Innoveren door te reguleren: een bespreking van het rapport inzake de regulering van AI

Mr. S.W. van de Ven *

Recentelijk heeft het Europees Parlement een voorstel voor een verordening gepubliceerd ten aanzien van de regulering van artificial intelligence (AI). Dit artikel bespreekt de bepalingen en implicaties van het voorstel en geeft daarmee een inkijk in Europese AI-regelgeving die in het verschiet ligt.

1 Inleiding

De regulering van artificiële intelligentie, robotica en verwante technologieën (hierna gezamenlijk: AI) staat al enige tijd hoog op de agenda van de EU.¹ Zo stelde Ursula von der Leyen bij haar aantreden als voorzitter van de Europese Commissie in 2019 dat ze binnen honderd dagen zou komen met ‘wetgeving voor een gecoördineerde Europese aanpak ten aanzien van de menselijke en ethische implicaties van AI’.² Kort hierna publiceerde de Europese Commissie een witboek over AI dat de door de Europese Commissie geformuleerde beleidsopaties voor de regulering van AI bevat (hierna: het Witboek).³

De noodzaak tot regulering van AI vloeit voort uit de veronderstelling dat AI optimaal kan worden benut als de burger vertrouwen heeft in AI.⁴ Europese regelgeving op basis van fundamentele waarden moet dat vertrouwen geven en bedrijven aanmoedigen AI-toepassingen te ontwikkelen. De regulering van AI past tevens bij andere wetgevingsinitiatieven van de EU, zoals de bredere Digitale Strategie, waarin onder

meer een interne markt voor data binnen de EU is beoogd.⁵ Ook houdt AI-regelgeving verband met de Europese Green Deal, aangezien de verwachting is dat de ontwikkeling van AI, mits goed gereguleerd, kan bijdragen tot het verwezenlijken van de duurzaamheidsdoelstellingen van de Europese Green Deal.⁶

Gezien deze achtergrond en het feit dat de Europese Commissie sinds het in februari 2020 uitgebrachte Witboek nog niet met een conceptvoorstel is gekomen, heeft het Europees Parlement in oktober 2020 een rapport uitgebracht omtrent de regulering van AI (hierna: het Rapport).⁷ Met het Rapport verzoekt het Europees Parlement de Europese Commissie om een voorstel voor een verordening in te dienen op basis van de tekst die is opgenomen in de bijlage van het Rapport (hierna: het Voorstel).⁸ Aangezien het Voorstel sterk overeenkomt met de eerder door de Europese Commissie geformuleerde hoofdlijnen voor de regulering van AI, is het waarschijnlijk dat het Voorstel een inkijk biedt in Europese AI-regelgeving die in het verschiet ligt.⁹

Deze bijdrage bespreekt de bepalingen en implicaties van het Voorstel. Eerst zullen de reikwijdte en relevante definities worden behandeld. Hierna zullen de eisen die het Voorstel stelt aan AI worden belicht, met de nadruk op de aanvullende eisen die gelden voor hoogrisico-AI. Verder plaatst deze bijdrage een aantal kanttekeningen bij het Voorstel, waarbij de praktische haalbaarheid van sommige vereisten centraal staat. Als laatste wijdt deze bijdrage nog enige woorden aan de ogenschijnlijke spanning tussen regulering en innovatie. Zoals zal blijken, speelt deze spanning bij het Voorstel mogelijk minder dan in eerste instantie zou worden verwacht.

* Mr. S.W. van de Ven is advocaat bij Linklaters te Amsterdam. De auteur bedankt E.J. van Praag en L.J.J. van den Ende voor hun waardevolle commentaar op een eerdere versie van dit artikel.

1. Alhoewel de begrippen ‘artificiële intelligentie’, ‘robotica’ en ‘aanverwante technologieën’ niet dezelfde betekenis hebben, zijn er wel sterke overeenkomsten tussen deze begrippen. Vandaar dat de overkoepelende term AI (*artificial intelligence*) wordt gebruikt in dit artikel.
2. U. von der Leyen, A Union that strives for more. My agenda for Europe. Political guidelines for the next European Commission 2019-2024, p. 13 (<https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission-en.pdf>).
3. Europese Commissie, Witboek over kunstmatige intelligentie – een Europese benadering op basis van excellentie en vertrouwen, februari 2020.
4. Zie https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_nl.

5. Resolutie van het Europees Parlement van 20 oktober 2020 met aanbevelingen aan de Commissie betreffende een kader voor ethische aspecten van artificiële intelligentie, robotica en aanverwante technologieën (2020/2012(INL)), oktober 2020, p. 11.
6. Idem, p. 5 en 133.
7. Idem.
8. Het Rapport is uitgebracht krachtens art. 225 VWEU, op basis waarvan het Europees Parlement de Europese Commissie kan verzoeken passende wetgevingsvoorstellen in te dienen inzake aangelegenheden die naar het oordeel van het Europees Parlement besluiten van de EU vergen.
9. Witboek, p. 17.

2 Het Voorstel in het kort: definities, doelomschrijving en territoriale reikwijdte

2.1 Definities en doelomschrijving van het Voorstel

Het Voorstel definieert artificiële intelligentie als ‘een door software gestuurd of een in hardware geïntegreerd systeem dat intelligent gedrag vertoont, onder meer door het verzamelen en verwerken van gegevens, het analyseren en interpreteren van zijn omgeving, en het – met een zekere mate van autonomie – ondernemen van actie om specifieke doelstellingen te verwezenlijken’.¹⁰ Het Voorstel voorziet tevens in definities van ‘robotica’ en ‘aanverwante technologieën’, waaruit blijkt dat deze technologieën nauw verwant zijn aan artificiële intelligentie.¹¹

Naast definities van de relevante technologieën, bevat het Voorstel meerdere ondersteunende definities, zoals die van de ‘ontwikkeling’, de ‘uitrol’ en het ‘gebruik’ van AI.¹² Voor een volledig overzicht van de gehanteerde definities wordt verwezen naar artikel 4 Voorstel.

De brede reikwijdte van het begrip AI moet bewerkstelligen dat ook toekomstige ontwikkelingen worden gedekt.¹³ Dit streven naar toekomstbestendigheid komt tevens duidelijk naar voren in de doelomschrijving van het Voorstel.¹⁴ Hieruit volgt dat het Voorstel vertrouwen beoogt te scheppen door te voorzien in een toekomstbestendig raamwerk van ethische beginselen en wettelijke verplichtingen op het gebied van AI. De achterliggende ratio achter het opzetten van een dergelijk raamwerk is mede gelegen in het verbeteren van de concurrentiepositie van de EU. Er zijn namelijk aanwijzingen dat de EU in vergelijking met China en de Verenigde Staten achterloopt bij de ontwikkeling van AI.¹⁵ Het Europees Parlement hoopt de aantrekkingskracht van de EU op het gebied van AI te ver-

sterken door een voorloper te zijn in verantwoorde AI.¹⁶ Zo zouden Europese bedrijven en consumenten sneller geneigd kunnen zijn om AI te gebruiken, doordat het Voorstel de betrouwbaarheid van Europese AI garandeert. Tevens zou de EU investeringen in AI kunnen aantrekken door als een van de eerste partijen juridische duidelijkheid te verschaffen aan ontwikkelaars.

