

# Wat kan artificiële intelligentie betekenen voor de kwaliteit van de forensische advisering?

*Margriet Mutsaers en Maaïke Kempes\**

In het strafproces en de tenuitvoerlegging wordt op meerdere momenten geadviseerd over de persoon van de verdachte of de veroordeelde. Denk bijvoorbeeld aan de rapportage pro Justitia, over de toerekenbaarheid van een delict, en de indicatiestelling voor forensische zorg. Het Nederlands Instituut voor Forensische Psychiatrie en Psychologie (NIFP), een landelijke dienst van de Dienst Justitiële Inrichtingen (DJI) van het Ministerie van Justitie en Veiligheid, heeft op beide gebieden coördinatie- en uitvoeringstaken. Daarbij gaat het steevast om doelgroepen met complexe psychische problematiek en/of ernstige delicten.

Genoemde adviezen kunnen grote gevolgen hebben voor betrokkenen en maatschappij; bovendien zijn rechtsgelijkheid en rechtsbescherming belangrijke uitgangspunten. Daarom is van belang dat deze advisering zorgvuldig wordt uitgevoerd. Naast methodische werkwijzen, meer-ogen-beoordeling en periodieke deskundigheidsbevordering is het NIFP dan ook altijd op zoek naar mogelijkheden tot verdere objectivering van zijn advisering.

Met de opmars van artificiële intelligentie (AI) is de vraag gerezen of en hoe deze technologie kan bijdragen aan de kwaliteitsverbetering van de forensische advisering. Enerzijds wordt dan gedacht aan vervanging van de menselijke handeling. Anderzijds zoekt men juist naar aanknopingspunten om de mens efficiënter en effectiever te kunnen inzetten. Daarbij wordt met een schuin oog gekeken naar sectoren waar AI al succesvol wordt ingezet. Het gaat dan vaak om situaties waarin relatief eenvoudige beslissingen worden genomen, zoals het

\* Drs. M. Mutsaers is senior beleidsadviseur bij het Nederland Instituut voor Forensische Psychiatrie en Psychologie (NIFP). Prof. M. Kempes is hoofd van de afdeling Wetenschap en Opleidingen bij het Nederland Instituut voor Forensische Psychiatrie en Psychologie (NIFP). Zij is tevens bijzonder hoogleraar forensische neuropedagogiek aan de Universiteit Leiden.

verzekeringswezen (Śmietanka e.a. 2021), en aspecten van de gezondheidszorg, zoals de interpretatie van radiologische scans (Yu e.a. 2018).

In dit artikel gaan we in op de ervaringen met een exploratief onderzoek naar de mogelijkheden van door TNO ontwikkelde programmatuur voor de ondersteuning van de indicatiestelling voor klinische forensische zorg door het NIFP. Dat betreft een relatief overzichtelijk proces, op basis van reeds bestaande diagnostiek. Het doel was om daarvoor een prototype van een beslissingsondersteunend instrument te formuleren. Dit zou niet alleen opzoekwerk kunnen overnemen van de indicatiestellers, maar ook hun beoordeling moeten faciliteren. Om de uitkomsten van het project te kunnen begrijpen, gaan we eerst in op de taakstelling van indicatiestelling klinische forensische zorg en de aard van de beslissing, en daarna op de opzet, randvoorwaarden en uitvoering van het onderzoek. Vervolgens worden de opbrengst en lessen besproken. In de slotparagraaf beschrijven we hoe deze lessen kunnen worden verbreed naar de toepassing van AI in andere beslissingsprocessen in het justitiële veld en de ontwikkeling daarvan.

## **Indicatiestelling klinische forensische zorg door het NIFP**

### *Taakstelling*

Forensische zorg is door de rechter opgelegd door middel van een zogenaamde forensische zorgtitel, zoals een voorwaardelijke veroordeling en een maatregel Inrichting Stelselmatige. Vereiste is dat er sprake is van zogenaamde forensische relevantie: een relatie tussen het delictgedrag en psychische problematiek. De aanname is dat behandeling en begeleiding van deze problematiek recidive zal tegen gaan. Daarmee is sprake van een tweede voorwaarde voor oplegging: de problematiek kan worden aangepakt, door beïnvloeding van uitlokkende risicofactoren en met gebruikmaking van beschermende factoren.

Forensische zorg kan zowel klinisch en ambulante als in een instelling voor begeleid wonen plaatsvinden en toeleiding en uitvoering zijn bij wet geregeld.<sup>1</sup> Om belangenverstrengeling te voorkomen is functie-

1 Wet forensische zorg, *Stb.* 2018, 38.

scheiding het organiserende principe: indicatiestelling, plaatsing, zorgverlening, toezicht en inkoop moeten onafhankelijk van elkaar en objectief geschieden. Daarbij is het uitgangspunt passende én gepaste zorg: aansluitend bij de te behandelen of te begeleiden problematiek en niet zwaarder (duurder) dan noodzakelijk.

