Universiteit
Leiden
The Netherlands

Technologie (AI & blockchain) en ESG; de Europese benadering van de digitaliseringstransitie: deel II

Wuisman, I.S.

Citation
Wuisman, I. S. (2023). Technologie (AI & blockchain) en ESG; de Europese benadering van de digitaliseringstransitie: deel II. *Ondernemingsrecht, 2023*(10/11), 429-448. Retrieved from https://hdl.handle.net/1887/3728371

Version:	Publisher's Version
License:	Leiden University Non-exclusive license
Downloaded from:	https://hdl.handle.net/1887/3728371

Note: To cite this publication please use the final published version (if applicable).

Iris Wuisman[1]

Technologie (AI & blockchain) en ESG; de Europese benadering van de digitaliseringstransitie – Deel II

Ondernemingsrecht 2023/64

In Deel I[2] van wat inmiddels is uitgegroeid tot een drieluik, heb ik besproken op welke wijze digitalisering en dan met name kunstmatige intelligentie en blockchain kunnen bijdragen aan de duurzaamheidstransitie en heb ik uiteengezet met welke ESG-uitdagingen ondernemingen geconfronteerd kunnen worden bij de ontwikkeling en het gebruik van deze technologieën. In dit tweede deel van het drieluik bespreek ik de belangrijkste digitaliseringswetgeving(sinitiatieven) en daarmee samenhangende andere reguleringsinstrumenten en hoe deze zich tot ESG verhouden. De focus ligt op data en de voorgestelde regulering van kunstmatige intelligentie omdat de ontwikkelingen op het gebied van deze technologie nu zeer rap gaan. Zowel op het niveau van soft law als op dat van verordeningen en richtlijnen zijn er allerlei Europese initiatieven en voorstellen gepubliceerd die een verantwoordelijke ontwikkeling en inzet van technologie trachten te borgen. Er vindt een ontwikkeling plaats waarbij beginselen van betrouwbare AI (aandacht voor de effecten op de mens) verschuiven naar beginselen van duurzame AI (aandacht voor de effecten op het brede spectrum van ESG). Ook worden deze beginselen in toenemende mate in de wetgeving zelf opgenomen. De hiervoor gekozen systematiek zou echter niet moeten leiden tot een situatie waarin alleen aanbieders en gebruikers van AI-systemen met een hoog risico deze beginselen in acht nemen. Aanbieders van foundationmodellen, modellen die aan een veelheid van AI-systemen in diverse domeinen ten grondslag kunnen liggen, ontlopen straks niet meer de dans. Verschillen in de voorgestelde aanpak van risico's door actoren in de AI-waardeketen zijn echter niet altijd goed te plaatsen. In Deel III ga ik in op de duurzaamheidswetgeving in het licht van de digitaliseringstransitie en sluit ik het drieluik af met een reflectie op de ontwikkelingen in de arena van technologie en ESG. Aansprakelijkheidsregels laat ik buiten beschouwing.

1. Inleiding

Na een periode met veel aandacht voor blockchaintechnologieën, is de focus de afgelopen jaren geleidelijk verschoven naar kunstmatige intelligentie met een explosie in de laatste paar maanden. De ontwikkeling van kunstmatige intelligentie heeft een vlucht genomen vanwege de beschikbaarheid van veel meer (computer)capaciteit en data. ChatGPT, gebaseerd op GPT-3,5 en de opvolger GPT-4,[3] hebben in de afgelopen tijd veel losgemaakt.[4] Deze generatieve kunstmatige intelligentie, dat wil zeggen kunstmatige intelligentie die in reactie op een door de gebruiker ingegeven opdracht nieuwe content creëert op basis van bestaande data, opent nieuwe mogelijkheden voor allerlei soortige contentcreatie en werkwijzen.[5] De ontwikkeling van deze generatieve modellen zorgt aan de ene kant voor een jubelstemming, gelet op alle kansen die deze software biedt, aan de andere kant creëert het ook veel zorgen. Zo verbood Italië in eerste instantie de technologie vanwege de privacy van gebruikers. Het verbod is inmiddels weer opgeheven na onder andere verbetering van de software door toegenomen transparantie over gegevensverwerking en het ter beschikking stellen van zogenoemde 'opt-out'-rechten (bijvoorbeeld het recht dat conversaties met de software niet worden gebruikt voor trainingsdoeleinden).[6] Maar ook na dit soort aanpassingen blijft er veel onduidelijkheid over de onderliggende technologie bestaan omdat

1 Prof. mr. drs. I.S. (Iris) Wuisman is hoogleraar Ondernemingsrecht aan de Universiteit Leiden, partner bij ACE Management Consulting en redacteur van dit tijdschrift. Dit onderzoek is onderdeel van het Society Artificial Intelligence and Life Sciences (SAILS)-project van de Universiteit Leiden. Het onderzoek is afgesloten op 26 juni 2023.

2 *Ondernemingsrecht* 2023/54.

3 https://openai.com/product/gpt-4. Deze versie kan ten opzichte van GPT-3,5 qua input zowel tekst als afbeeldingen verwerken (daarmee is de software multimodaal geworden; de output is wel nog steeds alleen tekst), kan in plaats van 3000 woorden 25.000 woorden als input verwerken, scoort veel hoger op correcte uitkomsten en is in staat om complexere teksten te analyseren: *Financial Times* (2023), 'GPT-4 from OpenAI shows advances – and moneymaking potential', 19 maart 2023; https://www.ft.com/content/af75643c-9ddb-49ad-8df9-2ffb02484e38.

4 GPT staat voor 'Generative Pre-trained Transformer'-model. Deze modellen zijn in staat om door middel van zelf-aanpassingsmechanismen bestaande data om te zetten in nieuwe content.

5 De testen die OpenAI, het bedrijf achter de GPT-software, heeft uitgevoerd met de nieuwste versie die in maart 2023 is gelanceerd, resulteerden in met de mens vergelijkbare uitkomsten. Bij het afnemen van examens zoals de US Bar Exam en de SAT (de toets die afgelegd dient te worden voordat iemand in de Verenigde Staten als advocaat aan de slag mag respectievelijk de Amerikaanse schoolexamens) behaalde het taalmodel scores behorend bij de *top* 10%. GPT-3,5 zat daarentegen bij de laagst scorende 10%. OpenAI (2023), 'GPT-4 Technical Report', p. 1, en *Financial Times* (2023), 'GPT-4 from OpenAI shows advances – and moneymaking potential', 19 maart 2023; https://www.ft.com/content/af75643c-9ddb-49ad-8df9-2ffb02484e38. Zie ook: https://openai.com/research/gpt-4.

6 Andere maatregelen die genomen moesten worden, waren het opstellen en implementeren van een plan voor het voorkomen van gebruik van de software door kinderen onder de 13 jaar zonder ouderlijke toestemming en de toezegging van een mediacampagne om Italiaanse inwoners te informeren over de persoonlijke gegevensverwerking door de software; https://www.politico.eu/article/chatgpt-italy-lift-ban-garante-privacy-gdpr-openai/.

OpenAI niet zo 'open' is over haar AI-tool.[7] En daarin staat dit bedrijf niet alleen. Ook andere techgiganten houden de kaarten het liefst tegen de borst. Ook bestaat er nog veel onduidelijkheid over de (langetermijn)effecten van AI op de maatschappij alsook de impact op het klimaat en milieu vanwege het bijbehorende energieverbruik. De afgelopen jaren heeft er echter een maatschappelijke verschuiving plaatsgevonden waarbij de druk op bedrijven om transparanter en duurzamer te ondernemen, is toegenomen. Zowel op het niveau van soft law als op dat van verordeningen en richtlijnen zijn er allerlei initiatieven en voorstellen op het gebied van digitalisering gepubliceerd.[8] Deze reguleringsgolf heeft tot diverse discussies geleid. Niet alleen vanwege de toename in regeldruk en de mogelijke gevolgen daarvan voor de concurrentiepositie van ondernemingen in Europa,[9] maar bijvoorbeeld ook als gevolg van verschillende opvattingen over wat in Europa acceptabel wordt geacht als het gaat om inzet van technologie. In dit Deel II van het drieluik over technologie en ESG bespreek ik de belangrijkste digitaliseringswetgeving(sinitiatieven) en daarmee samenhangende andere reguleringsinstrumenten tegen de achtergrond van ESG. De focus ligt op kunstmatige intelligentie, omdat de ontwikkelingen op het gebied van deze technologie nu zeer rap gaan. Aansprakelijkheidsregels laat ik in dit artikel buiten beschouwing.

2. Digitaliseringswet- en regelgeving

Regulering van technologie is een veelzijdige aangelegenheid. Het Europese recht bevat daardoor veel verschillende soorten 'digitale regels' die al zijn aangenomen of nog in de maak zijn.[10] Het gevolg is dat er een web aan regels bestaat, waarbij de regels in veel gevallen gestapeld toegepast dienen te worden. Het betekent ook dat de Europese wetgever er goed voor dient te zorgen dat deze verschillende sets van regels compatibel met elkaar zijn en er geen leemtes ontstaan. Dat blijkt nog best een uitdaging.[11] Inmiddels zijn de Algemene Verordening Gegevensverwerking ('AVG'),[12] de Datagovernanceverordening,[13] de Verordening betreffende markten in cryptoactiva (MiCa),[14] de Digitaledienstenverordening ('DSA'),[15] de Digitalemarktenverordening ('DMA'),[16] de Digital Operational Resilience Act ('DORA')[17] en de Cyberbeveiligingsverordening[18] aangenomen of al in werking getreden. Andere wetgeving is momenteel nog in voorbereiding, zoals het voorstel voor de AI Act,[19] het voorstel voor een richtlijn betreffende de aanpassing van de regels inzake niet-contractuele civielrechtelijke aansprakelijkheid aan artificiële intelligentie[20] en de AI-productaansprakelijkheidsrichtlijn.[21] Ook de voorstellen voor de Dataverordening,[22] de cyberweerbaarheid[23] en de verbetering van de arbeidsvoorwaarden van platformwerk (richtlijnvoorstel platformwerk)[24] zijn in voorbereiding. Het veld is dus in beweging. De Europese Commissie heeft daarnaast een aantal andere instrumenten ontwikkeld om verantwoordelijke ontwikkeling en inzet van technologie te bevorderen zoals de Ethische richtsnoeren voor betrouwbare kunstmatige intelligentie[25] en

7 OpenAI (2023), *supra* noot 5, p. 2: "*Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.*" Inmiddels heeft de European Data Protection Board een taskforce opgericht om de zorgen ten aanzien van privacy te onderzoeken en samenwerking tussen de lidstaten en mogelijke handhavingsacties te coördineren; https://edpb.europa.eu/news/news/2023/edpb-resolves-dispute-transfers-meta-and-creates-task-force-chat-gpt_en en https://www.euractiv.com/section/artificial-intelligence/news/european-data-protection-authorities-launch-task-force-on-chatgpt/.

8 Zie voetnoten 12 t/m 26.

9 Deze discussies betreffen vragen zoals: wat is het gevolg als Europa vergaande regels introduceert om de ontwikkeling en het gebruik van kunstmatige intelligentie in goede banen te leiden, terwijl de Verenigde Staten en China deze regels niet, of althans niet in deze mate, opleggen, of wat is het gevolg van de enorme ijsberg aan duurzaamheidsrapportageverplichtingen die Europese ondernemingen opgelegd krijgen ten opzichte van buitenlandse concurrenten waarop dit niet van toepassing is?

10 Bepaalde regels zijn gericht op specifieke technologieën waarbij aansprakelijkheid veelal apart wordt geregeld. Andere regels zijn gericht op (specifieke soorten) gegevens, weer andere op veiligheid en weerbaarheid, en er zijn bijvoorbeeld regels die van toepassing zijn op het verlenen van bepaalde digitale diensten (waarbij het soort bedrijf een rol speelt, zoals grote online platforms) en de werkomstandigheden bij die bedrijven.

11 CEPS (2022), 'The AI Act and Emerging EU Digital Acquis: Overlaps, gaps and inconsistencies', september 2022-02; https://www.ceps.eu/wp-content/uploads/2022/09/CEPS-In-depth-analysis-2022-02_The-AI-Act-and-emerging-EU-digital-acquis.pdf.

12 Verordening (EU) 2016/679, *PbEU* L 119/1, 4 mei 2016.

13 Verordening (EU) 2022/868, *PbEU* L 152/1, 3 juni 2022. Deze verordening is van toepassing met ingang van 24 september 2023: zie artikel 38.

14 Het voorstel voor een verordening betreffende markten in cryptoactiva (MiCA) (COM(2020) 593 final). Inmiddels is de trialoog geëindigd nadat het EP zijn positie had aangenomen op 20 april 2023 waarna de Council de finale tekst heeft aangenomen op 16 mei 2023. De wetgeving zal binnen een jaar in werking treden.

15 Verordening (EU) 2022/2065, *PbEU* L 277/1, 27 oktober 2022.

16 Verordening (EU) 2022/1925, *PbEU* L 265/1, 12 oktober 2022.

17 Verordening (EU) 2022/2554, *PbEU* L 333/1, 27 december 2022.

18 Verordening (EU) 2019/881, *PbEU* L 151/15, 7 juni 2019. Zie ook de NIS 2-Richtlijn (Richtlijn (EU) 2022/2555, *PbEU* L 333/80, 27 december 2022).

19 Europese Commissie (2021), 'Voorstel voor een Verordening van het Europees Parlement en de Raad tot vaststelling van geharmoniseerde regels betreffende Artificiële Intelligentie (Wet op de Artificiële Intelligentie) en tot wijziging van bepaalde wetgevingshandelingen van de Unie', COM(2021)206 final, 21 april 2021 (hierna: 'Voorstel AI Act').

20 Europese Commissie (2022), 'Voorstel voor een Richtlijn van het Europees Parlement en de Raad betreffende de aanpassing van de regels inzake niet-contractuele civielrechtelijke aansprakelijkheid aan artificiële intelligentie (AI)', 2022/0303(COD), 28 september 2022.

21 Europese Commissie (2022), 'Voorstel voor een Richtlijn van het Europees Parlement en de Raad inzake aansprakelijkheid voor producten met gebreken', 2022/0302(COD), 28 september 2022.

22 Europese Commissie (2022), 'Voorstel voor een Verordening van het Europees Parlement en de Raad betreffende geharmoniseerde regels inzake eerlijke toegang tot en eerlijk gebruik van data (Dataverordening)', COM(2022) 68 final, 23 februari 2022 (hierna: 'Voorstel Dataverordening').

23 Europese Commissie (2022), 'Voorstel voor een Verordening van het Europees Parlement en de Raad betreffende horizontale cyberbeveiligingsvereisten voor producten met digitale elementen en tot wijziging van Verordening (EU) 2019/1020, COM/2022/454, final, 15 september 2022.

24 Europese Commissie (2021), 'Voorstel voor een Richtlijn van het Europees Parlement en de Raad betreffende de verbetering van de arbeidsvoorwaarden bij platformwerk', 2021/0414(COD), 9 december 2021 (hierna: 'Richtlijnvoorstel platformwerk').

25 Deskundigengroep op hoog niveau inzake kunstmatige intelligentie (2019), 'Ethische richtsnoeren voor betrouwbare kunstmatige intelligentie'; https://op.europa.eu/nl/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1.

de Europese verklaring over digitale rechten en beginselen voor het digitale decennium.[26] In dit artikel beperk ik mij bij de bespreking van digitaliseringswetgeving tot de AVG, het Voorstel Dataverordening, het Voorstel AI Act en ten slotte de Europese digitale rechten en beginselen. Ik bespreek deze hoofdzakelijk vanuit de optiek van ESG.