Naast de hierboven genoemde beoogde concurrentievoordelen, kan het Voorstel worden beschouwd in het licht van het *Brussels effect*. Hiermee wordt bedoeld op het vermogen van de EU om unilateraal de regels voor de wereldeconomie voor te schrijven.¹⁷ Mogelijkerwijs wordt getracht het Voorstel te laten gelden als de internationale standaard voor AI-regulering. Mocht dit streven slagen, dan zou het de geopolitieke slagkracht van de EU op het gebied van AI – en daarmee haar concurrentiepositie – sterk ten goede komen.

2.2 Hoogrisico-AI

Kenmerkend voor het Voorstel is de risicogerichte aanpak die wordt gehanteerd. Een dergelijke benadering wordt noodzakelijk geacht om toekomstige innovatie niet te belemmeren en geen onnodige lasten te creëren voor het bedrijfsleven.¹⁸ Tevens is het door de diversiteit aan AI-toepassingen lastig om te komen tot één uniforme oplossing voor alle mogelijke risico’s.¹⁹ Het Voorstel bevat vandaar een tweetal generieke vereisten voor elke vorm van AI en specifieke regelgeving voor hoogrisico-AI.

Het Voorstel definieert een ‘hoog risico’ als:

‘een aanzienlijk risico dat verbonden is aan de ontwikkeling, de uitrol en het gebruik van artificiële intelligentie, robotica en aanverwante technologieën dat schade of letsel wordt berokkend aan personen of de maatschappij, in strijd met de grondrechten en veiligheidsvoorschriften zoals vastgelegd in het recht van de Unie, rekening houdend met [1] het specifieke gebruik of het specifieke doel ervan, [2] de sector waarin zij worden ontwikkeld, uitge-

10. Art. 4 sub (a) Voorstel.

11. Art. 4 sub (c) Voorstel definieert robotica als ‘technologieën die automatisch geregelde, herprogrammeerbare multifunctionele machines in staat stellen in de fysieke wereld handelingen uit te voeren die traditioneel door mensen werden uitgevoerd of geïnitieerd door mensen, onder meer aan de hand van artificiële intelligentie of aanverwante technologieën’ en art. 4 sub (d) Voorstel definieert aanverwante technologieën als ‘technologieën die software in staat stellen met gedeeltelijke of volledige autonomie een fysiek of virtueel proces te beheersen, technologieën die in staat zijn biometrische, genetische of andere gegevens te detecteren, en technologieën die menselijke eigenschappen kopiëren of er anderszins gebruik van maken’.

12. Ofschoon de ontwikkeling, de uitrol en het gebruik van AI afzonderlijk gedefinieerde begrippen zijn in het Voorstel, zal ik ten behoeve van de leesbaarheid in het vervolg van dit artikel slechts de termen ‘ontwikkelen’ en ‘gebruiken’ of ‘ontwikkelaar’ en ‘gebruiker’ hanteren.

13. Rapport, p. 5 en 6.

14. Art. 1 Voorstel en Rapport, p. 33.

15. Europese Commissie, Digital Transformation Monitor: USA-China-EU plans for AI: Where do we stand?, januari 2018, p. 4.

16. Rapport, p. 8, waarin het volgende staat: ‘(...) meent dat een dergelijk op waarden gebaseerd regelgevingskader een meerwaarde zou vormen doordat het de Unie een uniek concurrentievoordeel verschaft en een significante bijdrage levert aan het welzijn en de welvaart van de burgers en bedrijven in de Unie doordat het de interne markt versterkt; benadrukt dat een dergelijk regelgevingskader voor AI ook meerwaarde heeft wat betreft het bevorderen van innovatie op de interne markt.’

17. Zie A. Bradford, *The Brussels effect*, Oxford: Oxford University Press 2020. Doordat de EU een grote interne markt heeft, kiezen bedrijven er vaak vrijwillig voor om aan de veelal strenge Europeesrechtelijke standaarden te voldoen, zelfs indien de producten of diensten niet in de EU worden aangeboden. Dit is gelegen in het feit dat het voor bedrijven financieel en operationeel gezien vaak niet aantrekkelijk is om gedifferentieerde standaarden te hanteren, waardoor zij vrijwillig voor de zwaarste standaard opteren. Om die reden zou de EU in staat zijn om de wereldwijde standaard voor regulering vast te stellen.

18. Rapport, p. 6.

19. Rapport, p. 6.

rold of gebruikt en [3] de ernst van het letsel of de schade die naar verwachting optreedt'.²⁰

Om invulling te geven aan criteria 1 en 2 bevat de bijlage bij het Voorstel een uitputtende lijst van zowel risicovolle toepassingen als risicovolle sectoren. Deze lijst dient regelmatig door de Europese Commissie te worden bijgewerkt (hierna: de Risicolijst).²¹ De Risicolijst bestempelt onder meer de bancaire sector, de gezondheidszorg en de publieke sector als risicovolle sectoren, en kredietverlening, politieke campagnes en geautomatiseerd rijden als risicovolle toepassingen. Als een sector of toepassing op de Risicolijst staat, betekent dit niet dat alle AI die is ontwikkeld in die sector of voor die toepassing direct kwalificeert als hoogrisico-AI. Een AI-model dat bijvoorbeeld klantenfeedback automatisch groepeerd en doorstuurt naar de relevante afdeling binnen de bank is immers vanuit risicoperspectief niet vergelijkbaar met een AI-model dat het kredietrisico van de bank voorspelt, hoewel beide modellen in de bancaire sector zijn ontwikkeld. De beoogde sector en toepassing dienen derhalve in samenhang te worden beschouwd.

Naast de sector en de beoogde toepassing is 'de verwachte ernst van het letsel of de schade' het derde criterium dat van belang is voor de risicobeoordeling. Die ernst moet worden bepaald op basis van de omvang van het mogelijke letsel of de schade, het aantal getroffen personen, de totale waarde van de veroorzaakte schade en de schade voor de samenleving als geheel.²² Als voorbeelden van ernstige soorten letsel en schade noemt het Voorstel schendingen van de rechten van kinderen, consumenten of werknemers die, gezien hun omvang, het aantal getroffen personen of hun impact op de samenleving als geheel, een aanzienlijk risico met zich meebrengen op schending van Unierechtelijke grondrechten en veiligheidsvoorschriften.²³ Ofschoon de bewoording van deze voorbeelden niet uitblinkt in duidelijkheid, legt dit criterium de lat relatief hoog. Enkel AI-toepassingen die een aanzienlijk risico met zich meebrengen op ernstig letsel of schade aan het individu of de samenleving voldoen immers aan dit criterium.

Middels het derde criterium lijkt het Voorstel gedeeltelijk tegemoet te komen aan de zorgen die veertien lidstaten, waaronder Nederland, recentelijk hebben geuit ten aanzien van het Witboek.²⁴ Zij zijn kritisch op de risicobeoordeling van AI die de Europese Commissie voorstelt in haar Witboek, die enkel is gebaseerd op de sector en de beoogde toepassing van de AI-applicatie.²⁵ Volgens de veertien lidstaten kunnen ernstige risico's niet alleen worden bepaald aan de hand van deze twee

criteria. Dit zou namelijk te veel AI als hoog risico bestemmen en een statisch beeld geven van de sectoren waarin en de toepassingen waarvoor AI tot nu toe is ontwikkeld.²⁶ De veertien lidstaten pleiten voor een risicobeoordeling die is gebaseerd op een objectieve methodologie, waarbij wordt gekeken naar de waarschijnlijkheid dat de risico's zich voordoen en de potentiële impact van de risico's.²⁷

Gezien de zorgen van de veertien lidstaten is 'de verwachte ernst van het letsel of de schade' als derde criterium een zinvolle toevoeging. Het is immers voor de verdere ontwikkeling van AI in de EU onwenselijk om elke vorm van AI waaruit enig risico voortvloeit te onderwerpen aan vergaande voorschriften. Een dergelijke regeldruk zou jonge bedrijven ernstig kunnen benadelen, omdat zij in tegenstelling tot hun gevestigde concurrenten vaak niet de capaciteit en mankracht hebben om hieraan te voldoen.