Ook de reclassering en de multidisciplinaire zorgteams in de justitiële inrichtingen indiceren voor forensische zorg, maar alleen het NIFP adviseert voor de klinische variant, te weten opname in een behandelvoorziening,<sup>2</sup> en daarmee de zwaardere doelgroepen. De indicatiestelling is de basis voor de matching met een geschikte zorginstelling, het plaatsingsbesluit en de start van de zorg. Na zes weken is de zorgaanbieder verplicht om een behandel- of begeleidingsplan te formuleren.

### *Uitgangspunten*

In het geval van klinische forensische zorg adviseert de indicatiestelling over de gewenste verblijfsintensiteit en het noodzakelijke beveiligingsniveau. De Nederlandse Zorgautoriteit (NZa) heeft beide begrippen geoperationaliseerd en daarbinnen verschillende niveaus afgebakend. Zo gaat het bij ‘verblijfsintensiteit’ onder meer om het leefklimaat en de inzet van personeel en bij ‘beveiligingsniveau’ om de locatie van het behandelaanbod – binnen of buiten de beveiligde ring – en drugs- en kamercontroles.<sup>3</sup>

Het NIFP indiceert voor de zwaardere zorg. Van de zeven verblijfsintensiteiten, uitgedrukt in letters, indiceert het NIFP voor de meest intensieve vier; van de zes beveiligingsniveaus, uitgedrukt in cijfers, indiceert het NIFP eveneens voor vier. Indicaties voor klinische forensische zorg zijn dus combinaties als D/1 en F/2.2 (zie tabel 1).

### *Werkwijze*

Het NIFP indiceert op basis van een dossier van zogenaamde bron-documenten: per forensische zorgtitel is benoemd welke bron-documenten nodig zijn, zoals tenlastelegging of vonnis en reclasserings-

2 Besluit forensische zorg, *Stb.* 2019, 230 (gewijzigd *Stb.* 2021, 251 en *Stb.* 2021, 158).

3 Regeling dbbc's, zzp's en extramuraal parameters forensische zorg, NR/REG-2022a, *Stcrt.* 2020, 1188.

**Tabel 1** Zorgprestaties forensische zorg, met uitgelicht de typen waarvoor het NIFP indiceert

Beveiligingsniveau Verblijfsintensiteit	Geen (0)	Laag (1)	Gemiddeld- laag (2.1)	Gemiddeld- hoog (2.2)	Hoog (3)	Ze er hoog (4)
Licht (A)						
Beperkt (B)						
Matig (C)						
Gemiddeld (D)						
Intensief (E)						
Extra intensief (F)						
Ze er intensief (G)						

informatie. Omdat het gaat om het actuele toestandsbeeld en de (eventuele) aanpak van de actuele situatie, moeten gedragsdeskundige beoordelingen als rapportages pro Justitia, behandelverslagen en risicotaxaties relatief recent zijn. Het dossier wordt aangeleverd door de aanvragers van de indicatiestelling, ten behoeve van advisering aan de rechter of voor behandeling buiten de justitiële inrichting.

De twintig indicatiestellers verwerken de brondocumenten systematisch aan de hand van een checklist met zo'n vijftig mogelijke dynamische eigenschappen van een casus. Het is immers de bedoeling dat de te indiceren forensische zorg het delictgedrag positief beïnvloedt. Deze lijst is door middel van *conceptmapping* door professionals tot stand gekomen, getoetst aan wetenschappelijke literatuur en met behulp van statistische technieken onderverdeeld in clusters van kenmerken (Mulder e.a. 2014). Items als psychische problematiek en behandelmotivatie verwijzen naar de te adviseren verblijfsintensiteit; kenmerken als criminogene factoren en schending van voorwaarden refereren aan het te indiceren beveiligingsniveau. Hebben de eigenschappen een relatie met het delictgedrag waarvan een justitiabele wordt verdacht of waarvoor hij/zij is veroordeeld, dan worden zij als 'kritisch' aangemerkt: te beïnvloeden door de te adviseren klinische zorg.

### *Indicatiestellingsbrief*

Een indicatiestelling is dus een onderbouwde afweging. Dit is terug te zien in de opbouw van de indicatiestellingsbrief. Deze bestaat uit een sectie ‘Overwegingen’, met de als kritisch gewaardeerde eigenschappen van de casus, gevolgd door een sectie ‘Conclusie’, met een advies voor een verblijfsintensiteit en een beveiligingsniveau. Beide zijn het resultaat van een door een format geleid overzicht van de relevante aspecten in de levens-, hulpverlenings-, delict- en (forensische) behandelgeschiedenis.