2.1 *Europese verklaring over digitale rechten en beginselen voor het digitale decennium*

In het beleidsprogramma '*Path to the Digital Decade*'[27] staat centraal dat het digitaliseringsproces en de technologische innovaties en -oplossingen sterk geworteld dienen te zijn in de Europese waarden, gericht op de mens. Hoewel de originele principes van Industrie 4.0[28] gericht waren op sociale rechtvaardigheid en duurzaamheid is de nadruk steeds meer komen te liggen op digitalisering en AI-gedreven technologieën die efficiëntie en flexibiliteit van productie verhogen.[29] Bij de verschuiving van industrie 4.0 naar industrie 5.0 beweegt de nadruk nu weer terug naar de originele principes waarbij drie kernelementen centraal staan: mensgerichtheid, duurzaamheid en bestendigheid.[30] Hierbij is de vraag niet zozeer wat wij kunnen doen met de technologie, maar wat de technologie voor ons kan doen. Centraal daarbij staat hoe de technologie zodanig kan worden ontwikkeld dat deze gericht is op de persoon die ermee dient te werken en ook onze privacy, autonomie en menselijke waardigheid respecteert. In dat kader hebben de Europese Commissie, het Europees Parlement en de Raad in december 2022 de Europese verklaring over digitale rechten en beginselen voor het digitale decennium ondertekend.[31] Dit initiatief wordt door de EU-burgers breed ondersteund, met name door de jongere generatie.[32] Bijna negen van de tien respondenten vinden het belangrijk dat EU-burgers worden beschermd tegen risicovol of onethisch gebruik van digitale technologieën zoals kunstmatige intelligentie.[33] Marktdeelnemers dienen hun sociale verantwoordelijkheid te nemen en zich veilig te gedragen.[34] Met deze verklaring committeren de EU en haar lidstaten zich binnen hun bevoegdheden tot de bevordering en uitvoering van de beginselen. De verklaring dient tevens als referentiepunt voor ondernemingen als zij nieuwe technologieën ontwikkelen en uitrollen.[35]

2.1.1 *Digitale beginselen*

In de verklaring zijn zes categorieën van beginselen opgenomen, elk in een afzonderlijk hoofdstuk. Dit zijn (1) mensen centraal; (2) solidariteit en inclusie; (3) keuzevrijheid; (4) deelname aan de digitale openbare ruimte; (5) veiligheid, beveiliging en empowerment; en (6) duurzaamheid. Het eerste beginsel houdt in dat in de Europese digitale transformatie mensen centraal staan. Technologie moet ten dienste staan en ten goede komen aan alle mensen die in de EU wonen en hen in staat stellen hun ambities na te streven, in alle veiligheid en met volledige inachtneming van hun grondrechten.[36] De categorie 'solidariteit en inclusie' bevat beginselen die erop gericht zijn dat het ontwerp, de ontwikkeling, de uitrol en het gebruik van technologische oplossingen de grondrechten eerbiedigen, de uitoefening ervan mogelijk maken en solidariteit en inclusie bevorderen.[37] Het gaat dan bijvoorbeeld om transparantie over de inzet van kunstmatige intelligentie en het menselijk toezicht bij gebruik van algoritmen in het kader van werkuitvoering.[38] De categorie 'keuzevrijheid' staat in het teken van AI-systemen. Mensen dienen geïnformeerde keuzes te maken over het gebruik van AI en beschermd te worden tegen risico's met betrekking tot gezondheid, veiligheid en grondrechten. Ook hier gaat het om transparantie en menselijk toezicht, welke worden aangevuld met het gebruik van passende data, het voorkomen van manipulatie en het toepassen van ethische en betrouwbare normen. De categorie 'veiligheid, beveiliging en empowerment' is gericht op een beschermde, veilige en beveiligde digitale omgeving, privacy en individuele controle over gegevens en bescherming en empowerment van kinderen en jongeren in de digitale omgeving. Het hoofdstuk over duurzaamheid beschrijft dat digitale diensten en producten zodanig dienen te worden ontworpen, geproduceerd, gebruikt, gerepareerd, gerecycleerd en verwijderd dat de negatieve gevolgen voor milieu en samenleving worden beperkt en voortijdige veroudering wordt voorkomen. In het kader daarvan dient iedereen toegang te hebben tot accurate, gemakkelijk te begrijpen informatie over de milieueffecten en het energieverbruik van digitale producten

26 Gemeenschappelijke Verklaringen Europees Parlement, Raad, Europese Commissie (2023), 'Europese verklaring over digitale rechten en beginselen voor het digitale decennium', (2023, C 23/01), *PbEU* L 23/1, 23 januari 2023 (hierna: Verklaring Europese digitale rechten en beginselen).

27 European Commission (2021), 'Proposal for a decision of the European Parliament and of the Council establishing the 2030 Policy Programme "Path to the Digital Decade"', COM(2021)574 final, 15 september 2021.

28 De vierde industriële revolutie werd aangedreven door de opkomst van slimme technologieën zoals big data en AI-analyses, horizontale en verticale integratie, cloud computing, virtuele realiteit, industrieel internet der dingen, 3D-printing, autonome robots, digitale twins en cyberveiligheid.

29 European Commission (2021), 'Industry 5.0: Towards a sustainable, human centric and resilient European industry', *R&I Paper Series Policy Brief*; https://op.europa.eu/en/publication-detail/-/publication/468a892a-5097-11eb-b59f-01aa75ed71a1/.

30 Industrie 5.0 betreft niet zozeer nieuwe technologische innovaties, maar is gericht op de interactie tussen computer en mens waarbij menselijke creativiteit en welzijn centraal staan en automatisering wordt gepersonaliseerd.

31 Verklaring Europese digitale rechten en beginselen, *supra* noot 26.

32 European Commission (2021), 'Special Eurobarometer 518: Digital Rights and Principles', p. 8, 26 en 27. Zie voor de verschillen tussen generaties: https://europa.eu/eurobarometer/surveys/detail/2270, p. 27 (hierna: 'Eurobarometer 518').

33 Eurobarometer 518, p. 34.

34 Hoofdstuk I, II en III Verklaring Europese digitale rechten en beginselen.

35 Overweging 8 Verklaring Europese digitale rechten en beginselen.

36 Beginsel 1 Verklaring Europese digitale rechten en beginselen.

37 Dit zijn beginselen op het gebied van solidariteit en inclusie, connectiviteit, digitaal onderwijs en digitale opleiding en vaardigheden, rechtvaardige en billijke arbeidsomstandigheden en -voorwaarden en digitale overheidsdiensten.

38 De rol van vakbonden en werkgeversorganisaties wordt benadrukt.

en diensten en hun repareerbaarheid en levensduur, zodat zij verantwoorde keuzes kunnen maken.[39]

Dit is waar de twin transitions, waar technologie en ESG, in essentie over gaan. Deze verklaring is echter van declaratoire aard. In de overwegingen valt te lezen dat de beginselen derhalve als zodanig geen invloed op de inhoud van rechtsregels of de toepassing daarvan hebben.[40] Hoewel die invloed er mijns inziens wel kan zijn vanwege het mogelijke effect van deze beginselen op de invulling van open normen, zit de crux er dus in of, en zo ja op welke wijze, de in de verklaring beschreven rechten en beginselen zijn ingebed in de digitaliserings- en duurzaamheidswetgeving. Ik start met de bespreking van de digitale wetgeving(svoorstellen) die betrekking hebben op data. In het kader van dataverwerking zijn onder andere de AVG, de Verordening inzake een kader voor het vrije verkeer van niet-persoonsgebonden gegevens,[41] de Dataverordening en het Voorstel AI Act van belang. Ik richt me hierbij omwille van de beschikbare ruimte met name op dataverwerking in algemene zin en meer specifiek op datadeling en minder op de datakwaliteit en voornoemde transparantie in het kader van algoritmische besluitvorming. Daarna ga ik dieper in op het Voorstel AI Act.

2.2 *Wetgeving gericht op data*

Welk raamwerk van regels van toepassing is op de data die worden verzameld en gebruikt door ondernemingen hangt voor een groot deel af van de classificatie van de data (anoniem, pseudo-anoniem, geaggregeerd/statistisch en (bijzonder/gevoelig[42]) persoonlijk).[43] Zo zal op anonieme data een andere set met regels van toepassing zijn dan op (bijzondere) persoonsgegevens.[44] En die regels kunnen 180 graden de andere kant uitwijzen, omdat bij anonieme data de regels gericht zijn op het bevorderen van datadeling (het voorkomen van beperkingen met betrekking tot het vrije verkeer van de data)[45] en daarmee de ondersteuning van innovatie, terwijl bij persoonsgegevens de regels juist met name gaan om bescherming (van de fundamentele rechten en vrijheden)[46] van het datasubject in het kader van de dataverwerking en het vrije verkeer.[47] Het Voorstel AI Act heeft daarentegen betrekking op zowel persoonsgebonden als op niet-persoonsgebonden gegevens en bevat data- en datagovernanceregels die specifiek betrekking op bescherming tegen de risico's van AI-systemen en -modellen. In het wetgevingstraject tot nu toe is door verschillende partijen benadrukt dat uit de AI Act helder moet blijken dat het bestaande gegevensbeschermingswetgevingskader waaronder de AVG van toepassing blijft en de AI Act daarmee niet conflicteert.[48]

39 Het Europees Parlement, de Raad en de Europese Commissie hebben zich in dit kader ertoe verbonden om a) de ontwikkeling en het gebruik van duurzame digitale technologieën met minimale ecologische en sociale gevolgen te bevorderen, b) duurzame consumentenkeuzes en bedrijfsmodellen te stimuleren, en duurzaam en verantwoord ondernemingsgedrag in de mondiale waardeketens van digitale producten en diensten te bevorderen, onder meer ter bestrijding van dwangarbeid, c) de ontwikkeling, de uitrol en het actieve gebruik van innovatieve digitale technologieën met een positief effect op milieu en klimaat te bevorderen, teneinde de groene transitie te versnellen en d) duurzaamheidsnormen en -labels voor digitale producten en diensten te bevorderen.

40 Overweging 7 Verklaring Europese digitale rechten en beginselen.

41 De Verordening inzake een kader voor het vrije verkeer van niet-persoonsgebonden gegevens in de Europese Unie (Verordening (EU) 2018/1807, *PbEU* L 303/59, 28 november 2018) (hierna: 'Verordening inzake een kader voor het vrije verkeer van niet-persoonsgebonden gegevens').

42 Bijzondere gegevens worden gedefinieerd als persoonsgegevens waaruit ras of etnische afkomst, politieke opvattingen, religieuze of levensbeschouwelijke overtuigingen of lidmaatschap van een vakbond blijken, en de verwerking van genetische gegevens, biometrische gegevens met het oog op de unieke identificatie van een natuurlijke persoon, gegevens over gezondheid of gegevens met betrekking tot het seksuele gedrag of gerichtheid van een natuurlijk persoon; artikel 9 AVG.

43 De Verordening inzake een kader voor het vrije verkeer van niet-persoonsgebonden gegevens respectievelijk de AVG. Zie voor een mooi overzicht: B. van der Sloot, S. van Schenkel & C.A. Fontanillo López (2022), 'The influence of (technical) developments on the concept of personal data in relation to the GDPR', Tilburg Institute for Law, Technology and Society, Tilburg University in opdracht van het WODC, p. 14; *Kamerstukken II* 2022/2023, 32 761, bijlage. Van der Sloot et al. plaatsen kanttekeningen bij deze indeling, bijvoorbeeld ten aanzien van anonimiteit waarover de auteurs stellen dat er eerder sprake is van een spectrum van anonimiteit in plaats van een categorie 'anoniem' en een categorie 'pseudo-anoniem' of bijvoorbeeld omdat de gevoeligheid niet zozeer afhangt van de data die worden verwerkt, maar van de technologie die wordt gebruikt voor de verwerking: zie p. 21 respectievelijk 24.

44 Hierbij is van belang dat de huidige status van data niet de classificatie bepaalt. Dat wil zeggen dat de mogelijk toekomstige status van de data bepalend is. Het kan zijn dat door de ontwikkeling van technologieën bepaalde data op termijn identificeerbaar worden. Dit hangt af van de omstandigheden van het geval.

45 Artikel 1 Verordening inzake een kader voor het vrije verkeer van niet-persoonsgebonden gegevens.

46 Zie onder andere overweging 2, 4, 10, 51 en 71 AVG.

47 Artikel 1 AVG.

48 EDPB (2021), 'EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act)', 18 juni 2021, p. 8 en 16; https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_en. Zie ook de nieuwe overweging 2b van de door het Europees Parlement in het kader van de AI Act aangenomen tekst: European Parliament (2023), 'Artificial Intelligence Act: Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts', (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), P9_TA(2023)0236; https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html (hierna: 'EP-tekst').

Uit een recent WODC-onderzoek blijkt overigens dat het de vraag is hoe houdbaar het voornoemde dataclassificatiesysteem nog is. Door de stand van de huidige techniek kan een 'datum' of dataset van seconde tot seconde van categorie wisselen.[49] In de Verordening inzake een kader voor het vrije verkeer van niet-persoonsgebonden gegevens in de Europese Unie, staat in de overwegingen dat het snelgroeiende internet der dingen, kunstmatige intelligentie en machine-learning belangrijke bronnen van niet-persoonsgebonden gegevens zijn.[50] Daarbij vermeldt de overweging echter dat indien technologische ontwikkelingen het mogelijk maken om geanonimiseerde gegevenssets om te vormen tot persoonsgegevens, dergelijke gegevens behandeld moeten worden als persoonsgegevens en dus de AVG van toepassing zal zijn. En dit is door de ontwikkelingen op het gebied van kunstmatige intelligentie steeds vaker het geval.[51] Ook pseudo-geanonimiseerde gegevens vallen onder de reikwijdte van de AVG.

2.2.1 *AVG-beginselen*

Wanneer dit in het perspectief van ESG wordt geplaatst, zorgt het feit dat de AVG in het kader van dataverwerking door AI-systemen vaker van toepassing zal zijn (nu gegevens sneller classificeren als persoonsgegevens), aan de ene kant voor een positief effect op fundamentele rechten en vrijheden uit de 'S'-pijler vanwege de AVG-bescherming. Zonder te willen verdwijnen in de krochten van de AVG, bespreek ik een paar beginselen om deze bescherming te duiden. Allereerst is van belang dat verwerking van persoonsgegevens alleen mogelijk is als deze rechtmatig is. Er dient derhalve sprake te zijn van een grondslag.[52] Als grondslag voor de dataverwerking in het kader van big-data-analyse door ondernemingen is veelal alleen het gerechtvaardigd belang de aangewezen optie.[53] En dan dient er een afweging plaats te vinden tussen de belangen van de verwerkingsverantwoordelijke (of de derde) en de betrokkene(n).[54] Bij deze afweging wordt meegewogen welke maatregelen de verwerkingsverantwoordelijke heeft genomen om ongewenste gevolgen voor de betrokkenen te voorkomen of te beperken.[55] Ook is het van belang rekening te houden met het verschil tussen enerzijds verwerking in het kader van de training van het model en anderzijds het gebruik van de persoonsgegevens als de input van het model zelf. In het eerste geval zal de verwerking, gecombineerd met maatregelen zoals pseudo-anonimisering en daarnaast anonimisering na de training, bij de belangenafweging sneller in het voordeel van de verwerkingsverantwoordelijke werken.[56] Bij het tweede geval is dat nog maar de vraag. De Autoriteit Persoonsgegevens hanteert overigens vooralsnog het uitgangspunt dat een puur commercieel belang of winstmaximalisatie niet als gerechtvaardigd belang kan dienen.[57] Dat beperkt het gebruik van deze grondslag voor verwerking in het kader van kunstmatige intelligentie aanzienlijk.

