

Commentaar op Eric Tjong Tjin Tai, Recht tussen mens en techniek (preadvies). In A. Berlee e.a., *De toekomst van de jurist, de jurist van de toekomst*¹, Deventer: Wolters Kluwer 2020, p. 177-198.

Henry Prakken

(Uitgesproken tijdens het NJV-jubileumcongres “De toekomst van de jurist, de jurist van de toekomst”, Den Haag, 8 oktober 2021.)

Eric Tjong Tjin Tai en Anna Berlee lijken het er in hun preadviezen over eens dat de computer de rechter/jurist niet moet vervangen maar moet ondersteunen. Dat betekent dat de jurist de recente ontwikkelingen in de Artificiële Intelligentie (AI) niet mag negeren maar zich moet voorbereiden op een andere manier van werken, ondersteund door AI. Routinewerk wordt overgenomen door de computer, en de menselijke jurist blijft er voor het creatieve werk en de inbreng van menselijke aspecten, zoals gevoel voor het bijzondere van individuele gevallen.

Ik ben het hier in grote lijnen mee eens. Wat kan ik verder over Tjong Tjin Tai’s preadvies zeggen? Hij bespreekt veel onderwerpen, maar ik heb maar 10 minuten. Ik zal mij daarom beperken tot enkele kanttekeningen die hopelijk relevant zijn voor het debat van straks.

1. AI is niet gelijk aan (data-gedreven) machine learning.

Mijn eerste kanttekening is dat AI niet gelijk staat aan machine learning. Het valt mij op dat Tjong Tjin Tai het in zijn preadvies alleen heeft over machine-learningtoepassingen. Eerst moet ik heel kort uitleggen wat machine-learningalgoritmen zijn. Dergelijke algoritmen passen niet, zoals kennisgebaseerde systemen, door de menselijke programmeur ingevoerde kennis toe om een probleem op te lossen, maar leren de benodigde kennis zelf uit grote hoeveelheden data, vaak in de vorm van statistische verbanden.

Nu is het inderdaad zo dat machine learning de laatste jaren hot is, en dat alle recente doorbraken in de AI (zoals *AlphaGo*, dat een van de beste menselijke go-spelers versloeg, automatisch vertalen, zoals *Google Translate*, zelfrijdende auto’s) grotendeels of geheel te danken zijn aan vooruitgang in de machine learning, in combinatie met de steeds toenemende rekenkracht van computer (vooral als verbonden in netwerken) en de steeds grotere hoeveelheid beschikbare data, vooral op internet.

Maar er is ook kennisgebaseerde AI. Zo noemt Berlee in haar preadvies juridische kennissystemen waarbij rechtsregels omgezet worden in voor de computer hanteerbare ‘logische’ regels, waarna de computer al logisch redenerend die kan toepassen op (ook in logische vorm omgezette) feiten. Die worden behoorlijk succesvol toegepast, vooral bij de bestuursrechtelijke uitvoering van wetgeving in bijvoorbeeld het sociale-zekerheidsrecht en het belastingrecht.

¹ Te downloaden van <https://njv.nl/wp-content/uploads/2020/08/De-toekomst-van-de-jurist-de-jurist-van-de-toekomst.pdf> (geraadpleegd 11 oktober 2021).

Marlies van Eck heeft in haar proefschrift wel tekortkomingen ontdekt, o.a. wat betreft transparantie van systemen en uitlegbaarheid van hun beslissingen, maar ik waag het te beweren dat dit niet een inherent probleem met dit soort systemen is. Integendeel, ze kunnen heel transparant zijn, maar de problemen komen volgens mij doordat de systeemontwikkelaars de lessen van de wetenschap over hoe je dit soort systemen netjes en transparant kunt ontwerpen grotendeels negeren.

Verder zijn er (kennisgebaseerde) argumentatiesystemen, ook kort besproken door Berlee. Dat zijn, heel kort gezegd, systemen die op basis van kennis over juridisch relevante factoren argumenten voor en tegen een beslissing kunnen geven. Lang waren dit soort systemen een kwestie van de tekentafel en fundamenteel onderzoek, maar sinds kort zijn er succesvolle toepassingen van kennisgebaseerde argumentatiesystemen in de advocatuur. Katie Atkinson en Trevor Bench-Capon, twee bekende Britse AI&recht-onderzoekers van de Universiteit van Liverpool, hebben toepassingen gebouwd voor een groot advocatenkantoor in die stad, en dat kantoor is er naar verluidt heel tevreden over.

