*Hans Henseler**

# *Redactioneel*

# De opkomst van kunstmatige intelligentie in expertise en recht: de 'krassporen' van *deeplearning*

Op 8 oktober jl. gaf prof. dr. H.J. van den Herik zijn college 'De Kracht van de Blinde Vlek' ter gelegenheid van zijn afscheid als hoogleraar Recht en Informatica bij eLaw, het Centrum voor Recht en Digitale Technologie van de Faculteit der Rechtsgeleerdheid van de Universiteit Leiden. In 1993 ben ik bij Van den Herik gepromoveerd op het onderwerp artificiële neurale netwerken nadat ik in 1987 onder zijn leiding aan de TU Delft was afgestudeerd op het thema kunstmatige intelligentie. De carrière van Van den Herik is zeer indrukwekkend en de geïnteresseerde lezer verwijs ik graag naar het interview[1] dat hij onlangs heeft gegeven aan de nieuwsbrief van de BNVKI (BeNeLux Vereniging voor Kunstmatige Intelligentie), waarvan Van den Herik in 1981 medeoprichter was en die in 2021 haar 40-jarig bestaan vierde.

Van den Herik en ik zien elkaar sinds enkele jaren weer met enige regelmaat bij de leergang Data Science die vanuit de Universiteit Leiden wordt georganiseerd. In die leergang geef ik een halve dag les. Daarin leg ik uit wat *deeplearning* is, hoe het werkt, hoe het ontstaan is en welke toepassingen er zijn. In de afgelopen tien jaar heeft *deeplearning* voor stormachtige ontwikkelingen gezorgd in het vakgebied kunstmatige intelligentie. Die storm is nog niet voorbij en de algemene verwachting is dat de komende jaren nog veel interessante doorbraken zullen volgen. De eerste ontwikkelingen werden zichtbaar voor het grotere publiek in 2012 toen een *deeplearning* methode op basis van een *Convolutional Neural Network* (CNN) de *ImageNet Large Scale Visual Recognition Challenge 2012* (ILSVRC2012)[2] won. Voor deze competitie is een subset van de volledige *ImageNet database* beschikbaar gemaakt met 1000 categorieën en 1,2 miljoen afbeeldingen. Een CNN genaamd AlexNet deed het aanzienlijk beter dan de concurrentie bij de interpretatie van afbeeldingen en het categoriseren van voorwerpen die op de foto's waren afgebeeld.

Geïnspireerd door dit succes hebben onderzoekers de jaren daarop talloze verbeteringen bedacht en is ook het toepassingsgebied uitgebreid naar andere modaliteiten zoals audio, video en taalanalyse. Zonder hierover een complete verhandeling te willen geven zijn daarvan een aantal wel de moeite waard om te noemen. In 2014 publiceerde Ian Goodfellow over het onderwerp van General Adversarial Networks (GANs), waarmee tegenwoordig niet alleen *deepfakes* gemaakt kunnen worden met negatieve toepassingen maar die ook veel positieve toepassingen kennen (waarover later meer in dit redactioneel). Een andere belangrijke mijlpaal in de AI-wereld is de overwinning van AlphaGo in 2016 op Lee Sedol, wereldkampioen go.[3] Deze overwinning kwam tien jaar eerder dan verwacht. AlphaGo was een combinatie van zoekstrategieën in combinatie met uitgebreide training[4] aan de hand van eerder gespeelde partijen. Misschien nog wel groter was de verrassing toen een jaar later AlphaGo werd verslagen door AlphaGo Zero, een neuraal netwerk dat zonder enige voorkennis na 40 dagen tegen zichzelf spelen met slechts kennis van de spelregels, in staat bleek alle eerdere programma's te verslaan.[5]

Neurale netwerken zijn niet alleen goed in het classificeren van afbeeldingen of het spelen van bordspellen. Integendeel zelfs. In 2016 heeft Google de statistische machinevertaling, waarop Google Translate sinds 2006 was gebaseerd, vervangen door een machinevertaling die is gebaseerd op neurale netwerken.[6] Neurale netwerken zijn nog beter gaan presteren bij de verwerking van taal nadat Google in 2018 een nieuw type neuraal netwerk BERT (*Bidirectional Encoder Representations from Transformers*) introduceerde. Initieel was BERT vooral gericht op de verwerking van taal, maar inmiddels zijn er *transformer* netwerken die multimodale data kunnen verwerken. Dat wil zeggen dat ze geleerd hebben om tekst te associëren met beeld en geluid en vice versa. Overigens is zeker niet alleen Google bepalend op dit gebied. Ook Microsoft, Facebook, IBM, Apple en Amazon timmeren hard aan de weg. Daarnaast zijn er ook nieuwe bedrijven zoals OpenAI, het bedrijf dat oorspronkelijk als non-profit startte met steun van Elon Musk. Sinds 2019 is OpenAI for-profit en heeft Microsoft er een miljard euro in geïnvesteerd en heeft het in 2020 de exclusieve rechten verworven op het GPT-3 taalmodel van