2.3 Territoriale reikwijdte

De territoriale reikwijdte van het Voorstel is evenals de definities van de relevante technologieën ruim geformuleerd. Het Voorstel is reeds van toepassing als een gedeelte van de AI-toepassing in de EU wordt ontwikkeld, uitgerold of gebruikt, ongeacht waar de door de AI-toepassing gebruikte of geproduceerde software, algoritmen of gegevens zich bevinden.²⁸ Hiermee wordt bewerkstelligd dat de voorschriften van het Voorstel al van toepassing zijn als er enig raakvlak is met de EU gedurende het gehele proces van ontwerp tot implementatie.²⁹

3 Vereisten die gelden voor alle AI-toepassingen

3.1 Het generiek vereiste van artikel 5 lid 1 Voorstel

Artikel 5 Voorstel bepaalt de normen die gelden voor alle vormen van AI, ongeacht de risicoclassificatie die hieraan is toegekend. Het eerste lid van dit artikel bevat het generieke vereiste dat AI moet worden ontwikkeld en gebruikt in overeenstemming met het Unierecht en met volledige eerbiediging van de menselijke waardigheid, autonomie en veiligheid en andere in het Handvest verankerde grondrechten.³⁰ Dit vereiste stelt zodoende een ethische ondergrens vast waaraan AI in ieder geval dient te voldoen. Humanistische beginselen als menselijke waardigheid, autonomie en veiligheid staan daarbij centraal. Een mensgerichte benadering voor de manier

20. Art. 4 (e) Voorstel. De nummering van de drie criteria is door de auteur aangebracht.

21. Art. 14 lid 4 Voorstel in samenhang met de bijlage bij het Voorstel.

22. Overweging 11 Voorstel.

23. Overweging 11 Voorstel.

24. Innovative and trustworthy AI: Two sides of the same coin, position paper on behalf of Denmark, Belgium, the Czech Republic, Finland, France Estonia, Ireland, Latvia, Luxembourg, the Netherlands, Poland, Portugal, Spain and Sweden, oktober 2020 (hierna: Position Paper).

25. Witboek, p. 17.

26. Position Paper, p. 2.

27. Position Paper, p. 2.

28. Art. 3 Voorstel.

29. Het ontwikkelproces van een AI-toepassing behelst vaak meerdere van de volgende stappen: (1) verzamelen en voorbereiden van data, (2) modelkeuze en modelimplementatie, (3) trainen van het AI-model, (4) testen en, waar nodig, valideren van het AI-model, (5) implementatie, en (6) onderhoud.

30. Handvest van de grondrechten van de EU (2012/C 326/02) (hierna: het Handvest). Het generieke vereiste van art. 5 lid 1 Voorstel alsmede vele andere voorschriften van het Voorstel zijn ook van toepassing op de software, algoritmen en gegevens die worden gebruikt of geproduceerd door AI. Om de leesbaarheid van dit artikel te bevorderen verwijst AI waar nodig ook naar de software, algoritmen en gegevens die worden gebruikt of geproduceerd door AI.

waarop AI-toepassingen worden ontwikkeld of gebruikt, is eerder ook al gepropageerd in de Ethische richtsnoeren voor betrouwbare AI, die zijn opgesteld in opdracht van de Europese Commissie (hierna: de Ethische Richtsnoeren).³¹ De Ethische Richtsnoeren schrijven voor dat AI-toepassingen moeten worden gebruikt in dienst van de mensheid en het algemeen belang, met als doel de verbetering van het welzijn en de vrijheid van de mens.³²

Doordat het generieke vereiste van artikel 5 lid 1 Voorstel een directe verwijzing naar de grondrechten van het Handvest bevat, wordt een mogelijk gebrek weggenomen. De Europese Commissie heeft namelijk opgemerkt dat momenteel het Handvest niet van toepassing is in situaties waarbij uitsluitend private partijen betrokken zijn.³³ Het Handvest bepaalt dat het van toepassing is op instellingen van de EU en lidstaten, maar zegt niets over de toepasselijkheid op relaties tussen particulieren.³⁴ De afzonderlijke verplichting die het Voorstel creëert voor zowel publieke als private partijen om in overeenstemming te handelen met het Handvest verhelpt dit potentiële gebrek.³⁵

Fundamentele rechten van het individu als ethische ondergrens voor alle vormen van AI acht ik een juist uitgangspunt. Andere vereisten, zoals transparantie en verklaarbaarheid, zijn wel belangrijk, maar niet voor alle toepassingen van AI even noodzakelijk. Veel AI-toepassingen zijn immers niet inherent risicovoller dan standaard ICT-toepassingen, waardoor dergelijke aanvullende vereisten niet nodig zijn. Dit neemt niet weg dat in de praktijk bepaalde grondrechten van het Handvest in het bijzonder relevant zijn bij de ontwikkeling en het gebruik van elke vorm van AI. Ik denk hierbij onder meer aan het recht om niet gediscrimineerd te worden en het recht op gendergelijkheid.³⁶ Als voorbeeld kan worden gedacht aan een AI-toepassing die beslist over kredietaanvragen.³⁷ Het algoritme achter deze toepassing zal sneller geneigd zijn om een vrouw af te wijzen indien succesvolle kredietaanvragen van vrouwen onvoldoende zijn vertegenwoordigd in de door het algoritme bestudeerde trainingsdata. Het algoritme consta-

teert dan dat in het verleden weinig kredietaanvragen van vrouwen zijn toegewezen en trekt hieruit de conclusie dat aan vrouwen minder snel krediet moet worden verleend. Het nalaten te controleren of de trainingsdata voldoende representatief zijn, kan aldus leiden tot discriminerende uitkomsten. Het Rapport benadrukt om die reden de noodzaak van *ethics-by-design*.³⁸ Hiermee wordt bedoeld op de strategie dat al in de ontwikkelfase de relevante ethische aspecten worden meegenomen in het ontwerp van de AI-toepassing.

3.2 Het generieke vereiste van artikel 5 lid 2 Voorstel

Naast een ethische ondergrens schrijft artikel 5 Voorstel ook voor dat elke verwerking van persoonsgegevens bij de ontwikkeling en het gebruik van AI, met inbegrip van persoonsgegevens die zijn afgeleid van niet-persoonsgegevens en biometrische gegevens, wordt uitgevoerd in overeenstemming met Richtlijn 2002/58/EG en de AVG.³⁹ De noodzaak om deze bepaling in het Voorstel op te nemen is niet direct duidelijk, nu dit vereiste ook al voortvloeit uit die regelgeving zelf. Deze bepaling lijkt dan ook met name bedoeld te zijn om het belang van een juiste verwerking van persoonsgegevens nogmaals te benadrukken.

