (“bias”) enerzijds en vals positieve dan wel negatieve voorspellingen anderzijds. Om (big) datasets te kunnen analyseren wordt gebruik gemaakt van algoritmes, deze zijn het werk van mensen en zij kunnen vooroordelen hebben welke in het algoritme kunnen worden verwerkt. Leslie et al. (2020) wijzen erop dat de rol van ontwerpers in het ontwikkelen van machine learning modellen van groot belang is aangezien zij onwillekeurig eigen aannames in het model kunnen verwerken. Hierbij kan worden gedacht aan de keuzes voor welke datasets gebruikt worden, welke parameters en hoe de uitkomsten van het model worden geïnterpreteerd. Gillingham (2019) wijst er eveneens op dat het ontwerp van het algoritmische model en de keuzes voor de gebruikte datasets mensenwerk is, dat onbewust vooroordelen en aannames in het model kunnen sluiten en dus dat het model minder neutraal is dan wellicht wordt gedacht. Het onbedoeld inweven van vooroordelen in modellen kan, net als bij het gebruik van onvolledige en niet-representatieve data, leiden tot stigmatisering en versterken van diezelfde vooroordelen, in plaats van het geven van een realistische weergave van de werkelijkheid.

Nog los van eventuele vooroordelen, is het operationaliseren van kindermishandeling problematisch. Wanneer is een klap een pedagogische tik en wanneer wordt het mishandeling? Dergelijke definities zijn echter wel nodig om moderne digitale technologieën in te kunnen zetten. Om de principes van machine learning toe te passen in het opsporen van mishandeling, dient het model te weten welke uitkomst van belang is (“outcome variable”), te weten die gevallen waar mishandeling heeft plaatsgevonden. Dit veronderstelt een heldere definitie van mishandeling en een heldere beschrijving van waar in de data naar kan worden gezocht. Dit is minder onproblematisch dan het wellicht zou lijken (Gillingham, 2016; Keddell, 2015). In het eerder aangehaalde Predictive Risk Model dat in Nieuw-Zeeland is onderzocht, is het model getraind op basis van wat in de data stond als zijnde ‘bevestigde gevallen’ (“substantiated cases”). Hoewel dit intuïtief een begrijpelijke keuze lijkt, blijkt dat onderzoek naar wat dit ‘bevestigd zijn’ inhoudt, minder helder is (Gillingham, 2016). Gillingham (2016) wijst erop de term in de verschillende gebruikte databases op verschillende wijzen gehanteerd, variërend van bewezen gevallen tot kinderen die risico lopen tot cases waarin wordt ingeschat dat het kind bescherming nodig heeft (Gillingham, 2016). Zijn deze data dan nog wel met elkaar te vergelijken? Dit voorbeeld toont aan hoe belangrijk het is om met zeer afgebakende, welomschreven en eenduidige voorspellende variabelen (“outcome variables”) te werken.

Naast operationalisatie is de kwaliteit van de algoritmes, in elk geval bij predictive analytics, ook gebaseerd op hoe kloppend de voorspellingen zijn. Geen enkel model is 100% accuraat en zal dus foutieve voorspellingen doen. Dit wordt aangeduid met vals positieven en vals negatieven en deze uitkomsten zijn inherent aan het gebruik van voorspellende modellen (Lanier et al., 2020). Vals positieve uitslagen in dit geval zijn voorspellingen waarin het model aangeeft dat er een hoog risico op mishandeling is, terwijl dit feitelijk niet gebeurt (of gaat gebeuren). Vals negatieven zijn het omgekeerde: er wordt een laag risico aangegeven, maar er is of zal wel degelijk sprake van mishandeling zijn (Church & Fairchild, 2017; de Haan & Connolly, 2014; Henman, 2020). In sommige gevallen zijn vals positieve of negatieve voorspellingen weinig problematisch. Ter illustratie, wie op de “play something”-knop op Netflix klikt en de film die wordt opgestart is niet conform de wensen, kan dit gemakkelijk stopzetten. Wanneer het echter gaat om foutieve voorspellingen bij kindermishandeling en huiselijk geweld, dan kan dit juist forse gevolgen hebben. Enerzijds kunnen kinderen die wel degelijk slachtoffer zijn over het hoofd worden gezien, anderzijds kunnen er ook negatieve gevolgen zijn voor families die onterecht als “positief” zijn aangemerkt (vals positieven). Een onterecht aangegeven hoge risicoscore kan invloed hebben op hoe het gezin door instanties wordt gezien en hoe men het gezin benadert. Hulpverleners zouden zich verplicht kunnen voelen om in te grijpen in het gezin en gezinnen kunnen zich verplicht voelen om hup te accepteren die men eigenlijk niet nodig heeft (de Haan & Connolly, 2014).