* Dr. ir. J. Henseler is lector *E-Discovery* en *Digital Forensics* bij Hogeschool Leiden, senior wetenschappelijk onderzoeker bij het Nederlands Forensisch Instituut en tevens redacteur van dit tijdschrift. De auteur dankt mr. Diederik Aben, dr. ir. Harm van Beek, prof. ing. Zeno Geradts, mr. Gert Haverkate, mr. ing. Nico Keijser, prof. dr. Marjan Sjerps en dr. ir. Rolf Ypma voor hun waardevolle commentaar en aanvullingen op eerdere versies van dit redactioneel.

1. Interview met prof. dr. H.J. van den Herik in de Nieuwsbrief van de BNVKI: ii.tudelft.nl/bnvki/?p=1790.
2. ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012), image-net.org/challenges/LSVRC/2012/index.php.
3. AlphaGo versus Lee Sedol (2016), en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol.
4. Met training wordt hier bedoeld het aanpassen van parameters in een wiskundig model (het neurale netwerk in dit geval) aan de hand van feedback die aangeeft of de voorspelling van het model goed of fout is.
5. Mastering the game of Go without human knowledge (2017), doi.org/10.1038%2Fnature24270.
6. Found in translation: More accurate, fluent sentences in Google Translate (2016), blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate.

OpenAI (de derde generatie van de *Generative Pre-trained Transformer* netwerken).

GPT-3 en vergelijkbare netwerken blijken in staat te zijn om veel algemene kennis op te nemen. Deze netwerken hebben zo veel informatie gezien dat ze in staat zijn tot *zero-shotlearning*. Dat wil zeggen dat het netwerk taken kan uitvoeren (zoals vragen beantwoorden of afbeeldingen classificeren) zonder dat het specifiek die taak heeft geleerd. Dat levert indrukwekkende voorbeelden op zoals de mogelijkheid van CLIP[7] om tekst met afbeeldingen te associëren. Daarmee is het nu mogelijk om aan de hand van een omschrijving van een categorie te zoeken naar afbeeldingen zonder dat daar specifiek op getraind hoeft te worden. In Dall-E worden GPT-3 en CLIP gecombineerd tot een systeem dat gegeven een beschrijving een afbeelding kan creëren die lijkt op de omschrijving.[8] Zo zijn er inmiddels talloze systemen waarmee gebruikers eenvoudig zelf nieuwe ontwerpen kunnen genereren, afbeeldingen kunnen genereren van personen, dieren of voorwerpen die niet bestaan of die helpen bij het schrijven van teksten, maken van samenvattingen of slimmer zoeken. De computer heeft verbeeldingskracht gekregen zoals Jarno Duursma het verwoordt in zijn zeer lezenswaardig rapport uit 2019.[9] Het is maar een kleine greep uit de toepassingen die mogelijk zijn en die we de komende jaren waarschijnlijk gewoon gaan vinden.

De opkomst van *deeplearning* staat uitgebreid beschreven in de Turing-lezing van Bengio, LeCun en Hinton die voor hun baanbrekende onderzoek en doorzettingsvermogen op dit gebied de *ACM Turing Award* hebben ontvangen.[10] Maar de ontwikkelingen staan niet stil en lijken zelfs steeds sneller te gaan. De *transformer* modellen die zo succesvol zijn bij de verwerking van natuurlijke taal hebben in 2021 inmiddels ook hun weg gevonden naar beeldherkenning.[11] De zogenaamde *vision transformers* zijn nu beter dan de CNN netwerkmodellen die sinds 2012 jaarlijks zijn verbeterd. Nieuwe technieken en toepassingen worden beschreven in wetenschappelijke publicaties waarbij de broncode en datasets ook online beschikbaar worden gemaakt. Andere wetenschappers kunnen zo binnen korte tijd in de praktijk beschikken over dezelfde *state of the art* technieken om ze toe te passen of te verbeteren. Deze technieken vinden ook toepassing in de medische sector. De huidige Covid-19-pandemie heeft het onderzoek naar deze nieuwe technieken verder versneld en heeft geleid tot efficiëntere ontwikkeling van geneesmiddelen. Niet alleen voor Covid-19 maar ook voor andere ziekten zoals kanker. Na de verbeteringen in zoeken en vertalen lijkt nu ook Google Maps aan de beurt. DeepMind en Google onderzoeken nu *graph neural networks* die goed lijken te zijn in het voorspellen van het tijdstip van aankomst voor een gegeven route. Dit is een lastig probleem waarbij het netwerk aan de hand van spatio-temporele patronen moet leren voorspellen wat de reistijden zullen zijn van de afzonderlijke segmenten in een route.