3.3 Toezicht

Elke lidstaat dient een onafhankelijke toezichthouder aan te stellen die toezicht zal houden op naleving van het Voorstel.⁴⁰ Dit toezicht heeft betrekking op zowel de generieke bepalingen zoals besproken in deze paragraaf als de in de volgende paragraaf te behandelen aanvullende vereisten voor hoogrisico-AI.⁴¹ Voor Nederland zal dit waarschijnlijk betekenen dat een nieuwe toezichthouder moet worden opgericht, omdat de taak die het Voorstel toekent aan de nationale toezichthouders naar mijn weten geen raakvlak kent met een al bestaande Nederlandse toezichthouder. Hierbij zij opgemerkt dat het Voorstel geen heldere bevoegdheidsverdeling tussen de nationale toezichthouders vaststelt, waardoor het onduidelijk is welke nationale toezichthouder bevoegd is in grensoverschrijdende gevallen.⁴²

4 Aanvullende vereisten voor hoogrisico-AI

4.1 Inleiding

In hoofdstuk II van het Voorstel staan de aanvullende vereisten vermeld die gelden voor hoogrisico-AI. De aanvullende vereisten zullen bekend voorkomen voor eenieder die werkzaam is binnen het vakgebied. Zij dekken namelijk de principes voor verantwoorde AI die al eerder door verschillende partijen zijn geformuleerd in de vorm van *best*

31. Deskundigengroep op hoog niveau inzake kunstmatige intelligentie, Ethische richtsnoeren voor betrouwbare AI, april 2019, p. 12 en 13.

32. Idem.

33. Europese Commissie, Structure for the White Paper on artificial intelligence – A European approach, datum onbekend (gelekt), p. 11 (beschikbaar via: <https://euractiv.com/wp-content/uploads/sites/2/2020/01/AI-white-paper-EURACTIV.pdf>).

34. Art. 51 lid 1 Handvest, waarin staat dat de bepalingen van het Handvest zijn gericht tot de instellingen, organen en instanties van de EU alsmede, uitsluitend wanneer zij het recht van de EU ten uitvoer brengen, tot de lidstaten.

35. In de literatuur is beargumenteerd dat een bepaling van het Handvest wel degelijk rechtstreeks rechten en verplichtingen voor private partijen kan scheppen. Zie Asser/Hartkamp 3-I 2019/231f.

36. Art. 21 Handvest wat betreft het recht op non-discriminatie en art. 23 Handvest wat betreft het recht op gendergelijkheid.

37. Dit voorbeeld is gebaseerd op een voorbeeld vermeld in L.J.J. van den Ende & S.W. van de Ven, Voorschriften voor een verantwoord gebruik van artificial intelligence door financiële ondernemingen, FR 2020, afl. 11, p. 18.

38. Rapport, p. 8.

39. Verordening (EU) 2016/679 betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens en betreffende het vrije verkeer van die gegevens en tot intrekking van Richtlijn 95/46/EG.

40. Art. 18 lid 1 Voorstel.

41. Art. 15 lid 1 Voorstel.

42. Een dergelijke bevoegdheidsverdeling zou bijv. kunnen worden gebaseerd op de competentiebepaling van de AVG (zie art. 56 AVG).

practices.⁴³ Ook komen de aanvullende vereisten inhoudelijk grotendeels overeen met de aanbevelingen uit de Ethische Richtsnoeren. In het vervolg van deze paragraaf zal elk aanvullend vereiste voor hoogrisico-AI kort worden besproken.

4.2 Menselijk toezicht en controle (art. 7 Voorstel)

Het eerste aanvullende vereiste is dat hoogrisico-AI moet worden ontwikkeld en gebruikt op een manier die te allen tijde menselijk toezicht garandeert.⁴⁴ Dit betekent dat door AI genomen of op AI gebaseerde beslissingen onderhevig moeten zijn aan zinvolle menselijke toetsing, oordeelsvorming, tussenkomst en controle.⁴⁵ Het moet tevens altijd mogelijk zijn om de volledige menselijke controle te herwinnen. Hiermee wordt bedoeld op de mogelijkheid om op zijn minst de AI-toepassing om veiligheidsredenen te wijzigen of stop te zetten.⁴⁶ Een bekend voorbeeld van een dergelijke stopknop betreft de functie bij zelfrijdende auto's om op ieder moment de controle over de auto over te kunnen nemen als de autopilot-functie aanstaat.

4.3 Veiligheid, transparantie en verantwoording (art. 8 Voorstel)

Artikel 8 Voorstel heeft betrekking op de verantwoording voor en veiligheid en transparantie van AI-toepassingen. Het artikel begint met een aantal veiligheidsvoorschriften die trachten te bewerkstelligen dat AI op een veerkrachtige en veilige manier wordt ontwikkeld of gebruikt. Zo moet AI in overeenstemming zijn met minimale cybersecurityeisen die voorkomen dat technische kwetsbaarheden worden misbruikt en die evenredig zijn aan het vastgestelde risico.⁴⁷ Ook dient een uitwijkplan voor de AI-toepassing te worden opgesteld waarin maatregelen staan vermeld ingeval een veiligheids- of beveiligingsrisico zich voordoet.⁴⁸ Een mogelijke maatregel bij een veiligheidsrisico zou kunnen zijn dat een mens alle uitkomsten van de AI-toepassing handmatig moet goedkeuren of dat wordt teruggevallen op een niet op AI gebaseerde werkwijze.

Naast veiligheidsvoorschriften bevat artikel 8 Voorstel een aantal technische eisen die zien op de werking van de AI-toepassing. Zo moet een AI-toepassing komen tot betrouwbare en reproduceerbare uitkomsten die redelijkerwijs door de gebruiker kunnen worden verwacht.⁴⁹ Hiermee wordt

bedoeld dat de AI-toepassing functioneert zoals beoogd en ook in de toekomst zo blijft functioneren. Een andere technische eis die voortvloeit uit artikel 8 Voorstel is het voorschrift dat AI nauwkeurig dient te zijn. Nauwkeurigheid oogt als een redelijk vastomlijnd begrip, maar kent meerdere dimensies. Zo staat het bijvoorbeeld vast dat een AI-toepassing die bijstandsfraude opspoot aan de hand van voor een gemeente beschikbare data niet tot onaantvaardbaar veel foutieve voorspellingen mag komen.⁵⁰ Hiervoor is het echter van belang dat de gemeente en de ontwikkelaar eerst bepalen welk foutpercentage aanvaardbaar is. In de praktijk is dit niet altijd even gemakkelijk. Zo kan worden beargumenteerd dat een foutpercentage van 76% onacceptabel is, aangezien dit betekent dat in de meerderheid van de opgespoorde gevallen een persoon onterecht wordt bestempeld als fraudeur (zogenoemde foutpositieven). Tegelijkertijd kan worden betoogd dat een foutpercentage van 76% aanvaardbaar is als dit percentage minder bedraagt dan het foutpercentage bij een reguliere fraudemelding en de te volgen procedure na een melding niet anders is dan bij een reguliere fraudemelding.⁵¹ In dat geval is de AI-toepassing nauwkeuriger dan een reguliere fraudemelding, terwijl de consequenties van een foutpositief hetzelfde zijn. De eis van nauwkeurigheid kent zodoende verschillende aspecten waaraan de betrokken partijen vooraf vorm dienen te geven.