Verder bevat de AVG het doelbindingsbeginsel dat inhoudt dat persoonsgegevens alleen voor welbepaalde, uitdrukkelijk omschreven en gerechtvaardigde doelen mogen worden verzameld (doelspecificatie) en niet op een met die doeleinden onverenigbare wijze mogen worden verwerkt.[58] Met andere woorden; het doel moet specifiek, expliciet en rechtmatig zijn en verdere verwerking mag daar niet incompatibel mee zijn. In dat kader dient er een compatibiliteitstoets gedaan te worden waarbij onder andere gekeken wordt naar de gevolgen voor de betrokkene en de genomen maatregelen.[59] Uitzondering hierop bestaat als er toestemming is van de betrokkene of de ver-

49 Daarnaast zorgt het ruime toepassingsbereik van de AVG er nu al voor dat veel gegevens worden aangemerkt als persoonsgegevens. Zie: P. Wolters (2023), 'De invloed van de Data Act op de verschuivende balans tussen gegevensbescherming en het vrije verkeer van data', *Ars Aequi* januari 2023, p. 25-34.

50 Overweging 9 Verordening inzake een kader voor het vrije verkeer van niet-persoonsgebonden gegevens. De overweging geeft een aantal voorbeelden van niet-persoonsgebonden gegevens: geaggregeerde en geanonimiseerde gegevenssets die worden gebruikt voor big-data-analyses, gegevens over precisielandbouw waarmee het gebruik van pesticiden en water kan worden gemonitord en geoptimaliseerd, en gegevens over de onderhoudsbehoeften van industriële machines.

51 European Parliament (2020), 'The impact of the General Data Protection Regulation (GDPR) on artificial intelligence', Study by the Panel for Future of Science and Technology, *PE* 641.530, juni 2020, p. 36-37; https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)641530.

52 Dit kunnen onder andere zijn toestemming van de betrokkene, uitvoering van een overeenkomst, een wettelijke verplichting of taak van algemeen belang of bijvoorbeeld onder omstandigheden het gerechtvaardigd belang van de verwerkingsverantwoordelijke of een derde. Artikel 6 lid 1 AVG. Zie voor een uitleg van het toetsingskader: B.W. Schermer & J. Toornstra (2023), 'Handleiding Algemene verordening gegevensbescherming en Uitvoeringswet Algemene verordening gegevensbescherming', in opdracht van het Ministerie van Justitie en Veiligheid, p. 37; https://www.rijksoverheid.nl/documenten/rapporten/2018/01/22/handleiding-algemene-verordening-gegevensbescherming (hierna: *Handboek AVG/UAVG*).

53 E.W.S. Peperkamp & M.H.A. Voorboom (2020), 'Autoriteit Persoonsgegevens beperkt mogelijkheden data analytics met haar normuitleg grondslag 'gerechtvaardigd belang'', *Privacy & Informatie* 2020, afl. 4, p. 171-175.

54 De toets die moet worden uitgevoerd houdt het volgende in: (1) Het belang dat wordt nagestreefd moet een gerechtvaardigd belang zijn. (2) De verwerking van de persoonsgegevens is noodzakelijk voor de behartiging van dat gerechtvaardigd belang. Hierbij moet tevens worden getoetst aan het proportionaliteits- en het subsidiariteitsbeginsel: is de inbreuk voor de betrokkenen in verhouding tot het met de verwerking te dienen doel en kan het doel ook voor de betrokkenen op een minder nadelige wijze worden bereikt. En (3) De fundamentele rechten en vrijheden van de betrokkenen prevaleren niet. HvJ EU 4 mei 2017, nr. C-13/16, ECLI:EU:C:2017:336, r.o. 28 (*Rigas*) en HvJ EU 29 juli 2019, nr. C-40/17, ECLI:EU:C:2019:629, r.o. 95 (*Fashion-ID*).

55 https://www.autoriteitpersoonsgegevens.nl/themas/basis-avg/avg-algemeen/grondslagen-avg-uitgelegd#grondslag-gerechtvaardigd-belang.

56 European Parliament (2020), *supra* noot 51, p. 50.

57 Over deze uitleg is nogal wat te doen omdat deze niet in lijn zou zijn met de richtsnoeren van de WP29 en de EDPB en de rechtspraak van het HvJ EU. De Europese Commissie heeft de AP gevraagd haar normuitleg te herzien. Er zijn ook prejudiciële vragen gesteld door de rechtbank Amsterdam in de KNLTB-zaak: Rb. Amsterdam 22 september 2022, ECLI:NL:RBAMS:2022:5565, r.o. 6.

58 Artikel 5 lid 1 onderdeel b AVG. Zie over deze dubbele negatie: E.M.L. Moerel & J.E.J. Prins (2016), 'Privacy voor de homo digitalis', in: E.M.L. Moerel et al. (2016), *Homo digitalis (Handelingen Nederlandse Juristen-Vereniging 146e jaargang/2016-I)*, Deventer: Wolters Kluwer 2016, p. 12.

59 Artikel 6 lid 4 AVG.

werking op regelgeving berust die in een democratische samenleving een noodzakelijke en evenredige maatregel vormt.[60] Voor AI-toepassingen gebaseerd op een grote hoeveelheid data afkomstig uit verschillende bronnen waarbij het vragen om toestemming onbegonnen werk is en de voornoemde grondslag afwezig is, zal het doelbindingsbeginsel verwerking lastig of zelfs onmogelijk maken. Veel AVG-beginselen[61] en de daaruit voortvloeiende verplichtingen (en de mogelijkheid om een beroep te doen op een uitzondering daarop) vergen een belangenafweging gebaseerd op een beoordeling van risico's voor de betrokkenen en de daarop gerichte mitigatiemaatregelen. Daarnaast bevat de AVG ook nog een apart impactassessment voor specifieke gevallen.

2.2.2 *Data protection impact assessment*

Op de verwerkingsverantwoordelijke kan namelijk de verplichting rusten om een gegevensbeschermingseffectbeoordeling (DPIA) uit te voeren.[62] Deze verplichting bestaat als de verwerking van persoonsgegevens waarschijnlijk een hoog risico[63] inhoudt voor de rechten en vrijheden van natuurlijke personen, in het bijzonder als er nieuwe technologieën worden gebruikt bij de verwerking.[64] Deze beoordeling houdt onder andere in de beoordeling van de risico's voor de rechten en vrijheden van natuurlijke personen en de beoogde maatregelen om de risico's aan te pakken.[65] Deze rechten en vrijheden hebben voornamelijk betrekking op het recht op gegevensbescherming en privacy, maar kunnen ook andere grondrechten betreffen zoals vrijheid van meningsuiting, vrijheid van gedachte, discriminatieverbod etc.[66]

2.2.3 *Data, AVG en ESG*

Uit het bovenstaande blijkt dat de AVG dus een belangrijke bescherming biedt in het kader van de 'S'-pijler. Er kan ook een conflict binnen de 'S'-pijler ontstaan omdat het recht op gegevensbescherming met bijbehorende regels voor het gebruik van data ervoor kan zorgen dat de bescherming van andere mensenrechten en vrijheden wordt bemoeilijkt, omdat voor die bescherming nu juist een verwerking van persoonsgegevens vereist is. Een belangenafweging dient dan te worden uitgevoerd conform het evenredigheidsbeginsel.[67] Daarnaast zorgt de uitbreiding van de capaciteit en functionaliteit van AI-systemen voor een toenemende complexiteit als het gaat om het bevorderen van datagedreven oplossingen die bijvoorbeeld kunnen bijdragen aan klimaat en milieu. Dit komt doordat de te gebruiken data sneller in het bakje van 'bescherming' kunnen vallen vanwege de classificatie persoonsgegevens dan in het bakje 'open data'.[68] Er ontstaat dan als het ware een conflict tussen de 'E'-pijler en de 'S'-pijler. Enerzijds gaat het om de bevordering van een goed klimaat en milieu, anderzijds gaat het om het recht op gegevensbescherming van individuen en mogelijk in het verlengde daarvan het recht op privacy.[69] Een onderneming die AI-toepassingen in het kader van duurzaamheid wil ontwikkelen waarbij persoonsgegevens verwerkt dienen te worden, stuit dus op een aantal belemmeringen of hobbels zo je wilt. Het gaat het bestek van dit artikel te buiten om de complexiteit van deze spanning binnen de kaders van de AVG en de daarmee samenhangende Nederlandse uitvoeringsregels geheel te bespreken. Het lijkt alsof de Europe-

60 Artikel 6 lid 4 jo. artikel 23 AVG.

61 Andere beginselen zijn juistheid, vertrouwelijkheid en integriteit, behoorlijkheid en transparantie (in aanvulling op rechtmatigheid) en het beginsel van minimale gegevensverwerking. Deze laatste houdt in dat er niet meer gegevens worden gebruikt dan noodzakelijk is voor het bereiken van het doel. Minimalisatie van gegevensverwerking betreft de hoeveelheid verzamelde persoonsgegevens, de mate waarin zij worden verwerkt, de termijn waarvoor zij worden opgeslagen en de toegankelijkheid daarvan; artikel 5 onderdeel c AVG. Om gegevensbescherming te waarborgen dient de onderneming deze te verankeren in het ontwerp (*by design*) en in de standaardinstellingen (*by default*). Een "standaardinstelling" zoals doorgaans gedefinieerd in informatica (computerwetenschappen), verwijst naar de reeds bestaande of vooraf geselecteerde waarde van een configureerbare instelling die is toegekend aan een applicatie, computerprogramma of apparaat. Dergelijke instellingen worden ook "voorinstellingen" of "fabrieksinstellingen" genoemd, met name bij elektronische apparaten: zie EDPB (2020), 'Richtsnoeren 4/2019 inzake artikel 25: Gegevensbescherming door ontwerp en door standaardinstellingen', versie 2.0, 20 oktober 2020, paragraaf 2.2.1 en 2.2.2; https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en. De maatregelen die genomen moeten worden, dienen passend te zijn met het oog op de betreffende verwerking en de context daarvan. Hiervoor dient de onderneming de risico's die de verwerking voor de betrokkenen kan veroorzaken te identificeren en te vertalen in de maatregelen. *Handboek AVG/UAVG*, p. 59-61.

62 Artikel 35 lid 1 AVG. DPIA staat voor data protection impact assessment.

63 Om te bepalen of er mogelijk sprake is van een hoog risico hanteren de toezichthouders de onderstaande vuistregel dat sprake is van een hoog risico wanneer de verwerking aan twee of meer van de onderstaande negen criteria voldoet: (1) evaluatie van personen of scoretoekenning; (2) geautomatiseerde besluitvorming met rechtsgevolg of vergelijkbaar wezenlijk gevolg; (3) stelselmatige monitoring; (4) gevoelige gegevens of gegevens van zeer persoonlijke aard; (5) op grote schaal verwerkte gegevens; (6) matching of samenvoeging van datasets; (7) gegevens met betrekking tot kwetsbare betrokkenen; (8) innovatieve toepassing van nieuwe technologische of organisatorische oplossing; (9) blokkering van een recht, dienst of contract; https://www.autoriteitpersoonsgegevens.nl/themas/basis-avg/praktisch-avg/data-protection-impact-assessment-dpia. De Autoriteit Persoonsgegevens heeft ook een niet-uitputtende lijst opgesteld met verwerkingen waarvoor altijd een DPIA dient te worden uitgevoerd; https://www.autoriteitpersoonsgegevens.nl/documenten/besluit-lijst-verplichte-dpia.

64 Artikel 35 AVG bevat drie situaties waarin in ieder geval een gegevensbeschermingseffectbeoordeling moet worden uitgevoerd. De Groep Gegevensbescherming artikel 29 heeft in haar richtsnoeren voor DPIA's, die door het EDPB zijn bekrachtigd (Endorsement 1/2018), de criteria verduidelijkt aan de hand waarvan kan worden bepaald of voor een verwerking een DPIA moet worden uitgevoerd: Groep Gegevensbescherming artikel 29 (2017), 'Richtsnoeren voor gegevensbeschermingseffectbeoordelingen en bepaling of een verwerking "waarschijnlijk een hoog risico inhoudt" in de zin van Verordening 2016/679', WP 248 rev. 01, 4 april 2017.

65 Artikel 35 lid 7 onderdeel c en d AVG. In de richtsnoeren staat hierover: '*Er moet worden benadrukt dat om de risico's voor de rechten en vrijheden van natuurlijke personen te beheren, de risico's moeten worden geïdentificeerd, geanalyseerd, ingeschat, geëvalueerd, aangepakt (bijvoorbeeld afgezwakt) en regelmatig moeten worden herbeoordeeld.*'

66 Groep Gegevensbescherming artikel 29 (2017), *supra* noot 64, p. 7.

67 Overweging 4 AVG.

68 Vergelijk B. van der Sloot, S. van Schenkel & C.A. Fontanillo López (2022), *supra* noot 43, p. 26 en 27. Zie ook: P. Wolters (2023), *supra* noot 49.

69 B. van der Sloot (2021), 'Wij zijn twee vriendjes, jij en ik: de innige tango van de twee Europese hoven op het gebied van privacy en gegevensbescherming', *SEW* 2021, nr. 12, p. 582-594.

se Commissie in het Voorstel AI Act heeft getracht de hobbel van het doelbindingsbeginsel onder omstandigheden weg te willen nemen voor AI-systemen die gericht zijn op de klimaat- en milieuproblematiek.

2.2.4 *Gegevensverwerking voor een ander doel en andere dataregels in het Voorstel AI Act*

In dit voorstel heeft de Europese Commissie namelijk in artikel 54 lid 1 AVG opgenomen dat persoonsgegevens die zijn verzameld voor andere doeleinden, rechtmatig gebruikt mogen worden voor het doeleinde om bepaalde innovatieve AI-systemen te ontwikkelen en te testen in het kader van een zwaarwegend algemeen belang, zoals een hoog niveau van bescherming en verbetering van de kwaliteit van het milieu.[70] Het voorstel lijkt derhalve de grondslag voor deze verwerking te verschaffen voor de AI-testomgeving. Dit betreft de *regulatory sandbox* voor AI-toepassingen.[71] Deze testomgeving biedt mogelijkheden voor ondernemingen om AI-systemen te ontwikkelen die kunnen bijdragen aan het oplossen of verminderen van de klimaat- en milieuproblematiek. Hoewel nog niet helemaal duidelijk is welke applicaties in deze categorie vallen, zou dit mijns inziens bijvoorbeeld een softwareprogramma kunnen betreffen dat aanbevelingen doet over de plaatsing van windturbines of zonnepanelen gebaseerd op de impact daarvan op de leefomgeving en relevante betrokkenen. Om die impact te berekenen kan gebruik worden gemaakt van verschillende data afkomstig uit diverse bronnen. De idee is dat de verschillende EU-lidstaten een eigen of een gezamenlijke testomgeving opzetten.[72] Zij zijn daar in het voorstel van de Europese Commissie echter niet toe verplicht. In de tekst aangenomen door het Europees Parlement ('EP-tekst') daarentegen wel.[73] Inmiddels is het traject voor de eerste AI-testomgeving in juni 2022 in Spanje van start gegaan.[74] Er zit echter wel een grote 'maar' aan deze testomgevingen vast. Het Voorstel AI Act bevat een hele reeks aan voorwaarden waaraan de testomgeving moet voldoen.[75] Hierbij is niet geheel duidelijk hoe deze regeling zich verhoudt tot de AVG.[76] De EDPB heeft als feedback gegeven dat het niet duidelijk is onder welke omstandigheden en met gebruikmaking van welke criteria de belangenafweging in het kader van de verwerking persoonsgegevens voor een ander doel plaatsvindt en of deze AI-systemen alleen in de AI-testruimte gebruikt zullen worden (dit komt mijns inziens neer op de vraag of de grondslag ook geldt voor de verwerking door het AI-systeem nadat deze de testruimte verlaat).[77] In de compromistekst van de Raad ten aanzien van de AI Act is een nieuwe overweging opgenomen over de verhouding van dit artikel en de AVG waarbij duidelijk naar voren komt dat de AVG leidend is, maar het is onzeker wat dit betekent voor artikel 54.[78] De uiteindelijke tekst van de AI Act zou hier helderheid in moeten brengen. Het Voorstel AI Act bevat daarnaast verschillende andere regels op het gebied van data en datagovernance. In de EP-tekst is een algemeen beginsel opgenomen dat van toepassing is op alle AI-systemen en -modellen dat luidt: '*privacy and data governance' means that AI systems shall be developed and used in compliance with existing privacy and data protection rules, while processing data that meets high standards in terms of quality and integrity.*' Omdat de toepasselijkheid van de regels die in acht moeten worden genomen om compliant te zijn met dit beginsel afhankelijk zijn van de kwalificatie van het betreffende AI-systeem of -model, en dit systeem enige uitleg vergt, bespreek ik deze dataregels bij de uitgebreidere uitleg van het Voorstel AI Act in paragraaf 2.3.5.

