FORUM

DE EUROPESE AI-VERORDENING: EEN NIEUWE BOUWSTEEN VOOR DE DIGITALE RECHTSSTAAT

ANNE MEUWESE & JOHAN WOLSWINKEL - 21.04.2021

Vandaag presenteerde de Europese Commissie haar lang verwachte voorstel voor een AI-Verordening (waarvan een oudere versie overigens vorige week al was uitgelekt). Hiermee heeft Commissievoorzitter Ursula van der Leyen met enige vertraging haar belofte uit 2019 ingelost dat ze binnen 100 dagen wetgeving zou voorstellen voor een gecoördineerde Europese aanpak van de menselijke en ethische implicaties van AI. Nadat de Commissie begin 2020 in haar Witboek over kunstmatige intelligentie 'Een Europese benadering op basis van excellentie en vertrouwen' al de contouren van die Europese wetgeving had geschetst, ligt er nu dan een concreet voorstel in de vorm van een Europese verordening.

In de afgelopen vijf jaar lijkt *artificial intelligence* (AI) of kunstmatige intelligentie het nieuwe toverwoord te zijn geworden ter vervanging van big data. Tegelijk bevestigt de voorgestelde AI-Verordening dat AI-systemen gevoed zullen moeten worden door grote hoeveelheden data. Achter de noemer van *artificial intelligence* (AI) of kunstmatige intelligentie gaat een veelheid aan moderne computertechnologieën schuil (zoals *machine learning*, *deep learning*, *knowledge-based systems*) die met elkaar gemeen hebben dat het computersysteem steeds in meerdere of mindere mate zelfstandig kan opereren. Het potentieel van AI lijkt ongekend, maar tegelijk zijn er zorgen over de risico's die het gebruik van deze technologie meebrengt, zoals *bias* in het algoritme en het ontbreken van enige menselijke interventie.

Het reguleren van AI is om een aantal redenen bijzonder lastig. Ten eerste is niet altijd evident of een bepaalde technologie berust op traditionele, volledig voorgeprogrammeerde *rule-based* algoritmes zonder zelflerend element ("als *x*, dan *y*") dan wel op zelflerende *case-based* algoritmes (AI). Bestaande Nederlandse reguleringsinitiatieven, zoals de richtlijnen voor het toepassen van algoritmes door overheden van het ministerie van Justitie en Veiligheid en het recent gelanceerde Toetsingskader Algoritmes van de Algemene Rekenkamer, richten zich (daarom) op het meer generieke object 'algoritme'. Ten tweede zijn naleving en handhaving niet vanzelfsprekend, omdat degenen die bestuurlijk of juridisch verantwoordelijk zijn binnen een organisatie vaak geen zicht hebben op wat de technologieën die binnen de organisatie in allerlei processen in gebruik zijn, nu precies 'zijn' en 'doen'. Ten derde zijn velen het erover eens dat een balans tussen het inperken en

het stimuleren van innovatie noodzakelijk is, maar is er wetenschappelijk nog te weinig bekend over de (langetermijn)effecten van de inzet van AI om de politieke besluitvorming goed te kunnen voeden. Zowel 'overregulering' als een te lichte vorm van ingrijpen liggen op de loer.