Artikel 8 Voorstel stelt verder dat AI voldoende transparant dient te zijn. Transparantie is een driedig principe. Transparantie betekent allereerst dat het algoritme achter de AI-toepassing verklaarbaar is, zodat de technische processen kunnen worden beoordeeld en zo nodig herzien.⁵² Transparantie betekent ook dat de elementen, processen en stadia van de AI-toepassing, zoals het onderliggende model en de trainingsfase, grondig worden gedocumenteerd.⁵³ Deze documentatieplicht moet nationale toezichthouders in staat stellen om te beoordelen of een AI-toepassing voldoet aan de voorschriften van het Voorstel. Daarnaast houdt transparantie in dat gebruikers worden geïnformeerd over het feit dat ze te maken hebben met een AI-toepassing, waarbij de capaciteiten, nauwkeurigheid en beperkingen van de AI-toepassing worden bekendgemaakt.⁵⁴

Als laatste stelt artikel 8 Voorstel dat de ontwikkelaar of gebruiker verantwoordelijk is voor de naleving van de hierboven beschreven eisen en moet kunnen aantonen dat deze in

43. Zie bijv. (1) Organisation for Economic Co-operation and Development, OECD AI principles, mei 2019 (beschikbaar via <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>),

(2) de door Microsoft opgestelde 'Microsoft AI principles' (beschikbaar via www.microsoft.com/en-us/ai/responsible-ai) en (3) de door Google geformuleerde 'AI principles' (beschikbaar via <https://ai.google/principles>). In totaal zouden er al meer dan 84 van dit soort initiatieven zijn die ethische principes voor AI voorschrijven. Zie hierover B. Mittelstadt, Principles alone cannot guarantee ethical AI, *Nature Machine Intelligence* (1) 2019, p. 501.

44. Art. 7 Voorstel.

45. Overweging 10 Voorstel.

46. Overweging 10 jo. art. 8 lid 1 sub (g) Voorstel.

47. Art. 8 lid 1 sub (a) Voorstel.

48. Art. 8 lid 1 sub (b) Voorstel.

49. Art. 8 lid 1 sub (c) Voorstel.

50. Dit voorbeeld is ontleend aan de AI-toepassing van de gemeente Nissewaard die bijstandsfraude opspoot. Zie 'Gemeente Nissewaard spoort bijstandsfraude op met AI', in: *AI-magazine*, AI als instrument voor gemeenten, september 2020 (beschikbaar via <https://ai-magazine.vngrealisatie.nl/>).

51. Dit was het argument van de gemeente Nissewaard, zoals blijkt uit een gelekt document (beschikbaar via <https://bijvoorbeeldverdacht.nl/wp-content/uploads/2019/12/2019.07.01-Memo-handhaving-Participatiewet.pdf>).

52. Art. 8 lid 1 sub (e) Voorstel.

53. Art. 8 lid 2 Voorstel. Zie noot 29 voor een beschrijving van de verschillende ontwikkelingsfasen van een AI-toepassing.

54. Art. 8 lid 1 sub (f) Voorstel.

acht worden genomen.⁵⁵ Tevens moeten de maatregelen die zijn genomen om aan deze eisen te voldoen controleerbaar zijn voor de nationale toezichthouder of, indien van toepassing, een nationale of Europese sectorale toezichthouder.⁵⁶

4.4 Geen vooringenomenheid en discriminatie (art. 9 Voorstel)

Artikel 9 Voorstel betreft het vereiste dat een AI-toepassing niet mag discrimineren of vooringenomen mag zijn.⁵⁷ Dit vereiste is in paragraaf 3.1 al besproken in het kader van de ethische ondergrens die het Voorstel in het leven roept. Van daar dat niet verder op dit vereiste wordt ingegaan. Hierbij zij wel vermeld dat het Voorstel een uitzondering bevat op het gelijkheidsbeginsel. Een afwijkende behandeling door een AI-toepassing is namelijk toelaatbaar voor zover er geen minder ingrijpende alternatieven bestaan en er een objectief, redelijk en legitiem doel is dat zowel evenredig als noodzakelijk is.⁵⁸ Een voorbeeld van een situatie waarin deze uitzondering mijns inziens zou kunnen slagen, betreft een AI-toepassing die de distributie van vaccinaties onder de bevolking optimaliseert en hierbij leeftijd meeneemt als belangrijke variabele. Mocht op deze uitzondering worden vertrouwd, dan is het in het licht van de documentatieplicht raadzaam om zowel het legitieme doel als de alternatieven nauwkeurig vast te leggen.

4.5 Maatschappelijke verantwoordelijkheid en gendergelijkheid (art. 10 Voorstel)

Artikel 10 Voorstel vereist dat hoogrisico-AI niet mag worden ingezet om verkiezingen te hinderen of mag bijdragen aan de verspreiding van desinformatie. Daarnaast verplicht dit artikel dat AI de rechten van werknemers eerbiedigt, kwalitatief onderwijs en digitale geletterdheid bevordert, de genderkloof niet vergroot en intellectuele eigendomsrechten alsmede eventuele beperkingen of uitzonderingen daarop respecteert.⁵⁹ Deze vereisten zien op een breed spectrum aan maatschappelijke problemen waarvan wordt aangenomen dat AI deze (onbedoeld) kan verergeren. Zo zijn de AI-toepassingen van Cambridge Analytica gebruikt om verschillende democrati-

sche processen te beïnvloeden, waaronder de Amerikaanse verkiezingen van 2016.⁶⁰ Een voorbeeld waaruit volgt dat AI op gespannen voet kan staan met de rechten van werknemers betreft de repetitieve taken die worden verricht ten behoeve van AI.⁶¹ Het gaat dan onder meer om het handmatig labelen van afbeeldingen als trainingsdata voor AI-modellen of het identificeren van aanstootgevende inhoud in de nieuwsfeed van sociale media.⁶² Naast het feit dat werknemers veelal slecht worden beloond hiervoor, zijn zij gevoelig voor allerlei psychische problemen, bijvoorbeeld doordat zij aanstootgevend videomateriaal moeten beoordelen.⁶³

4.6 Milieuduurzaamheid (art. 11 Voorstel)

Uit artikel 11 Voorstel volgt dat hoogrisico-AI dient te worden beoordeeld op milieuduurzaamheid door de nationale toezichthouder of, indien van toepassing, andere nationale of Europese sectorale toezichthouders.⁶⁴ Hierbij moet worden gekeken of maatregelen zijn genomen om de algemene impact van de AI-toepassing op natuurlijke hulpbronnen, energieverbruik, afvalproductie, koolstofvoetafdruk, klimaatverandering en milieuvervuiling te verminderen.⁶⁵ Deze maatregelen dienen tevens te bewerkstelligen dat toepasselijke wetgeving alsmede door de EU aangegeven internationale milieuvverbintenissen worden nageleefd.

De door AI-toepassingen vereiste rekenkracht gaat normaal gesproken gepaard met veel elektriciteitsverbruik. Naast rekenkracht heeft AI, in het bijzonder *machine learning*-modellen, grote hoeveelheden data nodig om te kunnen functioneren. Die dataopslag kost eveneens veel elektriciteit. Zo zouden in de Verenigde Staten datacentra al ongeveer 2% van het totale elektriciteitsverbruik voor hun rekening nemen.⁶⁶ Een bepaling die voorschrijft dat bij de ontwikkeling en het gebruik van hoogrisico-AI de algemene impact op het milieu moet worden meegenomen, is derhalve een welkome aanvulling, mede gezien de verwachting dat de inzet van AI sterk zal toenemen.