Met deze inleiding kom ik aan bij de titel van dit redactioneel. Wat betekent de opkomst van *kunstmatige intelligentie* voor expertise en recht? Tot nu toe is er vooral veel aandacht voor kunstmatige intelligentie in het kader van *legal technology*, als hulpmiddel voor de wetgever, het Openbaar Ministerie en de advocatuur om efficiënter hun werk te kunnen doen of voor burgers om het recht toegankelijker te maken. Denk bijvoorbeeld aan *chatbots* die vragen in natuurlijke taal begrijpen en beantwoorden. De technieken op dit gebied liggen in het verlengde van het gebied dat Van den Herik al in 1991 opschudde met zijn voorspelling dat computers ooit zouden kunnen rechtspreken. *Legal technology* is een bloeiend vakgebied waarin veel ruimte is voor kunstmatige intelligentie bij *fact-finding*, geautomatiseerde juridische beslissingsprocessen, *e-discovery*, automatische classificatie van sporen, contractanalyse en het beheer van documenten. In het forensische domein wordt kunstmatige intelligentie ook ingezet voor bewijsevaluatie, bijvoorbeeld sprekervergelijking. Maar wat nu als kunstmatige intelligentie en in het bijzonder *deeplearning* zelf onderwerp van onderzoek wordt? Welke forensische expertise is dan nodig? Of anders gezegd, als *deeplearning* een gereedschap is waarvan gebruikers (bewust of onbewust) gebruikmaken, wat zijn dan de 'krassporen' die *deeplearning* veroorzaakt en hoe kunnen we die sporen herkennen en interpreteren?

*Deepfakes* zijn afbeeldingen, video's en geluiden die zijn gecreëerd met *deeplearning* technieken. Met dit soort bestanden kan opzettelijk onjuiste informatie verspreid worden. Om te kunnen bepalen of een afbeelding een *deepfake* is, moet gezocht worden naar sporen van *deeplearning*. In een strafrechtelijk onderzoek naar *deepfakes* is daarom forensisch onderzoek nodig naar dat soort sporen. Meta AI (voorheen Facebook) heeft in 2020 een *deepfake detection challenge dataset* gepubliceerd[12] en in de wetenschappelijke literatuur zijn verschillende benaderingen te vinden om *deepfakes* te ontmaskeren. Dat blijkt in de praktijk lastig en bovendien wordt de software die *deepfakes* genereert steeds beter. Bovendien is voor forensisch onderzoek alleen detectie onvoldoende en is er ook behoefte aan een verklaring van de sporen. Het NFI doet samen met de Universiteit van Amsterdam in het Innovatie Centrum voor AI (ICAI) onderzoek naar het herkennen van *deepfakes*.[13] Het bedrijf DuckDuckGoose heeft *DeepDetector* ontwikkeld[14] en detecteert *deepfakes* en visualiseert waarom een afbeelding verdacht is. Deze technologie is inmiddels geïntegreerd in een product voor forensische beeldanalyse en wordt nu ook gebruikt voor niet-forensische toepassingen. Zo gebruiken journalisten op dit moment

7. CLIP: Connecting Text and Images (2021), openai.com/blog/clip.
8. DALL·E: Creating Images from Text (2021), openai.com/blog/dall-e.
9. Machines met verbeeldingskracht. Een kunstmatige realiteit (2019), jarnoduursma.nl/boek/machines-met-verbeeldingskracht-een-kunstmatige-realiteit.
10. Yoshua Bengio, Yann LeCun & Geoffrey Hinton, 'Deep Learning for AI', *Communications of the ACM*, July 2021, Vol. 64, No. 7, pp. 58-65, 10.1145/3448250.
11. Voor een overzicht van de nieuwste ontwikkeling zie bijvoorbeeld het *State of AI Report 2021*, stateof.ai.
12. Deepfake Detection Challenge Dataset (2021), ai.facebook.com/datasets/dfdc.
13. UvA en NFI doen onderzoek naar herkennen deepfakes en verborgen berichten van criminelen (2021), forensischinstituut.nl/actueel/nieuws/2021/05/22/uva-en-nfi-doen-onderzoek-naar-herkennen-deepfakes-en-verborgen-berichten-van-criminelen.
14. DuckDuckGoose, How does DeepDetector work?, duckduckgoose.ai/detector.