### *Overeenkomstige beoordeling*

Met het oog op rechtsgelijkheid en rechtsbescherming is van belang dat de indicatiestelling niet afhankelijk is van de persoon van de indicatiesteller. Naast best practices, checklists, formats en periodieke scholing wordt dit bevorderd door driemaandelijks zogenaamde consensusbesprekingen. Dan beoordelen alle indicatiestellers dezelfde complexe casus en bespreken zij hun interpretaties en afwegingen. Over deze casussen tezamen wordt jaarlijks de mate van overeenkomstige beoordeling bepaald.

Doel van het project was te onderzoeken in hoeverre het mogelijk was het indicatiestellingsproces te faciliteren en de overeenkomstige beoordeling te bevorderen met behulp van AI-software. Concreet werd daarbij gedacht aan een adviessysteem dat een suggestie zou doen op basis van in het verleden gestelde indicaties bij vergelijkbare casuïstiek.

### **Randvoorwaarden**

Voordat de meerwaarde van (de beschikbare) AI kon worden onderzocht, moest zijn voldaan aan enkele randvoorwaarden op het gebied van met name de privacybescherming en de instructie van de programmatuur. We bespreken ze hier kort omdat ze relevant zijn voor de ontwikkeling van bruikbare AI.

### *Data Protection Impact Assessment*

Om de haalbaarheid van een adviessysteem te kunnen vaststellen, moest worden gewerkt met gerealiseerde indicatiestellingen. En omdat daarin medische gegevens en andere persoonsgegevens zijn opgenomen, is allereerst een Data Protection Impact Assessment (DPIA) opgesteld.

Hieruit kwam als belangrijk knelpunt naar voren dat de (ex-)justitiabelen van de te gebruiken indicatiestellingen daarvan niet op de hoogte waren en daardoor hun rechten niet konden uitoefenen. Omdat ex-justitiabelen vaak niet meer zijn te lokaliseren, was achteraf om toestemming vragen echter niet aan de orde. Bovendien zou dit verzoek een inbreuk op hun privacy kunnen maken wanneer de omgeving niet bekend was met het justitieverleden. Laatstgenoemde overweging wordt ook door de Algemene verordening gegevensbescherming (AVG) gezien als zwaarwegend belang om betrokkenen niet te informeren. Bovendien waren de risico's laag, mede door de gesloten Data Transfer Agreement.

### *Pseudonimisering en instructie systeem*

Om de indicatiestellingen onherleidbaar te maken was het uitgangspunt dat de namen van justitiabelen, andere betrokken personen, plaatsen, instanties en instellingen zouden worden verwijderd. Hetzelfde gold voor data waaraan strafzaken zouden kunnen worden herkend en adressen. Tegelijk moest de relevante informatie voor het onderzoek behouden blijven, zoals de aard van de betrokken organisaties en de eigenschappen van de casus volgens ten minste de Checklist.

Dat betekent dat algemene aanduidingen als 'Forensisch-Psychiatrische Afdeling' en 'verslavingskliniek' moesten worden gehandhaafd, terwijl specificaties als eigennamen moesten worden verwijderd. Gebruikt als toevoeging ging daarmee geen informatie verloren, maar op zichzelf genoemd moest vertaling plaatsvinden. De beschikbare AI begreep namelijk niet zomaar wat 'de Jellinek' en 'Exodus' zijn, laat staan de eigennamen van plaatselijke en regionale instellingen. Hetzelfde geldt voor de vele afkortingen in de justitiewereld, het forensisch domein, de (geestelijke) gezondheidszorg en het maatschappelijk middenveld. Om deze informatie begrijpelijk te maken voor geau-

tomatiseerde verwerking zijn daarom lijsten aangelegd van betekenissen van begrippen, eigennamen en afkortingen.

Als eerste is hiermee een bestaande tool van TNO gevoed, waarmee de te gebruiken indicatiestellingsbrieven zijn gepseudonimiseerd. In vijf rondes van toetsing en aanvullende programmering is deze tool maximaal aangepast. Dat leverde relatieve tijdwinst op ten opzichte van handmatige pseudonimisering: het resultaat noopte tot integrale screening achteraf. Hierbij zijn resterende (eigen)namen verwijderd of vertaald, is onterecht verwijderde informatie teruggezet, zijn de lijsten met betekenissen aangevuld, en zijn in het geval van aanbieders van hulpverlening en zorg met een breed aanbod de genoten begeleiding en behandeling gespecificeerd.