55. Art. 8 lid 2 Voorstel. Het in par. 4.2 besproken voorschrift om een stopknop te hebben valt ook onder deze veiligheidseisen.

56. Art. 8 lid 3 Voorstel.

57. Art. 4 sub (l) Voorstel definieert vooringenomenheid als 'een bevooroordeelde persoonlijke of sociale perceptie van een persoon of groep personen op basis van hun persoonlijke eigenschappen' en art. 4 sub (m) Voorstel definieert discriminatie als 'een afwijkende behandeling van een persoon of groep personen op een grond die niet objectief of redelijkerwijs te rechtvaardigen is en daarom door de Uniewetgeving wordt verboden'.

58. Art. 9 lid 2 Voorstel. Deze uitzondering komt sterk overeen met de uitzondering voor indirecte discriminatie in art. 2 lid 2 sub (b) onderdeel (i) van Richtlijn 2000/78/EG van de Raad van 27 november 2000 tot instelling van een algemeen kader voor gelijke behandeling in arbeid en beroep en art. 2 lid 2 sub (b) van Richtlijn 2000/43/EG van de Raad van 29 juni 2000 houdende toepassing van het beginsel van gelijke behandeling van personen ongeacht ras of etnische afstamming.

59. Het is mij niet gelijk duidelijk hoe van alle hoogrisico-AI kan worden vereist dat zij kwalitatief onderwijs en digitale geletterdheid bevordert, aangezien veel AI-toepassingen geen raakvlak hebben met het onderwijs. Een beter vereiste zou zijn dat hoogrisico-AI kwalitatief onderwijs en digitale geletterdheid niet mag belemmeren.

60. The Economist asks: Is democracy safe in the age of big data?, The Economist, augustus 2018.

61. Hierbij zij vermeld dat AI in veel gevallen juist wordt ingezet om repetitieve taken te automatiseren.

62. Europees Parlement, The ethics of artificial intelligence: Issues and initiatives, maart 2020, p. 9. Het handmatig labelen van data is noodzakelijk voor *supervised learning*. Dit is een leermethode voor bepaalde AI-modellen waarbij de mens per geval een classificatie heeft aangebracht (zoals 'hond' of 'kat'). Het AI-model bestudeert vervolgens deze geclassificeerde data en probeert, aan de hand van de door de mens gemaakte tweedeling, gemeenschappelijke eigenschappen te vinden waarmee het voorspellingen kan maken (zoals: op deze foto staat een hond en geen kat). Voor meer informatie hierover, zie Van den Ende & Van de Ven 2020, p. 15.

63. Idem.

64. Art. 11 Voorstel.

65. Art. 11 Voorstel.

66. Europees Parlement 2020, p. 28.

4.7 *Eerbiediging van privacy en bescherming van persoonlijke gegevens (art. 12 Voorstel)*

Artikel 12 Voorstel heeft betrekking op het gebruik van biometrische gegevens voor identificatiedoeleinden, zoals gezichtsherkenning, in openbare ruimten. Het Voorstel erkent dat aan een dergelijk gebruik specifieke risico's voor grondrechten kleven en stelt dat alleen openbare autoriteiten van de lidstaten hiertoe bevoegd zijn voor doeleinden van aanzienlijk openbaar belang. Ook dienen openbare autoriteiten transparant te zijn omtrent het gebruik van biometrische gegevens voor identificatiedoeleinden in openbare ruimten. Daarnaast moet een dergelijk gebruik beperkt blijven tot specifieke doeleinden en in tijd, in overeenstemming met het toepasselijke recht, met name de AVG, en met inachtneming van de menselijke waardigheid en autonomie en de in het Handvest opgenomen grondrechten.

4.8 *De controlelijst zoals opgenomen in de Ethische Richtsnoeren als implementatiekader*

Om invulling te geven aan de aanvullende vereisten voor hoogrisico-AI kan de in de Ethische Richtsnoeren opgenomen controlelijst voor betrouwbare AI worden doorlopen.⁶⁷ Deze controlelijst stelt meerdere vragen die als handvat functioneren om verschillende principes, waaronder menselijk toezicht, betrouwbaarheid en non-discriminatie, te duiden. Vanwege de overeenkomstige principes zal de controlelijst in veel gevallen een handig implementatiekader zijn voor ondernemingen die naleving van de aanvullende vereisten willen vormgeven.

5 Een Europees certificaat van ethische conformiteit voor hoogrisico-AI

5.1 *Het Europese certificaat van ethische conformiteit voor hoogrisico-AI*

Een van de meest ingrijpende bepalingen van het Voorstel is de eis dat hoogrisico-AI in het bezit dient te zijn van een Europees certificaat van ethische conformiteit. Dit certificaat wordt verschaft nadat de nationale toezichthouder zowel een risico- als conformiteitsbeoordeling van de AI-toepassing heeft uitgevoerd.⁶⁸ Krachtens artikel 14 Voorstel dient elke AI-toepassing eerst te worden onderworpen aan een risicobeoordeling.⁶⁹ Dit houdt in dat de nationale toezichthouder aan de hand van de in paragraaf 2.2 besproken criteria beoordeelt of de AI-toepassing kwalificeert als hoogrisico-AI. Indien het antwoord positief luidt, beoordeelt de nationale toezichthouder vervolgens of de AI-toepassing in overeenstemming is met de aanvullende vereisten die gelden voor hoogrisico-AI.⁷⁰ Mocht dit zo zijn, dan verleent de nationale toezichthouder het Europese certificaat van ethische conformiteit aan de desbetreffende AI-toepassing.⁷¹ Eenmaal

verkregen is het certificaat geldig in de gehele EU. Een partij die gecertificeerde hoogrisico-AI gebruikt in de EU zal zich ervan moeten vergewissen dat de AI-toepassing in overeenstemming blijft met de voorschriften van het Voorstel. Een over het Europese certificaat van ethische conformiteit beschikkende AI-toepassing dient namelijk doorlopend te voldoen aan de eerder beschreven generieke vereisten en de aanvullende vereisten voor hoogrisico-AI.⁷²

5.2 *Het vrijwillig verkregen Europese certificaat van ethische conformiteit*

Voor laagrisico-AI bestaat de mogelijkheid om vrijwillig een Europees certificaat van ethische conformiteit aan te vragen. Hiervoor zal dezelfde procedure moeten worden doorlopen als die voor hoogrisico-AI.⁷³ De wenselijkheid van een vrijwillig betrouwbaarheidslabel zal mijns inziens afhangen van de tijdsduur van de aanvraagprocedure en de kosten die hiervoor in rekening worden gebracht. Indien de tijdsduur lang is en de kosten relatief hoog zijn, bestaat het risico dat gevestigde namen een voordeel genieten. Zij beschikken immers eerder over de financiële middelen en de vereiste mankracht om de aanvraagprocedure succesvol te doorlopen. Met het vrijwillig verkregen betrouwbaarheidslabel zouden zij vervolgens partijen gemakkelijker ertoe kunnen verleiden om hun AI te gebruiken in plaats van AI zonder betrouwbaarheidslabel. Een laagdrempelige en efficiënte aanvraagprocedure zou dit nadelige effect kunnen verhelpen.