2.2.5 *Datadeling in de Data Act*

In Deel I van dit drieluik besprak ik hoe de EC door datadelingsinitiatieven de Greendeal-doelstellingen probeert te realiseren. Met het voorstel voor de Dataverordening probeert de EC een kader te scheppen waarmee zowel het delen van persoonsgegevens als niet-persoonsgegevens wordt ondersteund. Het verschaft in beginsel geen nieuwe

70 Dit is de Nederlandse vertaling van *environment* in het Voorstel AI Act. In de Compromistekst van de Raad zijn deze gronden uitgebreid en specifieker gemaakt en de verwijzing naar 'hoge' bescherming is verwijderd. In deze compromistekst staat in artikel 54 lid 1 onderdeel a AVG: '*(iii) protection and improvement of the quality of the environment, including green transition, climate change mitigation and adaptation; (iv) energy sustainability, transport and mobility*'. Europese Raad (2022), 'Compromis text AI Act', 25 november 2022, aangenomen op 6 december 2022, 14954/22; https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/nl/pdf (hierna: 'Compromistekst van de Raad').

71 Een regulatory sandbox voor AI is in het Voorstel AI Act beschreven als een testomgeving voor regelgeving voorzien in een gecontroleerde omgeving ter vergemakkelijking van het volgens een specifiek plan ontwikkelen, testen en valideren van innovatieve AI-systemen voor een beperkte duur voordat zij in de handel worden gebracht of in bedrijf worden gesteld; artikel 53(1) Voorstel AI Act.

72 De nationale testomgevingen dienen wel te voldoen aan algemene beginselen die in uitvoeringshandelingen worden vastgelegd. Deze betreffen de selectie van de deelnemers, de voorwaarden om deel te nemen aan de testomgeving en procedurele regels, artikel 53(6) Voorstel AI Act.

73 Artikel 53(1) EP-tekst.

74 https://portal.mineco.gob.es/es-es/comunicacion/Paginas/20220627-PR_AI_Sandbox.aspx.

75 Artikel 54(1) Voorstel AI Act: dit betreft onder andere het optuigen van monitoringsmechanismen om vast te stellen of zich tijdens de experimenten in de testomgeving hoge risico's voor de grondrechten van de betrokkenen kunnen voordoen evenals responsmechanismen om die risico's onmiddellijk te beperken en indien nodig de verwerking stop te zetten, het plaatsen van persoonsgegevens in een functioneel gescheiden, geïsoleerde en beschermde omgeving voor dataverwerking, het opstellen en bewaren van een volledige en gedetailleerde beschrijving van het proces en de onderbouwing van het trainen, het testen en het valideren van het AI-systeem en het publiceren van een korte samenvatting van het in de testomgeving ontwikkelde AI-project en de doelstellingen en verwachte resultaten ervan op de website van de bevoegde autoriteiten.

76 Vormt artikel 54 een afwijking van de regel dat er een compatibiliteitstest gedaan dient te worden en hoe verhoudt deze zich dan tot artikel 23 AVG? In artikel 54(2) Voorstel AI Act staat het volgende: 'Lid 1 laat wetgeving van de Unie of de lidstaten onverlet, uitgezonderd verwerking voor andere doeleinden dan die uitdrukkelijk vermeld in die wetgeving.' De voorwaarden die aan de testomgeving worden gesteld, komen niet geheel overeen met die van artikel 23 AVG.

77 EDPB (2021), *supra* noot 48, p. 18-19.

78 Overweging 72-a Compromistekst van de Raad. De AI Act is de juridische grond om persoonsgegevens te verwerken die voor andere gronden zijn verzameld, in lijn met artikel 6 lid 4 en artikel 9 lid 2 onderdeel g AVG.

rechten of verplichtingen tot datadeling en de regels van de AVG blijven nog steeds van kracht. Uitzondering hierop zijn data die worden verzameld of gegenereerd door producten die behoren tot het internet der dingen. In dat kader bevat deze voorgestelde wetgeving nieuwe rechten voor de gebruiker van dit soort producten. Deze gebruiker heeft recht op (rechtstreekse of continue en in real-time) toegang tot de data die door het product of de daaraan gerelateerde dienst worden gegenereerd, en de ontwerper, fabrikant en dienstverlener van de producten en gerelateerde diensten dienen met deze toegang rekening te houden bij het ontwerp en vervaardiging van het product en de verlening van de dienst.[79] Dit kan zeker leiden tot gunstige 'E'-effecten. Denk aan een boer die zelf veel gerichter feedback krijgt van de apparaten en werktuigen die worden gebruikt, zodat er bijvoorbeeld veel minder water benodigd is. De boer is voor deze informatie door deze voorgestelde regel niet meer afhankelijk van de producent van de apparaten of de ontwerper van de software.[80] Ook zorgt de voorgestelde verordening ervoor dat het bijvoorbeeld makkelijker wordt om data te verplaatsen van de ene aanbieder van verwerkingsdiensten naar een andere. Dus bijvoorbeeld van de ene cloud-provider naar een andere cloud-provider.[81] De voorgestelde wetgeving bevat daarnaast regels voor de dataruimten die ik in Deel I heb besproken. De voorgestelde Dataverordening stelt eisen aan de exploitanten van dataruimten om de data-uitwisseling te vergemakkelijken.[82] Om de interoperabiliteit[83] mogelijk te maken dienen zij informatie te verschaffen over datakwaliteit, -verzamelingsmethoden, -structuren, -formaten, vocabularia, classificatiethema's en taxonomieën etc.[84] Deze eisen kunnen door de EC nader worden gespecificeerd en de EC kan geharmoniseerde normen in de vorm van standaarden laten opstellen, zelf uitvoeringshandelingen vaststellen of richtsnoeren aannemen met interoperabiliteitsspecificaties zoals architectuurmodellen en technische normen.[85] Dit is een belangrijke stap in het creëren van de randvoorwaarden voor een werkbare datadeling, zodat de Common European Green Deal Data Space daadwerkelijk de beoogde doelstellingen kan realiseren.

2.3 *Wetgeving gericht op kunstmatige intelligentie: het Voorstel AI Act*

2.3.1 *Bescherming in het Voorstel AI Act*

Het voorstel voor de AI Act is wetgeving die gericht is op regulering van AI-toepassingen.[86] Inmiddels is de trialoog op 14 juni van dit jaar van start gegaan. De voorgestelde wetgeving bevat regels die eisen stellen aan de processen die het ontwerp, de ontwikkeling en het gebruik van bepaalde AI-toepassingen omgeven en verplichten tot het ter beschikking stellen van bepaalde informatie. De doelstelling van het Voorstel AI Act is om naast het ondersteunen van innovatie en het verbeteren van de interne markt, bescherming te bieden tegen bepaalde type risico's: risico's voor grondrechten, gezondheid en die gerelateerd aan veiligheid.[87] Voordat ik de structuur van de regels bespreek, is het van belang de reikwijdte van deze bescherming nauwkeuriger te duiden, aangezien deze bescherming zo op het eerste gezicht alleen betrekking lijkt te hebben op de 'S'-pijler en niet zozeer de 'E'-pijler. De 'E'-impact van AI is echter een niet te onderschatten 'risico',[88] zoals ik ook in Deel I heb beschreven.[89] De vraag rijst derhalve hoe de bescherming van klimaat en milieu zich

79 Artikel 3 en 4 Voorstel Dataverordening. Hierbij is het in het kader van persoonsgegevens wel de vraag of de gebruiker ook de betrokkene is op grond van de AVG. Dat hoeft namelijk niet het geval te zijn. De toegang tot de data is in dat geval beperkt tot situaties waarin er sprake is van een geldige rechtsgrondslag uit hoofde van artikel 6 lid 1 AVG en, in voorkomend geval, indien aan de voorwaarden van artikel 9 AVG is voldaan; artikel 4(5) Voorstel Dataverordening.

80 De gebruiker dient ook geïnformeerd te worden op welke wijze de fabrikant of de dienstverlener van plan is de data te gebruiken of een derde partij toe te staan de data te gebruiken, en hier zijn beperkingen aan gesteld.

81 Artikel 23 t/m 26 Voorstel Dataverordening.

82 Artikel 28 Voorstel Dataverordening.

83 In het Voorstel Dataverordening is de interoperabiliteit in artikel 2(19) als volgt gedefinieerd: '*"interoperabiliteit": het vermogen van twee of meer dataruimten of communicatienetwerken, -systemen, -producten, -toepassingen of -componenten om data uit te wisselen en te gebruiken teneinde hun functies te vervullen*'.

84 Artikel 28(1) Voorstel Dataverordening.

85 Artikel 28(2) en (3) Voorstel Dataverordening. Het Voorstel Dataverordening bevat de eisen waaraan deze specificaties en standaarden dienen te voldoen; artikel 29(1) Voorstel Dataverordening.

86 Het Voorstel AI Act gebruikt hiervoor de term AI-systeem. Over de definitie van AI-systemen is veel te doen geweest. De definitie in het Voorstel AI Act luidt: '*'artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with*'. In de Nederlandse vertaling: '*"artificiële-intelligentiesysteem" (AI-systeem): software die is ontwikkeld aan de hand van een of meer van de technieken en benaderingen die zijn opgenomen in de lijst van bijlage I en die voor een bepaalde reeks door mensen gedefinieerde doelstellingen output kan genereren, zoals inhoud, voorspellingen, aanbevelingen of beslissingen die van invloed zijn op de omgeving waarmee wordt geïnterageerd*'; artikel 3(1) Voorstel AI Act. In de Compromistekst van de Raad en de EP-tekst is de reikwijdte van de definitie beperkt. In de EP-tekst zijn naast de AI-systemen, foundationmodellen als aparte categorie toegevoegd. Ik voeg hier de definitie van AI-systemen uit de EP-tekst toe: '*'artificial intelligence system' (AI system) means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions that influence physical or virtual environments*'. De foundationmodellen bespreek ik in paragraaf 2.3.3 en 2.3.5.

87 Zie onder andere paragraaf 2.2 en 2.3 over subsidiariteit en evenredigheid. Zie bijvoorbeeld overweging 1, 5, 13, 27 en in het bijzonder 28 waarin de gezondheid aan mensenrechten en veiligheid wordt toegevoegd.

88 In termen van de CSRD zou dit worden aangeduid met 'impact' in plaats van 'risico', maar aangezien het Voorstel AI Act een risicobenadering hanteert en niet een impact-benadering, gebruik ik hier de terminologie van het Voorstel AI Act.

89 Een OECD-onderzoek uit 2022 geeft een goed overzicht van de verschillende aspecten van de 'E'-impact van AI op het gebied van energieconsumptie en de daarbij behorende CO_2-emissie, het waterverbruik en het gebruik van (zeldzame) grondstoffen: OECD (2022), 'Measuring the Environmental Impacts of Artificial Intelligence Compute and Applications; The AI Footprint', OECD Digital Economy Papers, november 2022, No. 341; https://www.oecd-ilibrary.org/docserver/7babf571-en.pdf?expires=1687282813&id=id&accname=guest&checksum=6F0BA0507037879785E46A3B02C2D154.

verhoudt tot deze grondrechten die de voorgestelde AI Act tracht te beschermen.

2.3.2 *Klimaat & milieu en grondrechten*

AI kan een negatieve impact hebben op bepaalde grondrechten opgenomen in het Handvest van de grondrechten van de EU. Dit betreft onder andere het recht op menselijke waardigheid, het recht op leven en eerbiediging van het privéleven en het familie- en gezinsleven, het recht op vrijheid van meningsuiting etc. In artikel 37 Handvest staat dat een hoog niveau van milieubescherming en de verbetering van de kwaliteit van het milieu in het beleid van de Unie moeten worden geïntegreerd en overeenkomstig het beginsel van duurzame ontwikkeling worden gewaarborgd. Dit is echter geen erkenning van een recht op een gezonde leefomgeving. Het Verdrag voor de Rechten van de Mens bevat ook geen recht op een gezonde leefomgeving.[90] In de Nederlandse Grondwet is een dergelijk recht ook niet opgenomen, maar staat een met artikel 37 Handvest vergelijkbaar artikel in de Grondwet.[91] Wel zijn in Nederland rechtstreeks werkende mensenrechten zoals artikel 2 en 8 EVRM breed geïnterpreteerd door de rechter.[92] Dit houdt in dat de rechter heeft geoordeeld dat het recht op leven en het recht op gezinsleven geraakt worden door de risico's van klimaatverandering. In een resolutie van het Europees Parlement over de gevolgen van klimaatverandering voor de mensenrechten is duidelijk omschreven op welke wijze de uitoefening, bescherming en bevordering van op menselijke waardigheid gebaseerde mensenrechten en een gezonde en duurzame planeet onderling afhankelijk zijn.[93] De Raad van Europa heeft dit tevens benadrukt in de laatste versie van zijn handleiding ten aanzien van mensenrechten en het milieu.[94] Daar komt bij dat de algemene vergadering van de Verenigde Naties in 2022 een universeel mensenrecht op een schoon, gezond en duurzaam leefmilieu heeft erkend en staten, internationale organisaties, ondernemingen en alle andere relevante stakeholders heeft opgeroepen om een schoon, gezond en duurzaam leefmilieu te verzekeren.[95] Benadrukt wordt dat dit mensenrecht belangrijk is voor

90 De Parlementaire vergadering van de Raad van Europa heeft in 2021 opnieuw een aanbeveling gedaan om dit recht toe te voegen aan dit verdrag nadat eenzelfde poging in 2009 was gestrand: Parliamentary Assembly of the Council of Europe (2021), 'Anchoring the right to a healthy environment: need for enhanced action by the Council of Europe', Resolution 2396 (2021); https://pace.coe.int/pdf/658d3f594762736ba3c0f378798b2c9529cf4be34aa45a8c38616ecd18fa80c0/res.%202396.pdf en Parliamentary Assembly of the Council of Europe (2009), 'Drafting an additional protocol to the European Convention on Human Rights concerning the right to a healthy environment', Recommendation 1885 (2009); https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=17777&lang=en.
Zie ook: https://pace.coe.int/en/news/8452/the-right-to-a-healthy-environment-pace-proposes-draft-of-a-new-protocol-to-the-european-convention-on-human-rights/. De Parlementaire vergadering van de Raad van Europa pakt ook de duurzaamheidsuitdagingen en digitalisering samen: "*While the threats of environmental degradation and climate change are among the biggest challenges facing humanity today, the Assembly views the unfettered use of certain new, man-made technologies (such as artificial intelligence, nanotechnology and genetic engineering) as a human rights challenge. It therefore considers that the Council of Europe should prepare a "Five Ps" convention on environmental threats and technological hazards threatening human health, dignity and life – in the spirit of the Stockholm Declaration. By preventing and prosecuting violations of the right to a safe, clean, healthy and sustainable environment, and protecting the victims, the contracting States would adopt and implement state-wide integrated policies that are effective and offer a comprehensive response to environmental threats and technological hazards, involving parliaments in holding governments to account for the effective implementation of environment-friendly pro-human rights policies.*"; Parliamentary Assembly of the Council of Europe (2021), 'Anchoring the right to a healthy environment: need for enhances action by the Council of Europe', Doc 15367, 13 september 2021, paragraaf 13; https://pace.coe.int/en/files/29409#trace-3.