Ethische overwegingen en sociale rechtvaardigheid

Bovenstaand zijn de nodige programmatische aandachtspunten benoemd. De kwaliteit van zowel de gebruikte data als het algoritme dient goed te zijn om betrouwbare resultaten op te kunnen leveren. In discussies over het al dan niet implementeren van moderne digitale technieken binnen het sociale domein, wordt het ethische perspectief echter ook benadrukt (Church & Fairchild, 2017; de Haan & Connolly, 2014; Keddell, 2015, 2019; Lanier et al., 2020; Leslie et al., 2020). Het met een ethische bril kijken naar het mogelijke gebruik van moderne digitale technologieën beperkt zich niet alleen tot toepassingen binnen het sociale domein. Er zijn diverse ethische kaders ontwikkeld waarin richtlijnen of principes worden geformuleerd om op ethische wijze met digitale technologieën om te gaan. De United States Association for Computing Machinery (USACAM 2017) heeft zeven principes voor verantwoordelijkheid/aansprakelijkheid (“accountability”) ontwikkeld opdat het gebruik van big data en algoritmes transparant en verantwoord gebeurt (Gillingham, 2019b). Daarnaast bestaat er de richtlijn “Desirable Documentation for Algorithmic Transparency” en de door FAT/ML - waarbij de afkorting staat voor Fairness, Accountability, and Transparency in Machine Learning - ontwikkelde principes (Lanier et al., 2020). Ook Keddell (2019) onderstreept het belang van statistische eerlijkheid (“statistical fairness”) en koppelt hier nog de notie van sociale rechtvaardigheid aan welke in ogenschouw dient te worden genomen. Het gebruik van modellen als het gaat om het voorspellen van kindermishandeling dient sociale rechtvaardigheid te borgen en dit lijkt in de huidige vorm nog niet het geval te zijn (Keddell, 2019).

Wat de verschillende ethische richtlijnen met elkaar in gemeen hebben is dat er in elk geval een dringende oproep wordt gedaan dat gebruikte modellen eerlijk en transparant zijn. Er zijn bezwaren rondom privacy en het gebruiken van data in andere contexten zonder dat hier toestemming voor is gevraagd (Coulton et al., 2015; Keddell, 2019; Leslie et al., 2020). Bij gebruik van big data en machine learning liggen vooroordelen, stigmatisering en onbedoelde over- (of onder)representatie van bepaalde kwetsbare groepen op de loer (Church & Fairchild, 2017; Gillingham, 2019a; Leslie et al., 2020; Wilson, Tumen, Ota, &

Simmers, 2015). Dit kan zowel binnen zowel de gebruikte data(set) als het gebruikte algoritmische model. Doordat processen geautomatiseerd verlopen, kan dit stigmatiseren of versterken van vooroordelen worden verergerd, hetgeen zowel vragen kan oproepen over eerlijkheid als (sociale) rechtvaardigheid (Keddell, 2019).