een prototype om afbeeldingen te controleren bij het browsen op internet. Bedrijven die klanten identificeren met een live video vanaf een smartphone hebben interesse getoond nu het mogelijk (en eenvoudig) blijkt om met *real-timedeepfake* technieken videobeelden te manipuleren. Dat dit niet vergezocht is bleek in april vorig jaar toen Nederlandse Kamerleden een gesprek dachten te voeren met een hooggeplaatste medewerker van de Russische oppositieleider Navalny. In werkelijkheid spraken zij met 'iemand' die zich als zodanig voordeed.[15]

*Deepfakes* zijn een probleem dat tot de verbeelding spreekt als het gaat om bewijskracht. Maar zoals blijkt uit de voorbeelden in de inleiding van dit redactioneel reikt de impact van *deeplearning* veel verder. *Deeplearning* is niet per se intelligent maar is wel in staat om veel meer soorten data te analyseren. Dankzij *deeplearning* kunnen systemen feilloos leren om ontbrekende gegevens in te vullen. Dat kan het zelfs al zo goed dat we tegenwoordige televisies in de huiskamer hebben die haarscherpe beelden in 8K-resolutie aan ons tonen terwijl de beelden maar in HD-kwaliteit (1K) worden uitgezonden. De 'verbeeldingskracht' van de computer in onze televisie is zo goed getraind dat die in staat is om natuurgetrouw de resolutie te verbeteren zonder daarbij de waarheid (merkbaar) geweld aan te doen. Het is een illustratie van het vermogen van deze nieuwe generatie systemen om patronen te leren op een schaal en met een snelheid die het menselijke vermogen ver te boven gaat. Dat betekent geenszins dat de computer intelligent is.

De meerwaarde van *deeplearning* zit vooral in taken waar netwerken bijna zo goed of ongeveer even goed in zijn als mensen, maar die door de netwerken wel veel sneller en goedkoper uitgevoerd kunnen worden. Nog meer waarde ontstaat er als mens en machine samenwerken en de kunstmatige intelligentie de intelligentie van mensen versterkt. Dit wordt ook wel aangeduid als *augmented intelligence*. Waarschijnlijk zullen mensen steeds meer (bewust of onbewust) geholpen worden door kunstmatige intelligentie. Die toepassingen gaan verder dan de spellen, zoeksystemen en sociale media op internet. Nu al zijn er aantoonbaar minder ongelukken met auto's die over een intelligente noodremfunctie beschikken (minder kop-staartbotsingen). Dankzij *deeplearning* is een revolutie ontketend in de manier waarop we medicijnen ontwikkelen. Zelflerende software helpt bij de analyse van CT-scans zodat radiologen hun werk sneller, preciezer en effectiever doen. Dit zijn maar een paar aansprekende voorbeelden en het aantal terreinen waar *deeplearning* wordt ingezet groeit snel. Denk bijvoorbeeld aan industriële toepassingen zoals ontwerpen en simuleren, weersvoorspellingen, planning en logistiek, preventief onderhoud en robotica. Het belang van deze ontwikkeling voor expertise en recht wil ik graag illustreren aan de hand van een voorbeeld.

Bij een onderzoek op de plaats delict kan een heel scala aan sporen worden bemonsterd. Een voorbeeld daarvan zijn fysieke krassporen in een kozijn bij een inbraak aan de hand waarvan mogelijk een dader kan worden geïdentificeerd. Een *deeplearning* systeem laat een ander soort sporen na zoals kleine details op grond waarvan een *deepfake* gedetecteerd kan worden. Maar hoe moeten we ons dit voorstellen in die andere toepassingen waarin *deeplearning* wordt gebruikt? Bijvoorbeeld een zelfrijdende auto krijgt een ongeluk, een verkeerde diagnose die is gesteld door een medisch apparaat, een brand die is ontstaan door een niet functionerende slimme stroomverdeler. Laat ik eerst een voorbeeld geven van een incident waarbij geen *deeplearning* systeem betrokken is. Bijvoorbeeld een ongeluk waarbij het dak boven de tribune in een voetbalstadion is ingestort. Zo'n constructie is vooraf uitgebreid doorgerekend en er is een bouwvergunning verstrekt. Je hebt een deskundige constructeur nodig die ter plaatse de situatie kan opnemen, kan vergelijken met het ontwerp en de bijbehorende berekeningen en die een antwoord kan geven op onderzoeksvragen. Is er een fout in de berekening gemaakt? Waren er extreme weersomstandigheden? Of heeft de aannemer materialen gebruikt die niet conform de specificaties waren? Nu terug naar een voorbeeld met *deeplearning*. Een bestuurder van een zelfrijdende auto heeft een dodelijk ongeluk veroorzaakt. Wie gaat dan het onderzoek doen? De auto heeft allerlei sensoren die het zelfrijdende gedrag bepalen door middel van een *deeplearning* systeem. De fabrikant heeft ongetwijfeld uitgebreid getest maar is niet onafhankelijk. Een onafhankelijk gerechtelijk deskundige is nodig die verstand heeft van *deeplearning* en die in staat is om aan de hand van beschikbare gegevens een uitspraak te doen over de werking van het systeem en de omstandigheden ten tijde van het incident.