91 Artikel 21 Grondwet.

92 HR 20 december 2019, ECLI:NL:HR:2019:2006 (*Urgenda*), r.o. 5.2.2. t/m 5.2.4 en 5.3.1. Zie ook: N.J. Schrijver (2023), 'Internationaal klimaatrecht. Een kolkende stroom', *RMThemis* 2023-1, p. 1-8 en de daarin opgenomen verwijzingen.

93 Europees Parlement (2021), 'De gevolgen van klimaatverandering voor de mensenrechten en de rol die milieuactivisten in dit kader spelen', Resolutie van het Europees Parlement van 19 mei 2021 over de gevolgen van klimaatverandering voor de mensenrechten en de rol die milieuactivisten in dit kader spelen (2020/2134(INI), P9_TA(2021)0245); https://www.europarl.europa.eu/doceo/document/TA-9-2021-0245_NL.pdf. Zie ook voor een beknopte beschrijving van de verhouding tussen het milieu en mensenrechten: Europees Parlement (2021), 'A universal right to a healthy environment'; https://www.europarl.europa.eu/RegData/etudes/ATAG/2021/698846/EPRS_ATA(2021)698846_EN.pdf. De Raad heeft dit onderstreept in zijn conclusies over de EU-prioriteiten in de VN-mensenrechtenfora. Conclusie 15: '*De EU zal maatregelen tegen de gevolgen van klimaatverandering, biodiversiteitsverlies en milieudegradatie voor het volledige genot van mensenrechten blijven steunen. De EU zal in dit verband de belangrijke rol van verdedigers van milieugerelateerde mensenrechten en mensenrechten van inheemse volkeren bevorderen. De EU zal actief deelnemen aan de besprekingen ter bevordering van het mensenrecht op een schoon, gezond en duurzaam milieu, en zal inclusie en non-discriminatie bevorderen. De EU zal zich actief blijven inzetten voor VN-resoluties over de samenhang van mensenrechten, klimaat en milieu, onder meer in het licht van de resultaten van de COP 27. De EU onderstreept het belang van toegang tot informatie, inspraak bij besluitvorming en toegang tot de rechter als het gaat om het milieu.*', persmededeling 20 februari 2023; https://www.consilium.europa.eu/nl/press/press-releases/2023/02/20/council-conclusions-on-eu-priorities-in-un-human-rights-fora-2023/. Zie ook: D.A. Dam-de Jong (2023), "Botsende' mensenrechten in klimaatbeleid: een pleidooi voor een inclusieve energietransitie', *RMThemis* 2023-1, p. 26-37.

94 Council of Europe (2022), 'Manual on Human Rights and the Environment (3rd edition): Principles emerging from the case law of the European Court on Human Rights and the conclusions and decisions of the European Committee of Social Rights', p. 9 en 13, met verwijzingen naar relevante jurisprudentie van het Europese Hof voor de Rechten van de Mens; https://rm.coe.int/manual-environment-3rd-edition/1680a56197. Ook in het Klimaatakkoord van Parijs is opgenomen dat staten bij hun acties om klimaatverandering tegen te gaan hun verplichtingen betreffende mensenrechten dienen te eerbiedigen, te bevorderen en in aanmerking te nemen; zie de negende overweging van de Overeenkomst van Parijs, *PbEU* L 282/4, 10 oktober 2016; https://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=CELEX:22016A1019(01).

95 United Nations General Assembly (2022), 'Resolution adopted on the human right to a clean, healthy and sustainable environment', 26 juli 2022, A/76/L.75; https://digitallibrary.un.org/record/3982508?ln=en. Hoewel deze resolutie niet juridisch bindend is, vormt het wel een belangrijk commitment van de leden.

het kunnen uitoefenen van andere mensenrechten.[96] Het is dus te beargumenteren dat de in het Voorstel AI Act benoemde bescherming van grondrechten tevens ook (indirect) bescherming van klimaat en milieu inhoudt en dus dat de in het Voorstel AI Act opgenomen verplichtingen, zoals het uitvoeren van een risicoassessment en het nemen van risicomitigatiemaatregelen, ook de risico's voor klimaat en milieu omvatten. Toch is in de EP-tekst de reikwijdte (grondrechten, gezondheid en veiligheid) uitgebreid met de bescherming van het 'klimaat en milieu'.[97] Hierdoor komt nadrukkelijk(er) het brede spectrum van ESG in zicht, omdat nu ook de 'E' expliciet een direct onderdeel wordt van de bescherming die deze wetgeving biedt. De toevoeging zorgt daarmee voor een betere aansluiting bij de Europese digitale rechten en beginselen en de verwevenheid van de twin transitions. In de inleidende tekst en de overwegingen van het Voorstel AI Act wordt overigens ook gewezen op de mogelijke positieve effecten die bereikt kunnen worden in de 'E'-pijler door de inzet van AI-systemen.[98] Zo wordt gesteld dat AI kan helpen bij het efficiënte gebruik van hulpbronnen en energie, en klimaatmitigatie en -adaptatie.[99] Ondersteuning van innovatie door gepaste wetgeving die niet te veel belemmert, dient zodoende bij te dragen aan positieve effecten in de 'E'-pijler.

2.3.3 *Verschillende soorten AI-systemen en AI-modellen*

De focus van het Voorstel AI Act ligt echter op risicobescherming. Het voorstel hanteert hiervoor een risicogebaseerde benadering van AI-systemen waarbij vier niveaus van risico's worden onderscheiden; verboden risico's, hoge risico's, beperkte risico's en verwaarloosbare risico's elk met een eigen set met regels.

Verboden en hoog risico AI-systemen

Sommige AI-systemen worden verboden omdat zij een onaanvaardbaar risico teweegbrengen.[100] Dit zijn bijvoorbeeld systemen die de betrouwbaarheid van mensen beoordelen op basis van sociaal gedrag of op basis van persoonlijke kenmerken in verschillende contexten ('social scoring').[101] De AI-systemen die een hoog risico met zich brengen (HRAIS) worden streng gereguleerd. Het voorstel bevat voor de aanbieders van HRAIS allerlei regels ten aanzien van het implementeren van risico- en kwaliteitsmanagementprocedures gericht op risico-identificatie, -beheersing en -evaluatie en daarnaast incidentrapportage. Ook dient er bijvoorbeeld een 'accountability framework' aanwezig te zijn waarin de verantwoordelijkheden van het management en ander personeel ten aanzien van de verschillende aspecten van de governance van AI zijn opgenomen.[102] Bovendien dient het HRAIS zelf aan allerlei vereisten te voldoen (zie paragraaf 2.3.5). Grofweg zijn er twee categorieën HRAIS. Deze zijn opgenomen in twee bijlagen.[103] De eerste groep betreft AI-systemen die zelf producten zijn die aan Europese productveiligheidswetgeving zijn onderworpen (of een veiligheidsonderdeel zijn van deze producten) en daarnaast onderworpen zijn aan een conformiteitsbeoordeling die door een derde partij dient te worden uitgevoerd (de 'bijlage II-HRAIS'). Voorbeelden zijn machines, medische hulpmiddelen en speelgoed. De tweede groep wordt gevormd door AI-systemen waarvan op basis van de toepassing van het AI-systeem en de daarmee samenhangede risico's besloten is dat deze bestempeld dienen te worden als HRAIS (de 'bijlage III-HRAIS').[104] Bij de vaststelling van deze lijst is gekeken naar de functie die het systeem vervult alsook het specifieke doel en de modaliteiten waarvoor het systeem wordt gebruikt.[105] Voorbeelden van bijlage III-HRAIS zijn bijvoorbeeld systemen die gebruikt worden voor de werving & selectie van nieuw personeel, die worden ingezet voor de beslissing over de toelating tot een opleiding aan een onderwijsinstelling of die de kredietwaardigheid van een persoon vaststellen op basis waarvan wel of niet toegang wordt verleend tot essentiële diensten zoals zorg, huisvesting en energie. Zowel de Compromistekst van de Raad als de EP-tekst bevatten uitbreidingen van deze lijst. Daarnaast introduceert de EP-tekst een belangrijke tweede toets voor de bijlage III-HRAIS.[106] Om als HRAIS te kwalificeren, dient ook vastgesteld te worden dat er daadwerke-

96 In de VN-resolutie is overigens geen definitie van dit recht opgenomen maar een rapport van de VN Mensenrechtenraad bevat wel een verwijzing naar vitale elementen daarvan. Dit zijn: schone lucht, schoon water en adequate sanitaire voorzieningen, gezond en duurzaam voedsel, een veilig klimaat en een gezonde biodiversiteit en ecosystemen. United Nations General Assembly (2019), 'Resolution on the issue of human rights obligations relating to the enjoyment of a safe, clean, healthy and sustainable environment', 8 januari 2019, A/HRC/40/55; https://digitallibrary.un.org/record/1663859?ln=en#record-files-collapse-header.

97 En daarnaast met de bescherming van de 'democratie en rechtsstaat'.

98 Paragraaf 1.1 Voorstel AI Act, p. 1: "*Door voorspellingen te verbeteren, verrichtingen en de toewijzing van middelen te optimaliseren en de dienstverlening te personaliseren kan het gebruik van artificiële intelligentie helpen om gunstige sociale en ecologische resultaten te behalen en belangrijke concurrentievoordelen op te leveren voor het bedrijfsleven en de Europese economie. Dergelijke maatregelen zijn vooral nodig in sectoren met een grote impact, zoals klimaatverandering, milieu en gezondheid, de overheidssector, financiën, mobiliteit, binnenlandse zaken en landbouw.*"

99 Overweging 3 Voorstel AI Act.

100 Artikel 5 Voorstel AI Act.

101 In het Voorstel AI Act was dit beperkt tot systemen ingezet door de overheid. Zowel in de Compromistekst van de Raad (p. 4 en artikel 5(1)(c)) als de EP-tekst (artikel 5(1)(c)) is dit uitgebreid naar private organisaties.

102 Artikel 17(1)(m) Voorstel AI Act.

103 Bijlage II en III van het Voorstel AI Act.

104 De lijst met HRAIS in bijlage III kan onder voorwaarden na inwerkingtreding nog worden gewijzigd of aangevuld door de Europese Commissie: artikel 7 Voorstel AI Act.

105 Voorstel AI Act, p. 15.

106 Artikel 6(2) EP-tekst. Voor bijlage II-HRAIS wordt de voorwaarde van een verplichte conformiteitsbeoordeling door een derde partij, beperkt tot conformiteitsbeoordelingen gerelateerd aan risico's voor de gezondheid of de veiligheid. De Europese Commissie dient volgens de EP-tekst richtlijnen op te stellen voor de praktische implementatie van deze toets, die door de aanbieders van de AI-systemen zelf uitgevoerd dient te worden (artikel 82b(d) EP-tekst). Indien de conclusie is dat het specifieke AI-systeem wel onder de in bijlage III genoemde toepassingen valt, maar niet tot het voornoemde significante risico leidt, zal de aanbieder deze conclusie met bijbehorende onderbouwing als notificatie sturen aan de relevante nationale toezichthouder. Artikel 6(2a) EP-tekst. De toezichthouder dient binnen drie maanden te beslissen of er sprake is van een misclassificatie. Hiertegen kan de aanbieder dan weer in beroep gaan.

lijk sprake is van een significant risico op schade met betrekking tot de gezondheid, veiligheid of fundamentele rechten van natuurlijke personen.[107] Het gaat hierbij dus specifiek om schade van natuurlijke personen. Voor een bepaalde groep bijlage III-HRAIS geldt een andere toets, namelijk of er sprake is van een significant risico van schade ten aanzien van het klimaat- en milieu.[108] Dit zijn HRAIS met een specifieke toepassing, namelijk AI-systemen die worden ingezet voor het management en functioneren van kritische infrastructuur.[109] Een voorbeeld van een dergelijk systeem is software die gebruikt wordt in het kader van het management van een waterdam. Deze software doet voorspellingen over de invloed van wijzigingen in de natuurlijke omgeving (zoals verzakkende grond) op basis van een simulatiemodel. Een foute voorspelling als gevolg waarvan te laat wordt ingegrepen, kan grote gevolgen hebben voor die natuurlijke omgeving.[110] Een AI-systeem met een grote 'E'-impact wordt dus niet vanwege deze 'E'-impact als HRAIS aangemerkt.

Beperkte en verwaarloosbare risico AI-systemen

Naast verboden AI-systemen en HRAIS, onderscheidt het Voorstel AI Act een groep specifieke AI-systemen die worden aangemerkt als systemen met een beperkt risico (zoals deepfakes, chatbots en bepaalde emotieherkenningssystemen). Deze systemen zijn aan transparantieverplichtingen onderhevig.[111] Reden hiervoor is dat het voor de betrokkenen duidelijk moet zijn dat het gaat om content gecreëerd door een AI-systeem of dat zij interacteren met een AI-systeem. Voor marktpartijen of organisaties die AI-systemen met een verwaarloosbaar of laag risico ontwikkelen, bevat het voorstel een aansporing om zichzelf gelijksoortige regels als die van toepassing zijn op HRAIS op te leggen via zelfregulering zoals gedragscodes.[112]

Foundationmodellen, generatieve AI, generatieve AI-systemen en general purpose AI-systemen

Recente ontwikkelingen met LLM's ('large language models') hebben ervoor gezorgd dat de Raad eind 2022 op de valreep nog bepalingen voor 'general purpose AI-systemen'[113] heeft opgenomen in de Compromistekst.[114] Het Europees Parlement introduceert in de EP-tekst aangenomen op 1 juni van dit jaar, verschillende begrippen en regels die samenhangen met deze LLM's (en breder zijn in scope dan deze LLM's). Dit zijn de begrippen 'foundation model', 'generative AI', 'generative AI system' en 'general purpose AI system'. Foundationmodellen, zoals GPT-4, zijn AI-modellen die inzetbaar zijn voor algemene toepassingen en het uitvoeren van een brede selectie van taken.[115] Zij kunnen bijvoorbeeld (eventueel na aanpassing) geïntegreerd worden in producten of websites en vormen derhalve een onderdeel van een waardeketen met downstream-toepassingen. Omdat foundationmodellen multi-inzetbaar zijn, kunnen dit dus vele waardeketens in verschillende domei-

107 'Significant risico' is gedefinieerd in artikel 3(1b) EP-tekst: "*significant risk' means a risk that is significant as a result of the combination of its severity, intensity, probability of occurrence, and duration of its effects, and its the ability to affect an individual, a plurality of persons or to affect a particular group of persons*'. In overweging 32 staat hierover het volgende: '*Such significant risk of harm should be identified by assessing on the one hand the effect of such risk with respect to its level of severity, intensity, probability of occurrence and duration combined altogether and on the other hand whether the risk can affect an individual, a plurality of persons or a particular group of persons. Such combination could for instance result in a high severity but low probability to affect a natural person, or a high probability to affect a group of persons with a low intensity over a long period of time, depending on the context.*'

108 Artikel 6(2) EP-tekst.

109 Dit betreft weg-, spoor-, luchtverkeer, water, gas, verwarming, elektriciteit en kritische digitale infrastructuur.

110 https://www.euractiv.com/section/next-generation-infrastructure/news/the-ai-acts-fine-line-on-critical-infrastructure/.