Een eerste impressie van dit voorstel laat zien dat de Europese Commissie aansluit bij een aantal typische kenmerken van eigentijdse Europese wetgeving om zo tussen deze drie reguleringsuitdagingen door te laveren. Zo werpt de Commissie zich op als pionier door een '*dedicated AI law*' te presenteren die voor alle sectoren geldt met expliciete verantwoordelijkheden voor zowel ontwikkelaars als gebruikers van AI-systemen. Die keuze kan geïnterpreteerd worden als een manier om een breder besef van verantwoordelijkheid voor de naleving van het nieuwe kader binnen organisaties aan te wakkeren. Een andere keuze is die om niet alle AI-systemen in gelijke mate aan te pakken; een keuze die kan worden beschouwd als een poging de proportionaliteit van het voorstel veilig te stellen. Enkele AI-toepassingen, zoals het beruchte *social scoring* van mensen maar ook, op een paar uitzonderingen na, het gebruik voor handhavingsdoeleinden van 'real-time' *remote* biometrische identificatiesystemen in de openbare ruimte, worden expliciet verboden omdat deze een 'onacceptabel risico' vormen in het licht van de Europese waarden. Verder richt de AI-verordening specifiek haar pijlen op '*high risk'* AI-systemen, een term die al eerder werd geïntroduceerd in het witboek, maar nog niet echt is ingeburgerd. Dit hoge risico ziet op gezondheids- en veiligheidsrisico's, maar ook op risico's op het vlak van fundamentele rechten. De bijbehorende risicobeoordeling gaat daarbij niet alleen uit van de functie van een AI-systeem als zodanig, maar ook van het specifieke doel waarvoor het ingezet wordt. Voor deze categorie van door de verordening aangewezen *high-risk* AI-systemen (en eventueel later door de Commissie aan te wijzen AI-systemen) formuleert de verordening een aantal algemene eisen ten aanzien van onder meer de kwaliteit van de gehanteerde datasets, de uitlegbaarheid, accuraatheid en robuustheid van het systeem en de mogelijkheid van menselijke tussenkomst. Het is vervolgens meestal aan de ontwikkelaar van het AI-systeem zelf om middels een conformiteitsbeoordeling te laten zien dat het AI-systeem aan deze eisen voldoet.

De keerzijde van deze reguleringsaanpak kan zijn dat de resulterende open normen ruim baan scheppen voor 'BigTech'-bedrijven, die AI al grootschalig inzetten om hun macht te vergroten om een geavanceerd juridisch woordenspel te beginnen over termen als 'voldoende transparant' en 'hoge mate van accuraatheid'. Toch is het alternatief, een veel verder dichtgeregeld voorstel, eigenlijk geen optie, aangezien de technologieën zo sterk in ontwikkeling zijn en in zekere zin ook zo uiteenlopen dat de concrete betekenis van de normen uiteindelijk voor elk AI-systeem afzonderlijk moet worden bepaald. Interessant in het licht van die differentiatie is verder dat voor bepaalde AI-systemen die specifieke manipulatierisico's met zich meebrengen (denk aan '*deep fakes*') speciale transparantievereisten gaan gelden, zoals de verplichting te vermelden dat het beeld

of de audio met behulp van AI gegenereerd is.

Binnen de *high-risk* AI-systemen vallen twee domeinen op die direct het optreden van de overheid raken: sociale zekerheid en asiel. Reden voor de Commissie om AI-toepassingen binnen deze domeinen als *high risk* aan te merken en daarmee te onderwerpen aan de hierboven besproken eisen inzake kwaliteit van datasets, uitlegbaarheid, etcetera, is dat het recht op menselijke waardigheid in het gedrang kan komen wanneer overheden dergelijke technologie inzetten bij mensen die buitengewoon afhankelijk zijn van de overheid. In het licht van de vele discussies over algoritmische besluitvorming binnen de Nederlandse overheid, is de lakmoesproef voor dit Commissievoorstel of de burger nu beter beschermd zal zijn tegen SyRI-achtige toepassingen bij de bestrijding van uitkeringsfraude, of tegen kinderopvangtoeslagpraktijken. Waar het Systeem Risico Indicatie (SyRI) 1.0 sowieso al sneuvelde bij de rechter, omdat de overheid geen inzage wilde geven in het gehanteerde systeem (en dus in het midden bleef of hier sprake was van AI en/of big data), dwingt de AI-Verordening overheden om hun omgang met (bepaalde) algoritmes serieus te nemen, hetzij als ontwikkelaar hetzij als gebruiker hiervan. Het is de vraag of het Nederlandse wetsvoorstel dat 'SyRI 2.0' mogelijk moet maken voldoende waarborgen bevat rond geautomatiseerde gegevensanalyse. De tekst zoals die nu voorligt bij de Eerste Kamer zet weliswaar sterk in op uitlegbaarheid, navolgbaarheid en controleerbaarheid van algoritmes, maar is nog niet afgestemd op de risicobenadering van het Europese voorstel. Vernieuwend aan de AI-Verordening is bovendien dat deze het belang van menselijke tussenkomst (en daarmee eigenlijk ook van de menselijke maat) expliciet benadrukt, een belang dat snel naar de achtergrond dreigt te raken bij algoritmische besluitvorming. In die zin moet de pioniersrol van deze AI-Verordening voor de ontwikkeling van de digitale rechtsstaat zeker niet worden onderschat.