## **Opzet en uitvoering**

### *Fasering*

Voor de ondersteuning van het beslisproces zou het te ontwikkelen adviessysteem zich moeten kunnen baseren op de brondocumenten voor de indicatiestelling. Om de AI beter te kunnen laten leren, is de exploratie van de mogelijkheden echter eerst uitgegaan van gerealiseerde indicatiestellingsbrieven. Hierin is de broninformatie verwerkt en zijn de kritische kenmerken reeds onderkend en conclusies getrokken. Als de TNO-programmatuur deze niet zou kunnen herkennen en combineren, zou bovendien de moeite kunnen worden bespaard van een voor de software ingewikkelder opgave: met meerdere brondocumenten per casus en nog ongekwalificeerde kenmerken.

In eerste instantie en bij wijze van beproeving van de AI is deze toegepast op zestien geselecteerde indicatiestellingsbrieven – één per mogelijke indicatie (Schadd e.a. 2022). De uitkomst is vervolgens getoetst aan vier willekeurige ‘echte’ indicatiestellingen. Daarmee kon sowieso niets worden gezegd over adequaatheid en betrouwbaarheid, maar de programmatuur leek wel in staat de indicatiestellingsbrieven te lezen op het niveau van woorden en begrippen.

Om te voorkomen dat het te ontwikkelen systeem op verkeerde gronden werkzaam zou worden geacht – een bedreiging van de integriteit van de uitvoeringspraktijk –, is de steekproef vergroot naar driehonderd random geselecteerde, gerealiseerde indicatiestellingsbrieven.

Daarbij is bewust gezocht naar een aantal waarmee met zo min mogelijk pseudonimiseringsinspanning de onderzoeksvragen zouden kunnen worden beantwoord. Om de slaagkansen te optimaliseren heeft het onderzoek met name gebruik gemaakt van de reeds op de kernzaken toegespitste secties ‘Overwegingen’ en ‘Conclusie’ van de indicatiestellingsbrieven.

### *Misverstand*

Aanvankelijk ging het onderzoek uit van indicatiestelling door rechtstreekse overname uit de brondocumenten van een beperkt aantal standaard omschreven kenmerken van personen, die in vaststaande combinaties – elkaar uitsluitend en oplopend naar zwaarte – staan voor een van de zestien mogelijke klinische indicaties. In werkelijkheid echter is het beslisproces complexer, zijn patiënten met dezelfde psychische problematiek geen homogene groep, en zijn er geen ‘enige echte’ uitkomsten als van een berekening.

De indicatiestelling voor klinische forensische zorg door het NIFP kent namelijk meerdere stappen, te beginnen met de detectie van relevante eigenschappen van de actuele situatie. Referentiekader zijn weliswaar de items van de Checklist, maar om de indicatiesteller in staat te stellen overige voor de casus belangrijke kenmerken te signaleren, kent deze meerdere onbepaalde rubrieken ‘anders’. Vervolgens moeten deze eigenschappen worden geïnterpreteerd om hun forensische relevantie te kunnen vaststellen, en tot slot moet het geheel aan kritische kenmerken worden gewogen. Daarbij telt niet het aantal, maar hun relatieve belang in de specifieke context van de casus en de andere kritische kenmerken. Ofwel: in het ene geval kunnen tien kritische eigenschappen een indicatiesteller leiden tot het advies E/2.1, terwijl in het andere geval drie kritische kenmerken genoeg zijn voor de indicatie G/3. Ook kan een indicatiesteller bij dezelfde combinaties verschillend adviseren.

Dat verklaart waarom eliminatie van een veelvoorkomend kenmerk als verslavingsproblematiek, een belangrijk uitgangspunt van de gebruikte programmatuur, geen geschikte weg bleek. Weliswaar onderscheidt een veelvoorkomende eigenschap op zichzelf niet de ene casus van de andere, maar in de specifieke context van de andere kenmerken heeft zij een bepaald belang, zoals zij op haar beurt bijdraagt aan dat van de andere eigenschappen.



### *Spanningsveld*

Anders gezegd: voor een adequate afweging moet methodische beoordeling plaatsvinden van juist zo veel mogelijk andere relevante informatie. Dit verdraagt zich echter slecht met belangrijke voorwaarden voor geautomatiseerde verwerking, zeker bij een relatief beperkte set van documenten: op basis van welomschreven labels met ieder een beperkt waardebereik.

Om hieraan tegemoet te komen is het waardebereik per item van de Checklist zo veel mogelijk teruggebracht tot twee of drie waarden: samen het domeinmodel voor het onderzoek. Hoewel de vraag is of een dergelijke vereenvoudiging recht doet aan de werkelijkheid, hoeft deze niet a priori een probleem te zijn. Het beoogde adviessysteem was bedoeld op basis van vergelijkingen, met het professionele oordeel van de indicatiesteller als doorslaggevende factor.