111 Artikel 52 Voorstel AI Act.

112 Artikel 69 Voorstel AI Act.

113 De definitie die de Raad heeft gekozen voor een 'general purpose AI system' is: '*an AI system that – irrespective of how it is placed on the market or put into service, including as open source software – is intended by the provider to perform generally applicable functions such as image and speech recognition, audio and video generation, pattern detection, question answering, translation and others; a general purpose AI system may be used in a plurality of contexts and be integrated in a plurality of other AI systems*'.

114 Artikel 4a en 4b Compromistekst van de Raad.

115 De definitie die in artikel 3 EP-tekst is opgenomen luidt: "*foundation model' means an AI system model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks*'. In de eerste versie van de EP-tekst die is aangenomen door de verantwoordelijke EP-commissies was het woordje 'system' geen onderdeel van de definitie. In de rest van de EP-tekst wordt echter een onderscheid gemaakt tussen foundationmodellen die wel en niet geïntegreerd zijn in AI-systemen en worden AI-systemen en foundationmodellen naast elkaar geplaatst. Niet helemaal duidelijk is daarom hoe 'AI system model' uit de definitie zich daartoe verhoudt. Committee on Internal Market and Consumer Protection and Committee on Civil Liberties, Justice and Home Affairs (2023), 'Draft Comprise Amendments on the Draft Report, Proposal for a regulation of the European Parliament and of the Council on harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts', COM(2021)0206-C9 0146/2021 – 2021/0106(COD), 9 mei 2023; https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf. Voor een goede uitleg over foundationmodellen zie: R. Bommasani et al. (2022), 'On the Opportunities and Risks of Foundation Models'; https://crfm.stanford.edu/report.html.
In overweging 60e staat het volgende: '*Foundation models are a recent development, in which AI models are developed from algorithms designed to optimize for generality and versatility of output. Those models are often trained on a broad range of data sources and large amounts of data to accomplish a wide range of downstream tasks, including some for which they were not specifically developed and trained. The foundation model can be unimodal or multimodal, trained through various methods such as supervised learning or reinforced learning. AI systems with specific intended purpose or general purpose AI systems can be an implementation of a foundation model, which means that each foundation model can be reused in countless downstream AI or general purpose AI systems. These models hold growing importance to many downstream applications and systems.*'

nen zijn.[116] Voorbeelden van applicaties die op basis van GPT-4 zijn gebouwd zijn een applicatie die financiële adviseurs in het domein van vermogensbeheer voorziet van gemakkelijk te begrijpen kennis en inzichten, antwoorden op door hen gestelde vragen en links naar documenten,[117] en bijvoorbeeld meta-GPT, een *text-to-app*-platform waarmee gebruikers gemakkelijk web-apps kunnen bouwen zonder dat er enige technische kennis voor nodig is.[118]

Foundationmodellen brengen een aantal risico's met zich, zoals het versterken van sociale bias, discriminatie en het verspreiden van fake news.[119] Door het (mogelijke) wijdverbreide gebruik van deze modellen in verschillende sectoren bestaat het risico van homogenisering en machtconcentratie. De keuzes die worden gemaakt bij het ontwerp en de ontwikkeling kunnen daardoor een ongewenst vliegwieleffect teweegbrengen in de maatschappij.[120] Het creëren van een foundationmodel vergt een enorme investering zowel in het kader van het trainen van het model als het gebruik en is daardoor slechts weggelegd voor een beperkt aantal techspelers. Door de bescherming van de AI Act alleen te richten op de toepassingen die gebruikmaken van deze foundationmodellen (AI-systemen) die veelal geen controle hebben over de foundationmodellen en niet betrokken zijn bij het ontwerp en de ontwikkeling van het model, wordt een belangrijke bron van risico's niet aangepakt. Dat is de reden dat er nieuwe regels voor deze modellen zijn geïntroduceerd in de EP-tekst. Foundationmodellen kwalificeren daarin niet als HRAIS, maar als foundationmodel en hebben een eigen set van regels.[121] Dit betekent dat naast het begrip 'AI-systeem' het begrip 'foundationmodel' wordt gehanteerd. Een foundationmodel kan, zoals gezegd, worden gebruikt voor een applicatie (een AI-systeem) die 'bovenop' het foundationmodel wordt gebouwd. Generatieve AI verwijst in de EP-tekst naar foundationmodellen zoals LLM's, die gebruikt worden in AI-systemen die specifiek gericht zijn op (met verschillende niveaus van autonomie) het genereren van content, zoals complexe tekst, afbeeldingen, audio of video.[122] Het is daarmee een specifieke soort foundationmodel omdat er ook foundationmodellen zijn die gebruikt worden voor niet-generatieve taken zoals het automatiseren van processen, classificatie of bijvoorbeeld informatie-extractie uit documenten.[123] Een voorbeeld van een 'generatief AI-systeem' is ChatGPT dat als applicatie op het foundationmodel 'GPT' is gebouwd, waardoor er interactie met een gebruiker kan plaatsvinden. In het kader van generatieve AI en generatieve AI-systemen gelden speciale regels gericht op contentgeneratie en transparantie over gebruik van beschermde trainingsdata.[124] Een generatieve AI-systeem kan ook voor verschillende specifieke praktische toepassingen ontwikkeld worden, zoals gebruik in recruitment en marketing. Dit generatieve AI-systeem zou derhalve door de specifieke toepassing als HRAIS kunnen kwalificeren. In de EP-tekst wordt in de overwegingen uitgelegd dat foundationmodellen daarnaast ook gebruikt kunnen worden door general purpose AI-systemen, maar dit zijn niet noodzakelijk AI-systemen die content genereren.[125] Deze systemen worden aangeduid als systemen die gebruikt worden in en aangepast worden voor een breed scala aan toepassingen waarvoor het systeem niet specifiek is ontworpen.[126] Dit is dus een categorie die breder is dan generatieve AI-systemen. Er zijn in de EP-tekst geen specifieke regels opgenomen voor general purpose AI-systemen als aparte categorie. De Raad heeft in tegenstelling tot het EP geen onderscheid gemaakt tussen de verschillende begrippen, maar gekozen voor één begrip; 'general purpose AI system'.[127] Specifieke regels gelden alleen als een general purpose AI-systeem als een HRAIS wordt gebruikt of als component van een HRAIS wordt ingezet. In die gevallen zouden volgens de Raad (sommige) regels die op HRAIS van toepassing zijn ook voor general purpose AI moeten gelden. De regels die samenhangen met foundationmodellen zoals voorgesteld door het EP en de Raad bespreek ik in paragraaf 2.3.5.

2.3.4 *Verschillende soorten partijen in de AI-keten*

Naast de verschillende regels voor de verschillende categorieën AI-systemen en AI-modellen, maakt het voorstel onderscheid in partijen in de waardeketen op wie die regels van toepassing zijn. De belangrijkste zijn de aanbie-

116 Foundationmodellen worden over het algemeen ter beschikking gesteld via een API ('application programming interface') die gecontroleerd wordt door de aanbieder van het foundationmodel of via open source aan andere marktpartijen, zodat zij hun applicaties als laag hierop kunnen bouwen. Bij de API-route draait het model nog steeds op de servers van de aanbieder van het foundationmodel. Dit betekent dat marktpartijen gebruik kunnen maken van het model zonder dat zij alle technische details van het model hoeven te kennen. Bij open source worden het foundationmodel en de daarbij behorende technische specificaties ter beschikking gesteld via een platform waar marktpartijen het model kunnen downloaden, waarna er in principe geen interactie tussen de aanbieder en de gebruiker van het model meer hoeft plaats te vinden. De gebruiker dient dan wel te beschikken over voldoende infrastructuur om het model te kunnen draaien, omdat dit dus niet draait op de servers van de aanbieder van het foundationmodel.

117 https://www.morganstanley.com/press-releases/key-milestone-in-innovation-journey-with-openai.

118 https://picoapps.xyz.

119 L. Weidingen et al. (2021), 'Ethical and social risks of harm from Language Models'; https://arxiv.org/pdf/2112.04359.pdf, en R. Bommasani et al. (2022), *supra* noot 115.

120 R. Bommasani et al. (2022), *supra* noot 115, p. 5 en 152-160.

121 Overweging 60g EP-tekst: '*These specific requirements and obligations do not amount to considering foundation models as high risk AI systems, but should guarantee that the objectives of this Regulation to ensure a high level of protection of fundamental rights, health and safety, environment, democracy and rule of law are achieved.*'

122 Artikel 28b(4) EP-tekst.

123 https://research.ibm.com/blog/what-is-generative-AI.

124 Artikel 28b(4)(a-c) en artikel 52(1) EP-tekst. Zie verder paragraaf 2.3.5.

125 Overweging 60e EP-tekst.

126 Artikel 3(1d) EP-tekst.

127 Zie noot 113.

der ('provider')[128] en de gebruiker ('user', deze wordt in de EP-tekst 'deployer' genoemd).[129] De verplichtingen richten zich met name op de aanbieders van AI-systemen en foundationmodellen en in mindere mate op de gebruikers.[130] De EP-tekst brengt daar in het kader van impactassessments bij HRAIS wel enige verandering in. Ik kom daar in paragraaf 2.3.5 op terug. Van belang is dat een gebruiker een rolverandering kan ondervinden waarbij onder omstandigheden deze gebruiker (tevens) wordt aangemerkt als een aanbieder van een HRAIS. Onder de EP-tekst gebeurt dit indien (1) de gebruiker zijn eigen naam of handelsmerk plaatst op een HRAIS dat al op de markt is gebracht of reeds in gebruik is gesteld; (2) de gebruiker een substantiële wijziging heeft aangebracht aan een HRAIS dat al op de markt is gebracht of reeds in gebruik is gesteld en blijft kwalificeren als een HRAIS; of (3) de gebruiker een substantiële wijziging heeft aangebracht aan een AI-systeem dat al op de markt is gebracht of reeds in gebruik is gesteld en dit AI-systeem door deze wijziging kwalificeert als een HRAIS.[131] De gebruiker dient dan te voldoen aan de verplichtingen die op een HRAIS-aanbieder van toepassing zijn. De oorspronkelijke aanbieder wordt niet meer aangemerkt als de aanbieder van dat AI-systeem. Wel dient de 'voormalige' aanbieder aan bepaalde verplichtingen te voldoen zoals het verschaffen van informatie en het verlenen van toegang en ondersteuning aan de gebruiker (nieuwe aanbieder) (met in achtneming van bedrijfsgeheimen en intellectueel eigendom) om deze in staat te stellen compliant te zijn met de verplichtingen die van toepassing zijn op een aanbieder van een HRAIS. Dit geldt ook als het gaat om een aanbieder van foundationmodellen die direct geïntegreerd zijn in een HRAIS of in een AI-systeem dat al op de markt was gebracht en door een substantiële wijziging door de gebruiker een HRAIS wordt.[132] In aanvulling op de aanbieders en de gebruikers onderscheidt de EP-tekst ook nog 'derde partijen' die instrumenten, diensten, componenten of processen verschaffen die gebruikt worden voor of geïntegreerd worden in HRAIS. Op de aanbieder en de derde partij rust de verplichting om contractueel af te spreken welke informatie, capaciteiten, technische toegang en/of andere ondersteuning moeten worden verschaft door de derde partij zodat de aanbieder aan de regels van de AI Act kan voldoen. Dit is van belang voor het kunnen nemen van risicomitigatiemaatregelen door de aanbieder.

2.3.5 *Verplichtingen in het Voorstel AI Act en ESG*

Algemene beginselen in AI-wetgeving

De voorgestelde wetgeving bevat dus specifieke sets met regels voor specifieke toepassingen gelinkt aan specifieke partijen. De EP-tekst introduceert echter ook een artikel waarin een algemene inspanningsverplichting ('best-effort-artikel') is opgenomen voor zowel aanbieders als gebruikers[133] om AI-systemen en foundationmodellen te ontwikkelen en te gebruiken in lijn met algemene Europese beginselen voor ethische en betrouwbare kunstmatige intelligentie. Een regel die dus van toepassing is op alle AI-systemen en foundationmodellen en op meerdere partijen in de waardeketen. Dit is een belangrijk artikel omdat hierin de beginselen zelf zijn opgenomen. Eén van deze principes is het beginsel van 'social and environmental well-being'. Dit houdt in dat AI-systemen ontwikkeld en gebruikt dienen te worden op een duurzame en milieuvriendelijke wijze alsook op een manier waarin alle mensen ervan profiteren en waarbij de langetermijnimpact op individuen, de maatschappij en de democratie wordt gemonitord en geanalyseerd.[134] In ditzelfde artikel zijn ook de andere principes opgenomen: menselijk toezicht, privacy & data governance, technische robuustheid en veiligheid, transparantie en diversiteit, non-discriminatie & rechtvaardigheid. De meeste beginselen zijn dus op de 'S'-pijler gericht, maar worden wel aangevuld met 'E'-aspecten. Om compliant te zijn met deze beginselen dienen de aanbieders en gebruikers volgens het betreffende artikel uit de EP-tekst te voldoen aan de regels van de AI Act die op hen van toepassing zijn. Dit betekent dus dat het uitgangspunt is dat deze beginselen voor alle AI-systemen en foundationmodellen ingebed zijn in de regels van de AI Act zodat, als zij daaraan voldoen, zij voldoen aan de beginselen. Het is de vraag of dit ook daadwerkelijk zo is, met name als het gaat om AI-systemen die geen HRAIS zijn. Dit bespreek ik in paragraaf 2.3.6.

128 *'"Aanbieder": een natuurlijke of rechtspersoon, overheidsinstantie, agentschap of ander orgaan die/dat een AI-systeem ontwikkelt of beschikt over een AI-systeem dat is ontwikkeld met het oog op het al dan niet tegen betaling in de handel brengen of in gebruik stellen ervan onder de eigen naam of merknaam*'; artikel 3(2) Voorstel AI Act. 'In de handel brengen' ('placing on the market') is niet gedefinieerd. 'In gebruik stellen' ('*putting into service*') daarentegen wel: '*'putting into service' means the supply of an AI system for first use directly to the deployer or for own use on the Union market for its intended purpose*'; artikel 3(11) EP-tekst. Dezelfde omschrijving is opgenomen in het Voorstel AI Act.

129 *'"Gebruiker": een natuurlijke of rechtspersoon, overheidsinstantie, agentschap of ander orgaan die/dat een AI-systeem onder eigen verantwoordelijkheid gebruikt, tenzij het AI-systeem wordt gebruikt in het kader van een persoonlijke niet-beroepsactiviteit*'; artikel 3(4) Voorstel AI Act.

130 Zie voor een Civil Society Statement waarin deze kritiek ook is opgenomen: EDRi et al. (2021), 'An EU Artificial Intelligence Act for Fundamental Rights: A Civil Society Statement', 30 november 2021; https://edri.org/wp-content/uploads/2021/12/Political-statement-on-AI-Act.pdf.

131 Artikel 28 Voorstel AI Act en EP-tekst. Een substantiële wijziging is als volgt gedefinieerd: '*'substantial modification' means a modification or a series of modifications of the AI system after its placing on the market or putting into service which is not foreseen or planned in the initial risk assessment by the provider and as a result of which the compliance of the AI system with the requirements set out in Title III, Chapter 2 of this Regulation is affected or results in a modification to the intended purpose for which the AI system has been assessed*'; artikel 3(23) EP-tekst.

132 Artikel 28(2) en artikel 28(1)(b a) EP-tekst. De verplichtingen die rusten op een aanbieder van een foundationmodel hebben betrekking op processen die dienen plaats te vinden voordat het foundationmodel op de markt wordt geplaatst of in gebruik wordt gesteld; artikel 28b EP-tekst.