Naast eerlijkheid is transparantie van belang. Hiermee wordt bedoeld dat het duidelijk is hoe een model tot zijn resultaten komt. Indien een gezin een hoog risicoprofiel heeft, is het van belang om te weten waarop dit profiel gebaseerd is. Het bieden van transparantie wordt echter om twee redenen bemoeilijkt. Enerzijds zijn het de ontwikkelaars die in sommige gevallen geen openheid van zaken willen geven over de werking van het model (Church & Fairchild, 2017; Gillingham, 2016). Dit is bij commerciële bedrijven niet ongebruikelijk, aangezien zij hebben geïnvesteerd in de ontwikkeling van het model (Church & Fairchild, 2017). Transparantie in machine learning is echter ook problematisch door de wijze waarop de modellen werken. Zoals reeds aangegeven is gaat het er bij machine learning om dat de computer zichzelf op basis van data en algoritmes traint om tot steeds betere resultaten te komen zonder tussenkomst van een programmeur. De gebruikte algoritmes in machine learning maken het soms zelfs voor een ervaren programmeur moeilijk om te weten wat er exact in de zogeheten “black box” gebeurt en dus om helder te hebben op basis waarvan een model tot een bepaalde uitkomst komt (Church & Fairchild, 2017).

Voorts roept het gebruik van deze technologieën ook vragen op rondom verantwoordelijkheid en aansprakelijkheid (“accountability”). Wie is er verantwoordelijk voor de resultaten of beslissingen genomen op basis van geïmplementeerde technologieën?

Lanier et al. (2019) wijzen in deze op het feit dat,

Algorithms and the data that drive them are designed and created by people – There is always a human ultimately responsible for decisions made or informed by an algorithm. “The algorithm did it” is not an acceptable excuse if algorithmic systems make mistakes or have undesired consequences, including from machine learning processes. – Fairness, Accountability and Transparency in Machine Learning (FAT/ML) ethical premise. (Lanier et al., 2019, p.1)

Als de mens uiteindelijk verantwoordelijk is en blijft voor zowel het ontwerp als de uitkomsten en of bepaalde beslissingen worden genomen, dan dient hij hiertoe wel in de gelegenheid te zijn. Door kritisch te blijven ten aanzien van het gebruik van moderne digitale technologieën, zeker als het gaat om onderzoek naar mishandeling, is het wellicht mogelijk deze te implementeren mits de kwaliteit van data en algoritmes cq. modellen geborgd zijn. Technologie neemt het werk en de beslissingsbevoegdheid echter niet over van de professional. Het kan een hulpmiddel zijn, niet een vervanging van de sociaalwerker. Het inzetten van moderne digitale technologieën vraagt echter wel om nieuwe competenties, waar ook de nodige aandacht voor dient te zijn.

De professional: kennis over moderne digitale technologieën en het nemen van beslissingen

Naast bovengenoemde praktische en ethische bezwaren is, liggen er aandachtspunten rondom de kennis van de professional. Zoals eerder aangegeven, zijn het domein van sociaal werk en dat van de digitale technologie nog gescheiden. Als ervoor kan worden gezorgd dat zowel data kwaliteit van data en algoritmes kan worden geborgd en de ethische overwegingen voldoende in ogenschouw worden genomen, dan nog is het van belang dat professionals voldoende geschoold zijn om met de implementatie van dergelijke technologieën om te kunnen gaan en zich – eventueel kritisch – hiertoe te verhouden (Lanier et al., 2020). Dit tot op heden niet het geval (Coulton et al., 2015; Gillingham, 2019b). Met