Wel is in een workshop getoetst aan welke kenmerken, mits kritisch, indicatiestellers meer gewicht toekennen. Zo bleken zij bijvoorbeeld 'verslaving' en 'verstandelijke beperking' relatief significant te vinden. Deze voorkeuren zijn verwerkt in de gebruikte analysetechnieken, waarmee het domeinmodel is aangescherpt. Dit zou moeten bijdragen aan de beoogde zoekfunctie naar vergelijkbare casuïstiek.

### *Onderzoeksvragen*

Conform de ambitie van de ontwikkeling van een instrument ter facilitering van zowel het handwerk van de indicatiestellers als hun beoordeling is achtereenvolgens onderzocht in hoeverre de met forensische domeinkennis verrijkte TNO-software:

1. indicatiestellingsbrieven kon inlezen en vertalen naar relevante eigenschappen van de casus;
2. kon bepalen welke kenmerken een relatie hadden met het indexdelict (kritisch waren);
3. vergelijkbare historische casuïstiek kon vinden; en
4. een adequate indicatie voor klinische forensische zorg kon adviseren.

Deze stappen zijn in deze volgorde voorwaardelijk voor elkaar. Daarbij zouden de eerste twee de doelmatigheid van het beslisproces dienen en zou het totaal de overeenkomstige beoordeling bevorderen.

Het project bestond erin om voor elk van deze stappen maten te definiëren voor de objectieve bepaling van de prestatie en de prestatie-niveaus vast te stellen. Hierbij stond de technische haalbaarheid voorop. Afhankelijk daarvan zou in een later stadium de meerwaarde in de praktijk worden getoetst.

## Resultaten

### *Detectie relevante kenmerken*

Om de juiste begrippen te kunnen vinden is AI gebaat bij standaard schrijfwijzen. Brondocumenten en indicatiestellingen zijn echter geschreven in de natuurlijke taal van vele auteurs, reden waarom de omschrijvingen veelal een grote variëteit kennen. Bovendien zijn de vaak conceptuele kenmerken van de Checklist niet altijd eenduidig gedefinieerd dan wel te definiëren. Zo werden in de driehonderd gebruikte indicatiestellingen ‘justitiële voorgeschiedenis’ en ‘gezinssituatie’ vaak verhalend beschreven, kennen ‘verstandelijke beperking’ en ‘huisvesting’ vele verschijningsvormen en werden ‘zorgmijdingsgedrag’ en ‘procriminele houding’ zelden aldus benoemd. Ook is de context vaak complex en genuanceerd. Is momenteel of was nou juist in het verleden sprake van middelenmisbruik, in welke mate, en betreft het de justitiabele zelf of was het toch zijn vader? Kan AI vaak nog wel uit de voeten met spelfouten en ontkenningen, per saldo was de conclusie dat de reguliere inleesmethodes de relevante kenmerken van de indicatiestelling klinische forensische zorg niet goed genoeg kunnen inlezen. Omdat er AI bestaat met de mogelijkheid van opsporing van ook synoniemen en meer conceptuele begrippen, was de aanbeveling deze te beproeven. Een belangrijke voorwaarde is dan wel een dataset van minimaal enkele tienduizenden (gepseudonimiseerde) indicatiestellingen.

### *Bepaling kritische kenmerken*

Om deze mogelijkheid te toetsen is uitgegaan van de aanname dat relatief vaak genoemde kenmerken in de sectie ‘Overwegingen’ van de indicatiestellingsbrief forensisch relevant (kritisch) zijn voor de te indiceren zorg. Omdat voor de bepaling daarvan ook de context

belangrijk is, is – weliswaar gegeven de geconstateerde onvoldoende inleesprestatie – gekozen voor geautomatiseerde samenvatting van de veelvoorkomende begrippen. Wanneer deze aanpak zou werken, zou toepassing op de onderbouwing van de indicatiestelling reeds een basis kunnen leveren voor de overwegingen waarop het advies wordt bepaald.

In het onderzoek is precies dat gedaan, waarbij de items in de door de AI gegenereerde samenvatting zijn vergeleken met de overwegingen van de indicatiestellers. Daarbij is gekeken naar de mate van overeenkomst en die van onterechte overname en weglating in de samenvatting. De conclusie moest zijn dat deze methode niet voldoet.