133 Alsook geautoriseerde vertegenwoordigers, de importeur en de distributeur, gezamenlijk genoemd: 'operators'; artikel 3(8) EP-tekst.

134 Artikel 4a(1)(f) EP-tekst.

Aanbieders van HRAIS

Kwaliteits- en risicomanagementsysteem

De bescherming die het Voorstel AI Act tracht te bewerkstelligen is voor het grootste gedeelte gericht op de negatieve sociale impact die HRAIS kunnen veroorzaken en de governance-aspecten die benodigd zijn om deze impact te mitigeren. Aanbieders van HRAIS dienen een kwaliteitsmanagementsysteem te hebben waar een risicomanagementsysteem onderdeel van uitmaakt.[135] Dit risicomanagementsysteem betreft een proces van (1) identificatie en analyse van bekende en voorzienbare risico's;[136] (2) het inschatten en evalueren van risico's die zich kunnen voordoen, gelet op het beoogde doel en voorzienbaar misbruik; (3) evaluatie van andere risico's die zich kunnen voordoen op basis van monitoringsdata; en (4) het vaststellen en implementeren van geschikte risicobeheersingsmaatregelen. Deze maatregelen dienen de risico's te elimineren of dienen deze zo goed als mogelijk te reduceren via het ontwerp en de ontwikkeling van het HRAIS. De EP-tekst voegt daaraan toe dat hierbij experts en stakeholders moeten worden betrokken indien relevant.[137] Ook dient rekening te worden gehouden met de technische kennis, ervaring, opleiding en training die een gebruiker nodig heeft.[138] Indien eliminatie niet mogelijk is, worden risico's gemitigeerd en dienen er controlemaatregelen te worden genomen. Risico's die overblijven (individueel en geaggregeerd beschouwd) dienen acceptabel te zijn en moeten worden gecommuniceerd met de gebruiker.[139] De EP-tekst voegt hieraan toe dat het moet gaan om een onderbouwde beoordeling van de aanvaardbaarheid van de overblijvende risico's. Dat is van belang omdat het de aanbieder zelf is die deze beoordeling uitvoert. Op de aanbieder rust ook de verplichting om ex post (dus na plaatsing op de markt) de risico's van het HRAIS te blijven monitoren en hiervoor een plan op te stellen.[140]

Risicobeheersingsmaatregelen te nemen door de aanbieder van HRAIS

Het Voorstel AI Act stelt vervolgens specifieke eisen aan de risicobeheersingsmaatregelen. Ik bespreek ze niet allemaal. De regels hebben onder andere betrekking op de data en datagovernance zoals de training, validatie en het testen van datasets.[141] Maatregelen die moeten worden genomen betreffen onder andere transparantie over het doel van dataverzameling, voorbereidende processen voor dataverwerking, analyse van de geschiktheid van data en het onderzoeken van mogelijke biases. De EP-tekst voegt daaraan toe dat deze maatregelen ook de mogelijke biases dienen te detecteren, voorkomen en mitigeren.[142] Dit is een belangrijke aanvulling omdat alleen een verplichting tot 'onderzoeken' niet voldoende is om bias ook daadwerkelijk tegen te gaan. Ook heeft de EP-tekst de nodige waarborgen toegevoegd voor de situatie waarin er persoonlijke data worden gebruikt voor het detecteren en corrigeren van negatieve bias.[143] Verder zijn er verplichtingen op het gebied van de technische documentatie en het bijhouden van logs. In de technische documentatie dient naast algemene informatie over het AI-systeem, uitgebreide info te worden opgenomen ten aanzien van het ontwerp en het ontwikkelproces alsook het gebruik van het HRAIS. Informatie over de geïdentificeerde voorzienbare risico's dient te worden gegeven.[144] In de EP-tekst dient ook informatie over het energieverbruik gedurende de ontwikkeling alsook het verwachte energieverbruik gedurende het gebruik te worden gegeven.[145] De EP-tekst gebruikt de vereiste loggingseigenschap om tegelijkertijd de verplichting op te leggen de energieconsumptie, het bronnenverbruik en de impact op klimaat en milieu gedurende de levenscyclus van het HRAIS te registreren.[146] Hieraan gekoppeld zijn twee belangrijke nieuwe overwegingen toegevoegd.[147] Zo staat er in deze overwegingen dat AI-systemen state-of-the-art-methoden in acht moeten nemen om energie te besparen, efficiënt gebruik te maken van bronnen en afval te reduceren. Niet duidelijk is waar deze verplichting in de artikelen voor HRAIS is opgenomen. Wellicht dat dit onderdeel is van de verplichte risicomitigatie, maar zoals de verplichting is opgenomen in de overweging lijkt het te gaan om een meer algemene verplichting die ook geldt op het moment dat energieconsumptie niet als risico is geïdentificeerd. Deze meer algemene verplichting is wel expliciet opgenomen bij de verplichtingen voor aanbieders van foundationmodellen.[148] De Europese Commissie zou volgens de EP-tekst daarnaast standaarden moeten opstellen waarin de methodologie is opgenomen voor het meten van de 'E'-impact van een AI-systeem.[149] Deze standaarden zouden ook behulpzaam kunnen zijn bij keuzes over eventuele toekomstige regulering op dit gebied.[150] Daarnaast zou de EC volgens de EP-tekst een methodologie kunnen opstellen om Key Performance Indicators (KPI's) te ontwikkelen op het gebied van duurzaamheid van AI-systemen, zodat targets kunnen worden vastgesteld. Dit zou kunnen leiden tot een rechtvaardige vergelijking van keuzes die zijn gemaakt bij het ontwerp en de ontwik-

135 Artikel 16 jo. 17 jo. 9 t/m 15 Voorstel AI Act.
136 In de EP-tekst is dit het brede spectrum van risico's met inbegrip van klimaat en milieu: '*(…) health or safety of natural persons, their fundamental rights including equal access and opportunities, democracy and rule of law or the environment (…)*'; artikel 9(2)(a) EP-tekst.
137 Artikel 9(4)(1a) EP-tekst.
138 Artikel 9(4)(2) EP-tekst.
139 Artikel 9(4) EP-tekst.
140 Artikel 61 Voorstel AI Act.
141 Artikel 10 Voorstel AI Act.
142 Artikel 10(2)(f) EP-tekst.
143 Artikel 10(5) EP-tekst.
144 Bijlage IV Voorstel AI Act en de EP-tekst.
145 Bijlage IV (3b) EP-tekst.
146 Artikel 12(2a) EP-tekst.
147 Overweging 46a en 46b EP-tekst.
148 Artikel 28b(2)(d) EP-tekst.
149 Overweging 87a en artikel 82b(h) EP-tekst.
150 "*Such common specifications on measurement methodology can develop a baseline upon which the Commission can better decide if future regulatory interventions are needed, upon conducting an impact assessment that takes into account existing legislation.*"

keling van AI-systemen en kan daarmee bijdragen aan de duurzaamheidsdoelstellingen van de EU.[151]

Voordat een HRAIS op de markt wordt geplaatst dient er een conformiteitsbeoordeling plaats te vinden. Deze zal voor bijlage II-HRAIS ten aanzien van risico's voor de gezondheid en veiligheid in beginsel door een externe partij worden uitgevoerd omdat dit een vereiste is voor de classificatie als HRAIS. In het geval van bijlage III-HRAIS zal de aanbieder deze conformiteitsbeoordeling over het algemeen zelf kunnen uitvoeren. Slechts een beperkt aantal HRAIS is onder omstandigheden[152] verplicht onderhevig aan een conformiteitsbeoordeling door een derde partij. Dit zijn AI-systemen die een biometrische identificatie van personen van een afstand uitvoeren of, zoals opgenomen in de EP-tekst, gevolgtrekkingen maken op basis van biometrische (of op biometrie gebaseerde) data zoals emotieherkenning.[153] Daarnaast kan de aanbieder van mening zijn dat de aard, het ontwerp, de constructie of het doel van het AI-systeem een verificatie door een derde partij vereist.[154] Indien een aanbieder van een HRAIS heeft voldaan aan de door de EC aangenomen AI-standaarden wordt de aanbieder verondersteld aan de vereisten van hoofdstuk 2 (voorschriften voor HRAIS) te voldoen.[155] Dit zijn de verplichtingen die te maken hebben met het systeem voor risicobeheer en de risicobeheersingsmaatregelen, maar omvat niet de (volledige) eisen die gesteld worden aan het kwaliteitsmanagementsysteem (dat breder is dan het risicomanagement). De aanbieder stelt vervolgens een conformiteitsverklaring op, brengt een CE-markering aan en registreert het bijlage III-HRAIS.[156]

Gebruikers van HRAIS

Naast de aanbieder kunnen, zoals gezegd, de gebruikers van de AI-systemen worden geplaatst. Zij dienen technische en organisatorische maatregelen te nemen om te waarborgen dat zij HRAIS gebruiken in lijn met de gebruiksinstructies. Indien de gebruiker controle heeft over het HRAIS, worden additionele verplichtingen toegevoegd zoals ten aanzien van (competent) menselijk toezicht, robuustheid, data en cybersecurity.[157] Indien een werkgever een HRAIS in de werkomgeving wil inzetten, verplicht de EP-tekst de gebruiker om voor de implementatie daarvan de werknemersvertegenwoordigers te consulteren met het oog op het bereiken van een overeenkomst en de betrokken werknemers te informeren. Op grond van de formulering van deze verplichting is het echter mijns inziens niet duidelijk of het sluiten van de overeenkomst een voorwaarde is voor de inzet. Helderheid daarover lijkt mij echter wel van belang. Indien een HRAIS wordt ingezet als ondersteuning bij beslissingen over natuurlijke personen of deze beslissingen zelf neemt, dient de natuurlijk persoon hierover ingelicht te worden.

Belangrijkste nieuwe verplichting voor gebruikers in de EP-tekst is de verplichting tot het uitvoeren van een impactassessment gerelateerd aan fundamentele rechten.[158] Dit assessment is alleen verplicht als het gaat om bijlage III-HRAIS, met uitzondering van HRAIS die te maken hebben met kritische infrastructuur (waarbij het gaat om impact op klimaat en milieu). Deze nieuwe verplichting is geïntroduceerd omdat gebruikers beter in staat zouden zijn om de daadwerkelijke risico's van het gebruik te analyseren dan aanbieders in het risicoassessment in de ontwerp- en ontwikkelfase.[159] Dit is een uitgebreid impactassessment gecombineerd met de verplichting een plan op te stellen voor de mitigatie van geïdentificeerde risico's en een wat onduidelijke verplichting ten aanzien van het governancesysteem. De uitkomsten van een eventuele DPIA kunnen worden ingezet voor dit impactassessment.[160] Opvallend is dat de voorzienbare impact op fundamentele rechten (welke niet gelimiteerd tot is negatieve impact) en de voorzienbare negatieve impact op het klimaat en milieu als apart element van elkaar zijn opgenomen in de lijst van elementen waaraan het assessment moet voldoen. De verplichting tot het opstellen van een mitigatieplan is echter alleen gericht op de mitigatie van de negatieve impact op de fundamentele rechten. Niet helemaal duidelijk is of het de bedoeling is dat dit onderscheid wordt gemaakt of dat hieronder ook wordt verstaan de negatieve impact op het klimaat en milieu. Om het assessment uit te voeren dient de gebruiker (met uitzondering van mkb-bedrijven) zo goed als mogelijk is, stakeholders of vertegenwoordigers daarvan te betrekken.[161] Dit stakeholderengagement vormt dus als zodanig een verplicht onderdeel van het assessment.[162] Verder is mijns inziens de verplichting ten aanzien van het governancesysteem niet geheel helder. Deze laatste verplichting luidt: '*(…) This assessment shall include, at a minimum, the following elements: (…) (j) the governance system the deployer will put in place, including human oversight, complaint-handling and redress.*' Deze formulering is wat verwarrend omdat het artikel eisen stelt aan het impactassessment, terwijl deze subverplichting het in-

151 Overweging 46b EP-tekst.
152 Deze omstandigheden zijn: de aanbieder past de in het *PbEU* gepubliceerde standaarden of specificaties niet toe of slechts gedeeltelijk, de voornoemde standaarden bestaan nog niet en algemene specificatie zijn niet beschikbaar, de standaarden zijn in beperkte mate beschikbaar.
153 Artikel 43(1) Voorstel AI Act. Zie ook de EP-tekst.
154 Artikel 43(1)(d) EP-tekst.
155 Indien die standaarden de vereisten van hoofdstuk 2 omvatten, artikel 40 Voorstel AI Act.
156 Artikel 48 respectievelijk 49 en 51 Voorstel AI Act.
157 Artikel 29(1a) EP-tekst.
158 Artikel 29a EP-tekst.
159 Overweging 58a EP-tekst.
160 Artikel 29(6) EP-tekst.
161 Artikel 29a(4) EP-tekst.
162 De gedachte van de mkb-uitzondering is waarschijnlijk dat in de afweging tussen de extra last voor het mkb die stakeholderengagement met zich brengt en de extra bescherming die dit zou kunnen bieden, het stakeholderengagement dat de aanbieder (in tegenstelling tot de gebruiker) dient uit te voeren, als voldoende bescherming wordt beschouwd. Niet duidelijk is overigens op welke midden- en kleinbedrijven deze uitzondering van toepassing is, omdat er geen definitie in de EP-tekst is opgenomen. Daarnaast worden mkb en start-ups apart in de tekst opgenomen; hoe het een zich tot het ander verhoudt is ook niet duidelijk.

stellen van een bepaalde governance verlangt. Waarschijnlijk houdt de verplichting ten aanzien van het governancesysteem in dat op basis van de analyse van de impact (het impactassessment) een gepast governanceraamwerk moet worden geïdentificeerd dat vervolgens verplicht geïmplementeerd dient te worden. Maar helemaal duidelijk is dit niet.