5.3 *De coördinerende rol van de Europese Commissie*

Het Voorstel kent een belangrijke rol toe aan de Europese Commissie. Zij heeft allereerst een coördinerende functie bij zowel de risico- als de conformiteitsbeoordeling.⁷⁴ Uniformiteit is cruciaal voor het succes van het Voorstel, aangezien een gefragmenteerde toepassing door nationale toezichthouders zou kunnen leiden tot regelgevingsarbitrage binnen de EU. Om dit te voorkomen heeft de Europese Commissie de taak ervoor te zorgen dat de risicobeoordeling consistent wordt uitgevoerd.⁷⁵ Hiervoor dient zij onder meer een lijst bij te houden van alle AI die is gekwalificeerd als hoogrisico-AI.⁷⁶ Nationale toezichthouders kunnen deze lijst als leidraad gebruiken bij het uitvoeren van hun risicobeoordelingen.

Verder dient te Europese Commissie bindende richtsnoeren vast te stellen over de toe te passen methodologie bij de conformiteitsbeoordeling, waardoor de beoordelingsvrijheid

67. Ethische Richtsnoeren, p. 32-41.

68. Art. 16 lid 1 Voorstel.

69. Alhoewel art. 14 Voorstel zelf niet direct spreekt over het tijdstip waarop de risicobeoordeling moet worden uitgevoerd, volgt uit het Rapport (p. 9) dat het een ex-antebeoordeling betreft.

70. Art. 15 Voorstel.

71. Art. 16 lid 1 Voorstel.

72. Dit volgt met zoveel woorden uit art. 5 lid 1 en 6 lid 2 Voorstel.

73. Art. 16 lid 2 Voorstel. Opvallend is dat deze bepaling ook aangeeft dat laagrisico-AI niet onderhevig is aan de risicobeoordeling van art. 15 Voorstel. Dit lijkt mij foutief, omdat het feit dat AI niet kwalificeert als hoogrisico-AI enkel kan worden vastgesteld op basis van deze risicobeoordeling.

74. Onder de Europese Commissie dient mede te worden verstaan alle relevante instellingen, organen, bureaus en agentschappen van de EU die voor dit doel kunnen worden aangewezen. Zie art. 14 lid 2 jo. art. 15 lid 1 Voorstel.

75. Art. 20 lid 1, eerste punt, Voorstel.

76. Art. 14 lid 3 Voorstel.

van nationale toezichthouders wordt ingeperkt.⁷⁷ De Europese Commissie is ook verantwoordelijk voor het opstellen van een aanvraagprocedure voor het Europese certificaat van ethische conformiteit. Een gemeenschappelijke aanvraagprocedure draagt bij aan uniformiteit, omdat partijen dezelfde informatie moeten aanleveren en gebonden zijn aan hetzelfde proces.

6 Beschouwing van het Voorstel

Als een van de eerste AI-regelgevingsinitiatieven betreft het Voorstel onontgonnen gebied. Om die reden kost het weinig moeite om een kritische noot te plaatsen bij het Voorstel. Daarnaast bevat het Voorstel, zijnde een voorstel voor conceptregulering, nog een veelvoud aan onderwerpen die verduidelijking behoeven.⁷⁸ In het vervolg van deze paragraaf zal ik mij daarom beperken tot enkele meer elementaire aandachtspunten die mij in het oog sprongen na lezing van het Voorstel.

Om te beginnen vraag ik mij af in hoeverre het mogelijk is om het gehele scala aan vereisten van het Voorstel mee te nemen in de conformiteitsbeoordeling. Deze vereisten hebben betrekking op een diversiteit aan onderwerpen, uiteenlopend van grondrechten en duurzaamheidsnormen tot technische veiligheidsvoorschriften. Hierdoor is het bij een toetsing lastig om met zekerheid vast te stellen dat een AI-toepassing aan alle vereisten voldoet, terwijl het certificaat van ethische conformiteit dit wel veronderstelt. Hier komt bij dat het gebruikelijk is dat componenten van een AI-toepassing zijn gebaseerd op *open source libraries*.⁷⁹ Een toetsing van een AI-toepassing zou dan ook een toetsing van deze doorlopend veranderende *libraries* moeten behelzen. Het valt te betwijfelen of een dergelijke omvangrijke beoordeling in de praktijk efficiënt en betaalbaar kan geschieden zonder dat Europese eindgebruikers, in tegenstelling tot de rest van de wereld, een lange tijd moeten wachten alvorens hoogrisico-AI kan worden ingezet. Ik neem daarom aan dat certificering voornamelijk als doel heeft om bewustwording te creëren en de overduidelijke risicogeveallen tegen te houden.

In aanvulling op het eerste punt vraag ik me af in hoeverre het gewenst is om het concept van certificering te gebruiken voor AI-toepassingen. Door het verlenen van een Europees certificaat van ethische conformiteit impliceren nationale toezichthouders namelijk dat de AI-toepassing in overeenstemming is met de eisen van het Voorstel. Bepaalde vormen van AI zijn echter zelflerend, waardoor ongewenste implicaties pas op een later moment in de AI-toepassing kunnen sluipen of zichtbaar

worden.⁸⁰ Hierbij kan worden gedacht aan een stemherkenningsdienst, zoals Amazons Alexa, die in allerlei verschillende situaties wordt gebruikt en op basis daarvan voortdurend evolueert. Een ex-anteconformiteitsbeoordeling zal dan slechts schijnzekerheid bieden aan het publiek.⁸¹ Kortom, door de sterk veranderende aard van sommige AI-toepassingen lijkt certificering niet altijd geschikt voor AI.

Een derde kanttekening ziet op de praktische haalbaarheid van de risicobeoordelingen. Het Voorstel schrijft nu voor dat een nationale toezichthouder voor elke AI-toepassing moet bepalen of deze kwalificeert als hoogrisico-AI. Gezien de vele reeds bestaande AI-toepassingen en de verwachting dat AI de komende jaren een enorme vlucht gaat nemen, is dit een haast onmogelijke opgave.⁸² Hierbij is het relevant om te vermelden dat de Autoriteit Persoonsgegevens, zijnde een niet-sectorale toezichthouder overeenkomstig de nationale toezichthouder krachtens het Voorstel, sinds de inwerkingtreding van de AVG in 2018 al een enorme achterstand heeft opgebouwd wat betreft de klachtenafhandeling van burgers en de goedkeuring van de doorgifte van persoonsgegevens naar derde landen.⁸³ Het lijkt mij daarom noodzakelijk dat een zekere schifting plaatsvindt, zodat AI-toepassingen die evident een laag risico hebben niet hoeven te worden voorgelegd aan de toezichthouder. Het is om die reden onvermijdelijk om toch te vertrouwen op de risicobeoordeling van de ontwikkelaar zelf. Deze zelfevaluatie zou bij voorkeur moeten worden gedaan op basis van criteria die door de toezichthouder zijn vastgesteld, met strenge sancties voor foute classificaties.