Bovendien zijn sommige van de meest accurate algoritmische voorspellers van uitkomsten van rechtszaken (een toepassing die de laatste jaren veel aandacht heeft gekregen) niet gebaseerd op machine learning maar passen ze zo'n kennisgebaseerd juridisch argumentatiemodel toe. Het gaat hier vooral om het werk van Kevin Ashley en zijn promovendi. Dit soort systemen kan dus niet alleen een uitkomst van een rechtszaak voorspellen, maar ook juridisch relevante redenen voor en tegen de voorspelde uitkomst geven. En ze behoren tot de beste juridische uitkomstvoorspellers van dit moment.

2. Juridische noodrem? Niet bij uitkomstvoorspellers.

Tjong Tjin Tai suggereert dat in de toekomst bepaalde beslissingen in eerste instantie aan de computer overgelaten kunnen worden, en dat menselijke juristen aan een juridische noodrem kunnen trekken als dat nodig is en de computer kunnen overrulen. Een toepassing waarbij anderen dat gesuggereerd hebben is die van de zojuist genoemde algoritmische uitkomstvoorspellers.

De hoop van velen lijkt te zijn dat als rechters zich in rechtszaken door algoritmische uitkomstvoorspellers laten informeren, dat dit de voorspelbaarheid en consistentie van hun beslissingen ten goede komt. De achterliggende gedachte lijkt te zijn dat voorspelalgoritmes de rechter vertellen wat de normale beslissing in de zaak zou zijn, een beslissing die overgenomen zou moeten worden tenzij de rechter bijzondere omstandigheden in de zaak kan vinden om van de voorspelling af te wijken (een voorbeeld van een juridische noodrem bij gebruik van AI).

Maar een probleem is dat dit helemaal niet de informatie is die voorspelalgoritmes geven, dus dit gebruik van voorspelalgoritmes is geen goed idee. En dit niet om juridische maar om technische redenen. Ik heb geen tijd om dat nu goed uit te leggen (zie hiervoor mijn recente artikel met Floris Bex²), maar de opvatting dat voorspelalgoritmes de normale uitkomst van een zaak geven

² F.J. Bex & H. Prakken, On the relevance of algorithmic decision predictors for judicial decision making. *Proceedings of the 18th International Conference on Artificial Intelligence and Law*, Sao Paulo (Brazil) 2021. New York: ACM Press 2021, 175-179.

berust op een verkeerd begrip van wat deze algoritmes doen, met name op een verkeerd begrip van de onderliggende statistiek en kansrekening.

Dat geldt trouwens niet alleen voor machine-learninggebaseerde maar ook voor kennisgebaseerde uitkomst voorspellers. Die laatsten hebben zoals gezegd wel het voordeel dat ze hun voorspellingen in juridische zin kunnen uitleggen, omdat ze niet naar statische maar naar redengevende verbanden kijken. De vraag rijst dan of machine-learninggebaseerde uitkomstvoorspellers dat misschien niet ook kunnen doen.

3. Statistische verbanden zijn niet hetzelfde als redengevende verbanden.

Tjong Tjin Tai verwacht dat machine-learninggebaseerde systemen in de toekomst conceptvonnissen kunnen produceren op basis van statistische verbanden tussen feiten en een uitkomst in eerdere zaken. Hij vindt dat zulke conceptvonnissen geen juridische autoriteit hebben. Daar ben ik het mee eens. Stel dat een algoritme statistische verbanden vindt tussen de wijk waarin een dader van een bepaald strafbaar feit is opgegroeid en krijgen van of een gevangenisstraf of een alternatieve straf. Dan willen we niet dat de rechter zegt: Jan-Karel krijgt een alternatieve straf want hij woont in het Statenkwartier, maar Nelis krijgt een gevangenisstraf want hij woont in de Schilderswijk. De jurist wil juist juridisch redengevende verbanden tussen feiten en straf.