het oog op transparantie en verantwoordelijkheid (accountability) is het van belang dat de professional begrijpt hoe bepaalde resultaten tot stand zijn gekomen en hoe deze invloed op zijn of haar beslissing (kunnen) hebben. Dat de professional nog niet is opgeleid in het gebruik van en zich verhouden tot technologieën, betekent echter niet dat men niet in staat is om dit al te kunnen. De-Arteaga, Fogliato en Chouldechova (2020) hebben onderzoek gedaan naar de vraag hoe professionals reageren wanneer zij geconfronteerd worden met foutieve algoritmische scores en naar aanleiding hiervan beslissingen dienen te nemen. Hun conclusie is dat professionals wel degelijk in staat zijn om de uitkomsten van een algoritmisch proces naast zich neer te leggen (De-Arteaga, Fogliato, & Chouldechova, 2020). Dit betekent echter wel dat professionals in de gelegenheid moeten zijn om autonoom te kunnen handelen. Zij zijn immers, al dan niet in een team, degene bij wie de eindbeslissing ligt en niet bij de ingezette technologie.

Conclusie

Dit artikel heeft als doelstelling inzicht te geven in de vraag welke potentie het gebruik van moderne digitale technologieën heeft als het gaat om het verminderen of voorkomen van kindermishandeling en het onderzoek hiernaar. Daarnaast is stilgestaan bij de vraag aan welke voorwaarden het gebruik dient te voldoen, waar uitdagingen liggen en welke ethische overwegingen in ogenschouw dienen te worden genomen. Er kan worden gesteld dat de technologieën zeker potentie hebben. We leven in een gedigitaliseerde wereld waarin we toegang hebben tot veel (digitale) data. De komst van big data en machine learning maakt het mogelijk om deze data met elkaar te verbinden en hieruit te leren, of dit nu om nieuwe patronen gaat (descriptive analytics) of voorspellingen over de toekomst (predictive analytics). Dit geldt ook voor toepassingen en/of onderzoek binnen het sociale domein en zo naar kindermishandeling en huiselijk geweld. Op dit moment zijn er echter nog de nodige programmatische uitdagingen als het gaat om zowel de kwaliteit van de data als dat van de gebruikte algoritmes. Met name vooroordelen en stigma lijken hier punt van aandacht vanwege het gebruik van onvolledige en niet-representatieve datasets en mogelijk onbewust invoeren van vooroordelen in keuzes bij het ontwerpen van algoritmische modellen. Het gebruik van big data en machine learning maakt het binnen het sociale domein, zeker als het gaat om het mogelijk voorspellen van kindermishandeling en huiselijk geweld, noodzakelijk om rekening te houden met ethische overwegingen, vraagstellingen zoals bijvoorbeeld worden gegeven bij de zeven principes van algoritmische verantwoordelijkheid of bij de FAT/ML-principes. Indien dit in onvoldoende mate gebeurt, kan dit leiden tot sociale onrechtvaardigheid en het verder marginaliseren van kwetsbare groepen. Het mogelijke gebruik of implementeren van deze technologieën is dus een serieuze aangelegenheid, zeker wanneer het gaat over een thema als kindermishandeling en de mogelijke gevolgen voor de gezinnen wiens data gebruikt worden. Mogelijke implementatie veronderstelt dat professionals globale kennis hebben van de werking van – bijvoorbeeld – big data en machine learning, opdat ze zich kritisch kunnen verhouden tot deze materie. Een kritische houding zorgt er tevens voor dat de professional zijn autonomie kan behouden en zo de verantwoordelijkheid kan dragen voor de beslissingen die hij mede op basis van door moderne digitale technologieën ingegeven resultaten neemt.