De vraag die we onszelf moeten stellen is: Welke 'krassporen' laat *deeplearning* achter en hoe moeten die sporen geëvalueerd worden? *Wie* zijn de gerechtelijk deskundigen die verstand hebben van deze sporen en die in staat zijn om daar forensisch onderzoek naar te doen en erover te rapporteren? Omdat *deeplearning* veel met computers te maken heeft, ligt het misschien voor de hand om te veronderstellen dat deskundigen in het deskundigheidsgebied Digitaal Forensisch Onderzoek (DFO)[16] er verstand van hebben. DFO kent zes verschillende deelgebieden maar geen daarvan voorziet in de juiste deskundigheid. Ook de deskundige op het deelgebied 008.2 Digital Forensics – Software forensics is ongeschikt want *deeplearning* is dan wel een algoritme maar het gedrag wordt bepaald door de parameters in het model dat is geleerd. Een rechter kan ook een *deeplearning* specialist vanuit het bedrijfsleven of de wetenschap als *ad hoc* deskundige benoemen maar het probleem blijft dat zij vooral gespecialiseerd zijn in toepassingen van *deeplearning*. Dat geldt ook voor deskundigen bij het NFI met de specialisatie Forensische Big Data

15. Arnout Brouwers & Laurens Verhagen, 'Kamerleden vergaderen met deepfake-imitatie van stafchef Russische oppositieleider Navalny', *de Volkskrant* 23 april 2021, volkskrant.nl/nieuws-achtergrond/kamerleden-vergaderen-met-deepfake-imitatie-van-stafchef-russische-oppositieleider-navalny~b04b5322.
16. J. Henseler & S.M. van Loenhout, 'Registratie gerechtelijk deskundigen Digitaal Forensisch Onderzoek', *EeR* 2016, afl. 5, p. 207.

Analyse (FBDA)[17] die op dat gebied samenwerken met de UvA.[18]

Het nieuwe deskundigheidsgebied Forensische Statistiek en Big Data Analyse (FSBDA)[19] biedt mogelijk meer perspectief. Dit deskundigheidsgebied is begin 2021 ontstaan uit de samenvoeging van de specialisatierichtingen FBDA en Forensische Statistiek & Methodologie. De deskundigen in dit team zijn gespecialiseerd in het redeneren met en kwantificeren van waarschijnlijkheden waarbij ze gebruikmaken van statistiek, kansrekening, kunstmatige intelligentie en big data-analyse en passen dit toe op forensisch onderzoek. Enkele voorbeelden van onderzoeksvragen waarmee FSBDA recentelijk aan de slag is gegaan: wat is de kans dat deze brand is aangestoken? Heeft de verdachte deze telefoon gebruikt? Wie is de auteur van deze e-mail? Centraal hierbij staat het redeneren met en het kwantificeren van waarschijnlijkheden waarbij de deskundigen zelf ook gebruikmaken van *deeplearning* technieken. Dat soort vaardigheden zijn nuttig bij het vaststellen van kenmerken van en tekortkomingen (beperkingen of onbedoelde voorkeuren) in *deeplearning* systemen. Omdat FSBDA niet gebonden is aan een forensisch domein werkt het samen met domeinexperts die vakinhoudelijke kennis inbrengen. Diezelfde samenwerking met domeinexperts komt goed van pas wanneer 'krassporen' van *deeplearning* onderzocht moeten worden.

17. Forensische big data analyse (FBDA) richt zich op technieken voor intelligente data-analyse om essentiële informatie te halen uit grote hoeveelheden digitale gegevens, forensischinstituut.nl/forensisch-onderzoek/forensische-big-data-analyse.
18. Zie noot 13.
19. NFI breidt uit met nieuw forensisch vakgebied FSBDA (2021), forensischinstituut.nl/actueel/nieuws/2021/02/03/nfi-breidt-uit-met-nieuw-forensisch-vakgebied-fsbda.