Aandachtspunten voor eventuele alternatieve AI zijn: zekerheid over werkzaamheid op de goede gronden, bruikbaarheid voor (ook) de Nederlandse taal en plaatsing van de indicatiestellingen op buitenlandse servers vanwege de grote benodigde rekenkracht. Deze voorwaarden zijn niet zomaar gerealiseerd en daarom was de aanbeveling deze stap (nog) in handen te laten van de indicatiestellers.

### *Vinden vergelijkbare casuïstiek*

Voor de toetsing van de mogelijkheid van vergelijking met historische indicatiestellingen kon dus niet worden teruggevallen op door de AI bepaalde kritische kenmerken. Om deze onderzoeksvraag toch te kunnen beantwoorden heeft TNO de sectie ‘Overwegingen’ van de indicatiestellingsbrieven in de steekproef handmatig gelabeld. Dit is gebeurd volgens het domeinmodel voor het onderzoek en met het domeinbegrip van dat moment.

Voor de opsporing van vergelijkbare casuïstiek in de database van het onderzoek bleek dat de methode van ‘gewogen tellen’ het beste resultaat gaf en goed presteerde. Hiermee wordt niet alleen bezien of reeds als kritisch aangemerkte eigenschappen overeenkomen, maar ook hun positie in de volgorde van toegekend belang. Een belangrijke voorwaarde is wel dat iedere indicatiestellingsbrief moet zijn voorzien van een herkenbaar profiel van kritische kenmerken (meerwerk).

### *Advisering*

Ook wat betreft de mogelijkheid van geautomatiseerde advisering van de gewenste verblijfsintensiteit en het noodzakelijke beveiligingsni-

veau is gebruik gemaakt van de door TNO handmatig gelabelde dataset. Overigens zonder gebruikmaking van de clustering in domeinen van de items volgens de Checklist is gekeken naar het verband tussen het samenstel van als kritisch aangemerkte kenmerken en de in werkelijkheid gestelde indicaties. Daarvan bleek de voorspelling onvoldoende, waarbij – conform de jaarlijkse bepaling door het NIFP van de mate van overeenkomstige beoordeling door de indicatiestellers – die van het beveiligingsniveau beter was dan die van de verblijfsintensiteit.

In een poging de prestatie van de AI te verbeteren zijn kleine aanpassingen van het domeinmodel doorgevoerd. Zo was de voorspelling beter bij beschouwing van de waarde ‘trekken van’ als persoonlijkheidsstoornis en toevoeging van de vaak genoemde aandachtsstoornis. Aanscherping van het domeinmodel leek dus potentie te hebben, zij het onder de voorwaarden dat deze beperkt zou blijven, forensische relevantie zou hebben en voldoet aan de gedragsdeskundige standaard. Dat was dan ook de aanbeveling, naast nader onderzoek naar methoden voor betere voorspelling van met name de verblijfsintensiteit.

### *Beschouwing*

De detectie van relevante en de bepaling van kritische kenmerken zijn onvoldoende haalbaar gebleken. Dit geldt alleen al technisch en na verwerking van de brondocumenten voor de indicatiestellingsbrieven. De conclusie moet daarom zijn dat de doelmatigheid van dit proces met toepassing van de beschikbare AI niet zal kunnen worden verbeterd. Daarbij vraagt de exploratie van de meerwaarde van alternatieve technieken zoveel extra inspanning over langere tijd dat deze zich slecht verdraagt met de onzekere uitkomst.

De technische haalbaarheid van geautomatiseerde suggesties voor klinische indicaties, zoals de ambitie was, lijkt weliswaar groter, maar ook hiervoor zijn grote investeringen nodig. Zo kan deze alleen aan de orde zijn bij wetenschappelijk getoetste (statistische) aanpassing van de Checklist en de permanente extra inspanning van handmatige labeling van de indicatiestellingsbrieven.

Naast de aanleg van een substantiële database is het laatste tevens een voorwaarde voor de meest kansrijk gebleken toepassing van de AI: de vergelijking met eerdere indicatiestellingen. Meerwaarde hiervan is

uitbreiding van het referentiekader van de indicatiesteller, die mogelijk overeenkomstige beoordeling bevordert. Gezien de functie van de indicatiestelling en de tijdelijkheid daarvan bij een voor diagnostiek acceptabele betrouwbaarheid tussen indicatiestellers (Kraemer 2014), is de meerwaarde voor de praktijk niet zonder meer overtuigend. Deze relativeringen nemen niet weg dat het project zijn opbrengst heeft. Door de analyse van de eenduidigheid van de items en de ont-sluiting van de rubrieken 'anders' heeft het NIFP suggesties gekregen voor aanscherping van de Checklist. En doordat gebruik is gemaakt van de structuur van de indicatiestellingsbrief zijn er aandachtspunten gebleken voor kwaliteitsverbetering van het product. DJI ten slotte is wijzer geworden wat betreft zijn verwachtingen van de automatisering van zijn meer complexe beslisprocessen.