Literatuur

- Blauw, S. (2021). *Het bestverkochte boek ooit*. Amsterdam: de Correspondent.
- Cash, S. J. (2019). Child Maltreatment and Technology. *International Journal of Child and Adolescent Resilience / Revue internationale de la résilience des enfants et des adolescents*, 6(1), 45-51.
doi:<https://doi.org/10.7202/1069075ar>
- Church, C. E., & Fairchild, A. J. (2017). In Search of a Silver Bullet: Child Welfare's Embrace of Predictive Analytics. *Juvenile and Family Court Journal*, 68(1), 67-81.
doi:<https://doi.org/10.1111/jfcj.12086>
- Coulton, C. J., Goerge, R., Putnam-Hornstein, E., & de Haan, B. (2015). Harnessing big data for social good: A grand challenge for social work. *Cleveland: American Academy of Social Work and Social Welfare*, 1-20.
- De-Arteaga, M., Fogliato, R., & Chouldechova, A. (2020). A case for humans-in-the-loop: Decisions in the presence of erroneous algorithmic scores. Paper presented at the Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.
- de Haan, I., & Connolly, M. (2014). Another Pandora's box? Some pros and cons of predictive risk modeling. *Children and Youth Services Review*, 47, 86-91.
doi:<https://doi.org/10.1016/j.childyouth.2014.07.016>
- Eynikel, J. (2017). *Robot aan het stuur*. Leuven: Lannoo Campus.
- Gillingham, P. (2016). Predictive Risk Modelling to Prevent Child Maltreatment and Other Adverse Outcomes for Service Users: Inside the 'Black Box' of Machine Learning. *British Journal of Social Work*, 46(4), 1044-1058. doi:10.1093/bjsw/bcv031
- Gillingham, P. (2019a). Can Predictive Algorithms Assist Decision-Making in Social Work with Children and Families? *Child Abuse Review*, 28(2), 114-126.
doi:<https://doi.org/10.1002/car.2547>
- Gillingham, P. (2019b). Decision support systems, social justice and algorithmic accountability in social work: A new challenge. *Practice*, 31(4), 277-290.
- Hardyns, W., & Snaphaan, T. (2020). *Big data en innovatieve methoden voor criminologisch onderzoek*. Den Haag: Boom criminologie.
- Henman, P. (2020). Improving public services using artificial intelligence: possibilities, pitfalls, governance. *Asia Pacific Journal of Public Administration*, 42(4), 209-221.
- Hunt, X., Tomlinson, M., Sikander, S., Skeen, S., Marlow, M., du Toit, S., & Eisner, M. (2020). Artificial Intelligence, Big Data, and mHealth: The Frontiers of the Prevention of Violence Against Children. *Frontiers in artificial intelligence*, 3, 543305.
- Keddell, E. (2015). The ethics of predictive risk modelling in the Aotearoa/New Zealand child welfare context: Child abuse prevention or neo-liberal tool? *Critical Social Policy*, 35(1), 69-88.
doi:<https://doi.org/10.1177/0261018314543224>
- Keddell, E. (2019). Algorithmic justice in child protection: Statistical fairness, social justice and the implications for practice. *Social Sciences*, 8(10), 281.
- Lanier, P., Rodriguez, M., Verbiest, S., Bryant, K., Guan, T., & Zolotor, A. (2020). Preventing infant maltreatment with predictive analytics: Applying ethical principles to evidence-based child welfare policy. *Journal of family violence*, 35(1), 1-13.
- Leslie, D., Holmes, L., Hitrova, C., & Ott, E. (2020). Ethics Review of Machine Learning in Children's Social Care. Leslie, D., Holmes, L., Hitrova, C. & Ott, E. (2020). *Ethics review of machine learning in children's social care*. [Report] London, UK: What Works for Children's Social Care.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*: Crown.
- Russell, J. (2015). Predictive analytics and child protection: Constraints and opportunities. *Child Abuse & Neglect*, 46, 182-189. doi:<https://doi.org/10.1016/j.chiabu.2015.05.022>
- Verbeek, P.-P. (2014). *Op de vleugels van Icarus*. Rotterdam: Lemniscaat.
- Wilson, M. L., Tumen, S., Ota, R., & Simmers, A. G. (2015). Predictive modeling: potential application in prevention services. *American journal of preventive medicine*, 48(5), 509-519.
- WorldHealthOrganization. (2006). *Preventing child maltreatment : a guide to taking action and generating evidence / World Health Organization and International Society for Prevention of Child Abuse and Neglect*. In. Geneva: World Health Organization.