Een vierde kanttekening heeft betrekking op artikel 17 lid 1 Voorstel. Deze bepaling schrijft voor dat AI aan de relevante governancemaatstaven dient te voldoen die zijn vastgesteld door de nationale toezichthouders onder coördinatie van de Europese Commissie. Hieruit leid ik af dat nationale toezichthouders aanvullende maatstaven mogen vaststellen ten aanzien van AI, zelfs indien het laagrisico-AI betreft. Naast het feit dat dit zou kunnen leiden tot gefragmenteerde regelgeving in de EU, laat dit de deur open om laagrisico-AI alsnog te onderwerpen aan meer voorschriften. Deze conclusie lijkt haaks te staan op de risicogerichte aanpak van het Voorstel, waarbij alleen hoogrisico-AI moet voldoen aan specifieke regelgeving en laagrisico-AI enkel aan een tweetal generieke vereisten. Als voorbeeld van bestaande governancemaatstaven noemt het Voorstel de Ethische Richtsnoeren.⁸⁴ Zoals vermeld in paragraaf 4.1 komen de Ethische Richtsnoeren en de

77. Art. 15 lid 3 Voorstel.

78. Zo is het onduidelijk of er een *grandfathering*-regime komt op basis waarvan de regels van het Voorstel niet of beperkt zouden gelden voor bestaande AI-toepassingen. Ook is het bijv. niet duidelijk of de ontwikkelaar dan wel de partij die de AI-toepassing inzet verantwoordelijk is voor het aanvragen van het Europese certificaat van ethische conformiteit, mochten deze partijen verschillen.

79. Voorbeelden van *open source libraries* zijn het veelgebruikte Scikit-learn en het door Google ontwikkelde TensorFlow.

80. Het gaat hier met name om *machine learning*-toepassingen. *Machine learning* is een tak van AI waarbij algoritmen worden opgesteld die zelfstandig kunnen leren van input.

81. Art. 15 lid 2 Voorstel biedt wel de mogelijkheid om de conformiteitsbeoordeling nogmaals uit te voeren.

82. Zie voor een beschrijving waarom AI een enorme vlucht zal nemen in de komende jaren S.W. van de Ven, De veranderende taak van bestuurders in het tijdperk van AI, MvO 2020, afl. 7, p. 194.

83. Voorzitter Autoriteit Persoonsgegevens: 'De achterstanden zijn zo groot dat het lachwekkend is', Trouw 10 november 2020.

84. Overweging 41 Voorstel.

aanvullende vereisten voor hoogrisico-AI qua inhoud groten-deels met elkaar overeen. Het op risico gebaseerde onderscheid komt hierdoor te vervagen, nu alle vormen van AI in overeenstemming moeten zijn met vergelijkbare maatstaven. Met het oog op de uniformiteit van regelgeving en de beoogde risicogerichte aanpak is het aldus wenselijk om artikel 17 lid 1 Voorstel te wijzigen. De relevante governancemaatstaven zouden op Europees niveau moeten worden vastgesteld en enkel dienen te gelden voor hoogrisico-AI.

De laatste kanttekening is gerelateerd aan het gebrek aan sancties. Het Voorstel voorziet namelijk niet in een bepaling die toezichthouders de bevoegdheid geeft om normschending te sanctioneren.⁸⁵ Alhoewel de kwalificatie ‘tandeloze tijger’ niet optimaal is voor een voorstel voor conceptregelgeving die vertrouwen moet creëren in AI, kan de afwezigheid van een sanctieregime gelukkig eenvoudig worden geredificeerd in een finaal wetgevingsvoorstel. Het zou logisch zijn om dit sanctie-regime te laten aansluiten bij de AVG, gezien de bekendheid van marktpartijen hiermee en de deels overlappende vereisten.⁸⁶ Dit betekent forse administratieve geldboeten die kunnen oplopen tot EUR 20 miljoen of, indien hoger, 4% van de totale wereldwijde jaaromzet in het voorgaande boekjaar.⁸⁷ Een sanctieregime zou mijns inziens wel gepaard moeten gaan met een strikte afbakening van verantwoordelijkheden om de ontwikkeling van AI niet te ontmoedigen. Zo zou de ontwikkelaar van een AI-toepassing niet mogen worden gesanctioneerd indien die toepassing vervolgens, zonder medeweten van de ontwikkelaar, wordt ingezet in strijd met de voorschriften van het Voorstel.

7 Slotwoord

Met het Voorstel heeft het Europees Parlement een aanzet gegeven voor Europese regelgeving op het gebied van AI. Het is nu aan de Europese Commissie om te beoordelen in hoeverre zij de tekst van het Voorstel wil overnemen in haar conceptverordening, die in het eerste kwartaal van 2021 wordt verwacht.⁸⁸ Wezenlijke verschillen tussen die twee zijn niet te verwachten. De reden hiervoor is dat het Voorstel de structuur volgt die de Europese Commissie al had uiteengezet in haar Witboek. Net als het Witboek maakt het Voorstel een onderscheid tussen hoogrisico- en laagrisico-AI, waarbij uitgebreide voorschriften enkel gelden voor de hoogrisicocategorie. Tevens voorziet het Voorstel, overeenkomstig het Witboek, in verplichte certificering voor hoogrisico-AI en vrijwillige certificering voor laagrisico-AI.

Het is lastig vast te stellen of de door het Voorstel en het Witboek ingeslagen weg de juiste vorm van regulering behelst. Het

succesvol reguleren van nieuwe technologieën draait om het vinden van een delicate balans tussen veiligheid en innovatie. Een balans die volgens marktpartijen waarschijnlijk niet snel zal worden bereikt.⁸⁹ Tegelijkertijd hoeft regulering niet noodzakelijkerwijs op gespannen voet te staan met innovatie. Maatregelen die de transparantie en veiligheid van AI-toepassingen waarborgen, kunnen immers zelf een stimulerende factor voor innovatie vormen, doordat zij vertrouwen bij eindgebruikers scheppen. Door het creëren van vertrouwen in AI te beschouwen als primaire doelstelling van het Voorstel lijkt het Europees Parlement zich hiervan bewust. Het is goed mogelijk dat de voorschriften van het Voorstel, zoals de aanvullende vereisten voor hoogrisico-AI, aan deze doelstelling zullen bijdragen. Innoveren door te reguleren: als jurist klinkt het haast te mooi om waar te zijn.

85. Art. 13 Voorstel stelt wel dat elk persoon het recht heeft om verhaal te halen voor letsel of schade die is berokkend door de ontwikkeling of het gebruik van hoogrisico-AI, waarbij het Unierecht en de in het Voorstel vastgestelde verplichtingen zijn geschonden.

86. Zie par. 3.2 en 4.7 wat betreft de overlappende vereisten.

87. Zie art. 83 AVG.

88. Zie <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements>.

89. Uit consultatiereacties van onder meer Facebook, Google en Microsoft op het Witboek volgt dat zij sterk tegen een ex-anteconformiteitsbeoordeling zijn die wordt uitgevoerd door een toezichthouder. Niet geheel verrassend stellen zij een structuur voor die grotendeels is gebaseerd op zelfevaluatie. Zie Facebook, Facebook's comments on European Commission White Paper on artificial intelligence – A European approach, juni 2020, p. 18 (<https://ai.facebook.com/blog/collaborating-on-the-future-of-ai-governance-in-the-eu-and-around-the-world/>); Google, Consultation on the White Paper on AI – A European approach, mei 2020, p. 33 (www.blog.google/documents/77/Googles_submission_to_EC_AI_consultation_1.pdf); Microsoft, Response of Microsoft Corporation to the European Commission's White Paper on artificial intelligence, juni 2020, p. 13 (<https://blogs.microsoft.com/eupolicy/2020/06/15/microsoft-response-european-commission-white-paper-ai/>).