### **Algemene lessen**

Uit dit onderzoek leren we ook lessen over welke randvoorwaarden belangrijk zijn voor het kunnen inzetten (of onderzoeken van de inzet) van AI rondom beslisprocessen in het justitiële veld.

*AI is succesvol als er veel informatie beschikbaar is*

Bevestigd is de algemeen bekende notie dat AI meer kan betekenen als er veel input aanwezig is waarvan de systemen kunnen leren. Daarbij moet men eerder denken aan honderdduizenden documenten dan de paar honderd die zijn gebruikt in dit project. Dat geldt bij uitstek voor beslissingen op basis van verschillende combinaties van tientallen kenmerken, die ook nog eens meerdere gradaties of verschijningsvormen kennen en moeten worden gewogen ten opzichte van elkaar, zoals de indicatiestelling voor klinische forensische zorg. In het onderzoek is het onmogelijk gebleken een dergelijke rijkdom aan relevante informatie tot zijn recht te laten komen en een model te genereren dat op betrouwbare wijze de werkelijkheid reflecteert.

Bij de hoeveelheid aan documenten in het justitiële veld spelen meerdere factoren een rol. Ten eerste gaat het om zeer gevoelige informatie. Voor het testen en bouwen van een AI-applicatie moet men met geanonimiseerde data werken. Het is ondoenlijk honderdduizenden records voor dit doel te anonimiseren/pseudonimiseren. Een tweede

punt is dat het aantal forensische adviezen, ook over de jaren heen, beperkt is. Zo zijn er eerder tienduizenden dan honderdduizenden indicatiestellingen forensische klinische zorg en pro Justitia-rapportages gerealiseerd. Een en ander maakt het justitiële veld niet een logische plek om dit soort AI-applicaties, die complexe beslissingen aan moeten kunnen, te ontwikkelen.

### *Complexe data*

AI-applicaties kunnen niet werken zonder dat er input wordt geleverd. Aan de voorkant moet een helder omschreven set van items worden ingevoerd. Zeker als het aantal brondocumenten waar de AI-software op moet 'leren' beperkt is. Op basis van deze input en informatie over hoe mogelijke verbanden moeten worden begrepen, kan AI in de ruwe data verbanden gaan leggen en leren. Dit vraagt, zeker aanvankelijk, om minder complexe input dan veelledige en gelaagde forensische rapportages met een grote variatie. Deze data kennen een hoge mate van ongestructureerdheid en zijn op meerdere manieren te interpreteren. Dit maakt het voor AI-software al lastiger om goed te kunnen werken. In het beschreven onderzoek bleek de gebruikte AI-software onvoldoende te werken op de verhalende stukken tekst waarop een aantal interpretatieslagen moesten worden gemaakt. Samenvattend bleken de data in dit proces nog te complex om snel tot ondersteuning met AI-technologie te kunnen komen.

### *Complex werk*

Niet alle kennis in de hoofden van forensische adviseurs als indicatiestellers en pro Justitia-rapporteurs is makkelijk op papier in regels om te zetten. Weliswaar is in dit onderzoek bevestigd dat AI vaak wel door hen genoteerde informatie kan herkennen, maar hun manieren om deze te interpreteren en te wegen zijn nog niet zo makkelijk te programmeren.

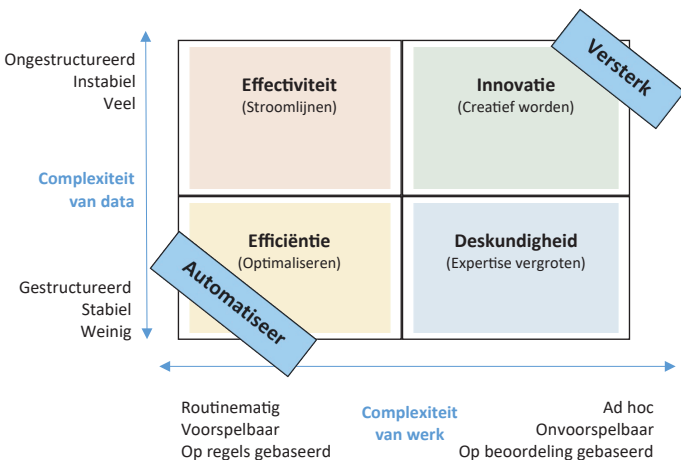
In zijn algemeenheid is AI-software wel goed in voorspellen, maar niet zo goed in beoordelen. Hoe complexer het werkproces en met name de aard van de beoordeling die moet worden gedaan, hoe lastiger dit is om aan AI-software over te laten. Daarnaast is in het besproken onderzoek gebleken dat het belangrijk is dat AI-softwareontwikkelaars een goed begrip moeten hebben van het daadwerkelijke beslisproces

in de praktijk. Misvattingen zoals beschreven staan de ontwikkeling van een functioneel product in de weg. Het is goed hier in toekomstige projecten rekening mee te houden en uitvoering en uitvoeringsbeleid op tijd te betrekken voor een gedegen inhoudelijk beeld van de te ondersteunen taak.