4. De nieuwste AI-trend: taalmodellen.

Maar toch is er recent onderzoek in de AI waarbij de grens tussen puur statistische en betekenisvolle verbanden lijkt te vervagen. Het gaat hier om zogenaamde taalmodellen, waarvan BERT van *Google* de bekendste is. Dat zijn enorm grote statistische modellen, geleerd uit grote hoeveelheden teksten, en die modellen gaan in essentie over niets anders dan over hoe vaak verschillende woorden bij elkaar in de buurt staan in soortgelijke contexten. Daarmee kunnen ze bijvoorbeeld voorspellen hoe een zin aangevuld moet worden. Denk bijvoorbeeld aan hoe *Whatsapp* de woorden in uw berichtjes probeert aan te vullen. Taalmodellen proberen zo de betekenis van woorden en uitdrukkingen te reduceren tot statistische verbanden over nabijheid in teksten.

Dat lijkt simplistisch maar het fascinerende is dat er verrassend veel impliciete kennis in dergelijke modellen zit. Zo zou een op zo'n taalmodel gebaseerd systeem bijvoorbeeld³ de zin

Erasmus was geboren in ...

kunnen aanvullen met

... Rotterdam⁴

³ Alle drie de volgende voorbeelden zijn door mij bedacht, hoewel de eerste twee voorbeelden een goed beeld geven van wat nu al technisch mogelijk is.

⁴ (Toegevoegd na mijn voordracht) Tijdens het uitspreken van mijn voordracht veronderstelde ik dat Erasmus inderdaad in Rotterdam is geboren. Maar een toehoorder zei dat Erasmus niet in Rotterdam maar in Gouda geboren

puur en alleen omdat deze woorden in teksten vaak bij elkaar in de buurt staan. Ook zou zo'n systeem de zin

Jan is ongetrouwd dus Jan is ...

kunnen aanvullen met

... een vrijgezel

alweer omdat de woorden 'ongetrouwd' en 'vrijgezel' vaak bij elkaar in de buurt staan. Het lijkt er zo op dat een dergelijk systeem in staat om eenvoudige gevolgtrekkingen te maken op grond van de betekenis van woorden.

Dan is het niet zo'n grote stap naar pogingen om een feitelijk verhaaltje juridisch proberen te kwalificeren door te kijken naar hoe vaak bepaalde juridische termen in de buurt staan van feitelijke termen. Bijvoorbeeld:

Marie zette een foto van Jan op Facebook, dus ... Marie heeft persoonsgegevens van Jan verwerkt

of als de context niet op privacyrecht maar op auteursrecht duidt:

... Marie heeft een werk geopenbaard.

Er zijn nu nog geen op taalmodellen gebaseerde systemen die dit goed kunnen, maar de eerste experimentele artikelen in deze trant zijn al verschenen. Ik ben er zeker van dat u hier in de komende vijf of tien jaar in de populaire juridische pers meer van gaat zien en horen, want dit is echt een van de nieuwste trends in de AI.

Maar is dit nu echt juridisch denken en redeneren? Hier geldt wat Tjong Tjin Tai volgens mij terecht opmerkt over veel AI-toepassingen: het algoritme heeft geen echt begrip van wat het doet, het heeft er geen gevoel bij, en het kan niet geloven in de juistheid van zijn antwoorden.

Maar goed, dit is een van de nieuwste trends in de AI, en ik denk dat wij als juristen, rechtswetenschappers en AI-onderzoekers deze nieuwe ontwikkeling heel kritisch moeten volgen. Ze is fascinerend maar ook vol gevaren als dit echt toegepast wordt in het recht. En Eric Tjong Tjin Tai 's preadvies bevat wat dat betreft een aantal zeer waardevolle overwegingen.

was. Hierop zei een andere toehoorder dat de Erasmus' werkelijke geboorteplaats betwist is. Dit voor mij op zich enigszins pijnlijke moment was een goede gelegenheid voor mij om te wijzen op een serieus probleem van taalmodellen, namelijk dat de correctheid van specifieke antwoorden moeilijk te valideren is.