*Beslissen of en waar AI bruikbaar is in een beslisproces*

De hiervoor genoemde complexiteit van de data en het werkproces blijken belangrijke factoren om te kunnen inschatten of en waar AI het beste kan worden gebruikt om van meerwaarde te zijn in bedrijfsprocessen rondom besluitvorming (Bataller & Harris 2016). Hoe deze twee aspecten samen richting kunnen geven aan de verwachting van de haalbare meerwaarde van AI in een bedrijfsproces is door de auteurs weergegeven in een schema (zie figuur 1).

**Figuur 1** Schema data- en werkcomplexiteit in relatie tot inzet AI-software (Bataller & Harris 2016)



Grofweg kan worden gesteld dat werkprocessen die bestaan uit lage data- en werkcomplexiteit zich goed lenen voor automatisering van processen. Dit betekent dat de computer werk van de mens kan overnemen. Werkprocessen met een hoge data- en werkcomplexiteit zijn beter geschikt om mensen in het proces te versterken.

Uit het beschreven onderzoek blijkt dat inzet van AI rondom beslissingsprocessen aangaande indicatiestelling als vorm van forensische advisering nog ver weg ligt. Op dit gebied is AI nog erg in ontwikkeling en zoals eerder benoemd is het justitiële veld niet de beste proeftuin. Het lijkt daarom zinvoller te kijken waar in het justitiële veld AI kan worden ingezet bij (deel)processen met lage data- en werkcomplexiteit. Dit maakt ook dat het werkproces makkelijker te begrijpen is voor softwareontwikkelaars, waardoor er eerder applicaties worden opgeleverd die de werkelijkheid representeren en functioneel bruikbaar zijn. Uit het beschreven onderzoek bleek bijvoorbeeld dat als de indicatiesteller in een aparte handelingsstap aan de computer zou aangeven welke factoren hij/zij forensisch relevant (kritisch) acht, de computer technisch in staat zou kunnen zijn vergelijkbare casussen te vinden.

### *Eindconclusie*

AI-technologie om laagcomplexe data- en werkprocessen te automatiseren is al gevorderd. Het is zinvol te bezien of beslissingsprocessen binnen het justitiële veld die aan deze kenmerken voldoen, kunnen worden geautomatiseerd of versterkt middels AI. Met betrekking tot processen die een hoger niveau van data- en werkcomplexiteit kennen, lijkt het verstandiger te wachten tot ontwikkelingen op dit gebied gevorderd zijn. Bovendien moet worden gekeken of de hoeveelheid beschikbare brondata toereikend is om AI-software ook functioneel te kunnen inzetten.

### **Literatuur**

#### **Bataller & Harris 2016**

C. Bataller & J. Harris, *Turning artificial intelligence into business value*. Today, Accenture, 2016, <https://files.stample.co/browserUpload/fc3be0d1-906c-4db4-b572-649edf4c73ac>.

#### **Kraemer 2014**

H.C. Kraemer, 'The reliability of clinical diagnoses: state of the art', *Annual Review Clinical Psychology* (10) 2014, p. 111-130.



**Mulder e.a. 2014**

H.J. Mulder, D.M. de Bruin, Th. Rinne & E. Bulten, 'Forensische indicatiestelling: naar een professioneel consistente praktijk', *Sancties* (5) 2014, p. 300-311.

**Schadd e.a. 2022**

M.P.D. Schadd, A.W. Burger, S.E. Lensink & D.O. Domela Nieuwenhuis Nyegaard, *Cattell: een AI systeem voor de ondersteuning van het indicatiestellingsproces van klinische forensische zorg bij het NIFP* (TNO rapport R12012), 2022.

**Śmietanka e.a. 2021**

M. Śmietanka, A. Koshiyama & P. Treleaven, 'Algorithms in future insurance markets', *International Journal of Data Science and Big Data Analytics* (1) 2021, p. 1-19.

**Yu e.a. 2018**

K.H. Yu, A.L. Beam & I.S. Kohane, 'Artificial intelligence in health-care', *Nature Biomedical Engineering* (2) 2018, p. 719-731.