Aanbieders van foundationmodellen, generatieve AI, generatieve-AI-systemen en general-purpose-AI-systemen

Op de aanbieder van een foundationmodel[163] rusten verplichtingen die onder andere te maken hebben met risicoidentificatie, -reductie en -mitigatie op het brede spectrum van risico's alsook documentatie, het betrekken van stakeholders en een uitgebreide registratie[164] van het model in de Europese database.[165] Deze verplichtingen zijn, zoals gezegd, op het laatste moment toegevoegd en lijken nog niet goed uitgewerkt. In het kader van risicoanalyse en -management dienen aanbieders van HRAIS een afweging uit te voeren ten aanzien van de aanvaardbaarheid van risico's. Deze verplichting is niet voor aanbieders van foundationmodellen opgenomen. Het is niet duidelijk waarom niet.[166] Ook deze aanbieders dienen mitigatiemaatregelen te nemen en dienen overblijvende risico's die niet te mitigeren zijn, te registreren.[167] Daarbij dienen zij te vermelden waarom de risico's niet te mitigeren zijn,[168] maar er is geen verplichting tot een aanvaardbaarheidstoets van deze overblijvende risico's. Overigens zal de identificatie alsook het uitvoeren van mitigatieactiviteiten geen eenvoudige opgave zijn, nu foundationmodellen het karakter hebben voor zeer uiteenlopende doeleinden te worden ingezet. Er wordt daarom ook wel gesteld dat de verplichting voor het instellen van een risicomanagementsysteem ertoe leidt dat Big Tech in de kaart wordt gespeeld omdat zij de middelen hebben om aan een dergelijke verplichting te voldoen in tegenstelling tot kleinere bedrijven. Hierdoor zou de Digitalemarktenverordening[169] worden ondermijnd.[170] Aanbieders van foundationmodellen dienen ook een kwaliteitsmanagementsysteem op te tuigen. Maar in tegenstelling tot het kwaliteitsmanagementsysteem van aanbieders van HRAIS waar zeer uitgebreide eisen aan worden gesteld, zijn de vereisten hier niet verder uitgewerkt. Sterker nog, er staat zelfs in het artikel dat aanbieders van foundationmodellen de mogelijkheid hebben om te experimenteren bij het voldoen aan dit vereiste. Anders dan dat foundationmodellen onderhevig zijn aan (snelle) ontwikkelingen, is er geen uitleg gegeven waarom deze vrijheid bij foundationmodellen bestaat. Dat vind ik zo opzichzelfstaand niet te begrijpen. Waarom zou bijvoorbeeld een 'accountability framework' niet verplicht moeten zijn bij een aanbieder van een foundationmodel, maar wel bij een aanbieder van een HRAIS? Het kan zijn dat de Big Tech-lobby een vinger in de pap heeft gehad bij het formuleren van deze regels. Aan de andere kant is er dan wel weer een zeer uitgebreide verplichting voor aanbieders van foundationmodellen als het gaat om de impact op het klimaat en milieu, die dan weer niet voor aanbieders van HRAIS bestaat (althans niet in de artikelen is opgenomen). Dat is voor foundationmodellen begrijpelijk omdat de aard daarvan een enorme computerkracht en daarmee energie vergt. Deze verplichting houdt in dat het AI-model zodanig moet worden ontworpen en ontwikkeld dat sprake is van een reductie van energieverbruik, bronnenverbruik en afval, en energie-efficiëntie en algehele efficiëntie van het systeem. De modellen moeten meet- en logcapaciteiten bezitten om energie- en bronnenverbruik te meten en te registreren en, indien technisch mogelijk, ook de impact van het gebruik van het systeem op het klimaat en milieu tijdens de gehele levenscyclus.[171]

Een andere belangrijke verplichting is dat aanbieders van foundationmodellen aan downstream-aanbieders van HRAIS voldoende technische documentatie en begrijpelijke instructies verschaffen, zodat zij kunnen voldoen aan hun verplichtingen onder de AI Act.[172] Op aanbieders van generatieve AI (dus aanbieders van foundationmodellen die (door andere partijen) worden gebruikt in AI-systemen die content genereren) en aanbieders die een foundationmodel in een generatieve-AI-systeem integreren zijn, zoals gezegd, speciale regels van toepassing, gericht op contentgeneratie en transparantie over gebruik van beschermde trainingsdata.[173] Uit een onderzoek door Stanford naar compliance van tien aanbieders van foundationmodellen met 12 belangrijke verplichtingen uit de EP-tekst blijkt dat aanbieders van foundationmodellen op dit moment nog aardig wat stappen moeten zetten om te voldoen aan de verplichtingen. Met name op het gebied

163 Ongeacht of dit model wordt aangeboden als een losstaand model, ingebed in een AI-systeem of een product, of onder vrije of open-sourcelicenties als een dienst of via andere distributiekanalen.

164 Dit is een uitgebreide registratie waarbij informatie moet worden verschaft over onder andere de capaciteiten, de bijbehorende risico's, genomen mitigatiemaatregelen, trainings-, performance-, testinformatie en optimalisatie van het model: bijlage VIII, Deel C EP-tekst.

165 Voor een mooi overzicht van de verplichtingen voor aanbieders van foundationmodellen onder de EP-tekst, zie: https://github.com/stanford-crfm/TransparencyIndex/blob/main/requirements.md.

166 '*For the purpose of paragraph 1, the provider of a foundation model shall: (a) demonstrate through appropriate design, testing and analysis the identification, the reduction and mitigation of reasonably foreseeable risks to health, safety, fundamental rights, the environment and democracy and the rule of law prior and throughout development with appropriate methods such as with the involvement of independent experts, as well as the documentation of non-mitigable risks after development.*'; artikel 28b(2)(a) EP-tekst.

167 Artikel 28b(2)(a) EP-tekst.

168 Bijlage VIII, Deel C, onder 6 EP-tekst.

169 Digitalemarktenverordening (Verordening (EU) 2022/1925, *PbEU* L 265/1, 12 oktober 2022).

170 P. Hacker (2023), 'Generative AI at the Crossroads', University of Oxford, faculty of law blogs, 12 juni 2023; https://blogs.law.ox.ac.uk/oblb/blog-post/2023/06/generative-ai-crossroads.

171 Artikel 28b(2)(d) EP-tekst.

172 Artikel 28b(2)(e) en artikel 60 jo. bijlage VIII, Deel C EP-tekst.

173 Artikel 28b(4)(a-c) en artikel 52(1) EP-tekst.

van risicomitigatie, performancemeting en -evaluatie, rapportage over energieverbruik en gebruik van beschermde data liggen uitdagingen.[174] Voor zover ik heb kunnen nagaan, bestaat er voor aanbieders van foundationmodellen onder de EP-tekst geen conformiteitsbeoordelingsverplichting zolang deze niet geïntegreerd in een HRAIS worden aangeboden, hoewel in artikel 40 wel is opgenomen dat indien foundationmodellen voldoen aan de op te stellen standaarden, zij geacht worden te voldoen aan artikel 28b dat de eisen stelt aan deze modellen.[175] In de overwegingen is opgenomen dat vanwege de aard van deze modellen expertise in conformiteitsbeoordelingen ontbreekt en methoden voor audits door derde partijen in ontwikkeling zijn. Als gevolg hiervan worden er veelsoortige interne assessmentactiviteiten ontwikkeld. Dat is wellicht de rede dat er nu geen verplichting voor een conformiteitsbeoordeling is opgenomen. De idee is dat de EC en de AI Office[176] periodiek het regelgevend kader voor foundationmodellen evalueren waardoor de verplichting wellicht later nog wordt ingevoerd.

De Raad heeft, zoals gezegd, gekozen voor één begrip: 'general purpose AI system'.[177] Specifieke regels gelden alleen als een general-purpose-AI-systeem als een HRAIS wordt gebruikt of als component van een HRAIS wordt ingezet. In die gevallen zouden volgens de Raad (sommige) regels die op HRAIS van toepassing zijn ook voor general purpose AI moeten gelden. Dit zou in een uitvoeringsbesluit moeten worden geregeld en in het kader daarvan zou een effectbeoordeling en een consultatie moeten plaatsvinden.[178] Ook zouden er uitvoeringsbesluiten moeten komen voor de samenwerkingsmodaliteiten tussen aanbieders van general purpose AI en andere aanbieders die deze systemen als HRAIS op de markt willen brengen, zodat deze laatste aan hun verplichtingen onder de AI Act kunnen voldoen.[179]

2.3.6 *Analyse ESG in het Voorstel AI Act en de EP-tekst*

Impact op klimaat en milieu

Door het gekozen classificatiesysteem worden AI-systemen die een significant risico vormen voor de milieu- en klimaatdoelstellingen, niet aangemerkt als AI-systemen die verboden zijn of die vanwege dit risico vallen in de HRAIS-categorie.[180] Het Voorstel AI Act heeft bij de vaststelling van de lijst van HRAIS met name gekeken naar de directe impact op mensen. Klimaat en milieu worden wel genoemd in de overwegingen van het Voorstel AI, maar zijn verder niet expliciet ingebed in de gekozen benadering van regels. Het Europees Parlement heeft dit proberen op te vangen door de best-effort-inspanningsverplichting ten aanzien van de beginselen voor ethische en betrouwbare AI. Hierbij wordt echter gesteld dat deze beginselen voor de verschillende soorten AI-systemen en -modellen uitgewerkt zijn in de artikelen die op hen van toepassing zijn.[181] Dit leidt tot een wat vreemde situatie. Aanbieders van HRAIS die als hoog risico zijn gekwalificeerd vanwege de mogelijke impact op mensen (en dus niet vanwege mogelijke impact op klimaat en milieu) dienen de 'E'-risico's mee te nemen in het risicoassessment en de risicomanagementactiviteiten. Indien er een negatieve (potentiële) 'E'-impact is geïdentificeerd in deze beoordeling, dan dient deze te worden voorkomen of gemitigeerd en dienen maatregelen te worden genomen.[182] Daarnaast bestaat er voor HRAIS een verplichting tot het geven van informatie over het energieverbruik gedurende de ontwikkeling en het te verwachten energieverbruik tijdens het gebruik. Deze activiteiten die gunstige 'E'-effecten hebben, lijken echter meer op een soort bijvangst. Daar komt bovenop dat gebruikers van HRAIS (in aanvulling op de verplichtingen die rusten op aanbieders) ook de negatieve impact van het HRAIS dienen te analyseren en hiervoor mogelijk een mitigatieplan dienen op te stellen voordat het HRAIS op de markt kan worden gebracht of in gebruik kan worden gesteld. Daardoor wordt de bijvangst zelfs nog groter. Dit betekent tegelijkertijd dat AI-systemen die (mogelijk) een grote impact hebben op milieu- en klimaat en die niet als HRAIS of foundationmodel kwalificeren, in de categorie van zelfregulering vallen. Standaarden en gedragscodes zijn niet verplicht. Voor deze AI-systemen betekent dit dat de wijze waarop aanbieders en gebruikers omgaan met mogelijke negatieve 'E'-effecten via de uitwerking van de best-effort-verplichting volledig afhankelijk is van zelfre-

174 R. Bommasani et al. (2023), 'Grading Foundation Model Providers' Compliance with the Draft EU AI Act', 15 juni 2023; https://crfm.stanford.edu/2023/06/15/eu-ai-act.html#fn:1.

175 Artikel 40(1) EP-tekst.

176 De AI Office is een door het EP in de EP-tekst geïntroduceerd EU-orgaan dat de geharmoniseerde toepassing van de AI Act dient te bevorderen door onder andere het publiceren van opinies, aanbevelingen en richtlijnen en het ondersteunen, het adviseren van en samenwerken met de Europese Commissie, de lidstaten, de toezichthoudende autoriteiten en andere Europese instituties. Hierbij worden verschillende stakeholders betrokken via een adviesforum. De AI Office heeft tevens de rol als mediator in discussies tussen de relevante autoriteiten in het kader van de toepassing van de AI Act. Het takenpakket bevat nog vele andere taken; overweging 76 en artikel 56a en 56b EP-tekst.

177 Zie noot 113.

178 Artikel 41 Compromistekst van de Raad.

179 Compromistekst van de Raad, paragraaf 3.1 en 3.2, p. 6, alsook overweging 12c en artikel 4b(5) Compromistekst van de Raad.

180 Bijlage III Voorstel AI Act.

181 Voor HRAIS; artikel 8 t/m 15 van hoofdstuk 2 en die van hoofdstuk 3 AI Act. Voor foundationmodellen zou dit in artikel 28 t/m 28b het geval zijn en voor alle AI-systemen zou compliance met deze beginselen via het toepassen van artikel 28, artikel 52, geharmoniseerde standaarden, technische specificaties en gedragscodes verlopen. Gedragscodes zoals opgenomen in artikel 69 AI Act.

182 Tevens is voor deze categorie AI-systemen ernstige schade voor het milieu geclassificeerd als een 'ernstig incident'. Artikel 3(44) Voorstel AI Act: AI-verordening: definitie van 'ernstig incident': '*elk incident dat direct of indirect leidt, kan hebben geleid of kan leiden tot: (a) het overlijden van een persoon of ernstige schade voor de gezondheid van een persoon, eigendom of het milieu (…).*'

gulering.[183] Het voorstel bespreekt dat ondernemingen in het kader van vrijwillige gedragscodes inzake het ontwerp, de ontwikkeling en het gebruik van AI-systemen rekening kunnen houden met ESG-aspecten. Het voorstel stelt: *'In deze codes kunnen ook vrijwillige verbintenissen zijn opgenomen met betrekking tot bijvoorbeeld milieuduurzaamheid, toegankelijkheid voor personen met een handicap, inspraak van belanghebbenden tijdens het ontwerp en de ontwikkeling van AI-systemen, en diversiteit binnen de ontwikkelingsteams.'*[184] Hoewel ik begrijp dat AI-systemen die een (grote) negatieve impact hebben op milieu en klimaat lastig zijn in te passen in het gekozen systeem voor de kwalificatie van HRAIS (die kwalificatie is gericht op het doel en de functionaliteit van het AI-systeem, terwijl bij AI-systemen die een impact hebben op milieu en klimaat het doel en de functionaliteit niet van doorslaggevende betekenis zijn voor het risico dat het systeem met zich brengt), vallen deze nu enigszins buiten de boot. Dat lijkt mij niet wenselijk. Voor foundationmodellen gelden 'E'-verplichtingen zoals in paragraaf 2.3.5 uiteengezet. Deze lijken verder te gaan dan de verplichtingen voor aanbieders van HRAIS, omdat hierbij voor alle foundationmodellen een expliciete verplichting geldt om de modellen zo te ontwerpen en te ontwikkelen dat zij energieverbruik, bronnenverbruik en afval reduceren en de energie-efficiëntie en efficiëntie van het hele systeem verhogen.

De vraag rijst overigens hoe haalbaar het is om de brede 'E'-impact van HRAIS en foundationmodellen vanaf het moment van inwerkingtreding van de wetgeving te berekenen, te communiceren en te reduceren. Voor de foundationmodellen is derhalve bepaald dat de verplichting het energieverbruik, bronnenverbruik en afval te reduceren en de efficiëntie te verhogen pas geldt als de Europese standaarden beschikbaar zijn. Maar ook het meten, registreren en publiceren van de 'E'-impact bij een HRAIS is nog best een uitdaging. Dit is een gebied dat volop in ontwikkeling is, zoals uiteengezet in het eerste deel van dit drieluik. Mogelijkerwijs zou hier een ingroeiperiode uitkomst bieden, zodat er geen gat ontstaat tussen het moment dat de rapportageverplichting ingaat en de voornoemde methodologie voorhanden is. Of de bepalingen en overwegingen gerelateerd aan het klimaat en milieu zoals deze nu zijn opgesteld, de trialoog overleven is ook nog maar de vraag. Het zou echter vreemd zijn als de finale tekst van de AI Act niet meer aandacht zal besteden aan de E, dan het Voorstel AI Act doet. Daarvoor is mijns inziens nu in het Voorstel AI Act te weinig aandacht. De benadering in de EP-tekst lijkt daarnaast een belangrijke groep AI-systemen in het licht van mogelijke 'E'-impact over het hoofd te zien.

Impact op personen

Het Voorstel AI Act en de EP-tekst bevatten allerlei waarborgen om risico's voor personen te voorkomen of te mitigeren. Dit wordt echter alleen gerealiseerd door een beperkt aantal AI-systemen te verbieden en een beperkte groep AI-systemen te onderwerpen aan de HRAIS-regels. Dat is ook de reden waarom wordt benadrukt dat deze lijst van HRAIS niet in steen gebeiteld is en dat er een continue evaluatie moet zijn of de lijst aangevuld dient te worden met nieuwe categorieën AI-systemen. Ook het feit dat de aanbieders van HRAIS zelf degenen zijn die bepalen of er sprake is van conformiteit met de AI Act stuit op kritiek. Zo zijn er bijvoorbeeld partijen die van mening zijn dat alle conformiteitsbeoordelingen van HRAIS door derde partijen moeten worden uitgevoerd.[185] Over het algemeen zijn er door de gelaagde structuur van impactassessments in de EP-tekst met assessments door aanbieders en daarnaast gebruikers bij HRAIS verschillende checks ingebouwd alsook verplichtingen om te handelen als er negatieve impact op personen uitkomt. Deze impactassessments op grond van de AI Act worden in sommige gevallen ook nog aangevuld met de DPIA's op grond van de AVG. Ook in het kader van foundationmodellen zijn belangrijke verplichtingen opgenomen die impact op mensen trachten te voorkomen of verminderen. Op het gebied van de 'S'-pijler zal de voorgestelde wetgeving een belangrijke stap in de goede richting zijn. Echter, ook hier geldt dat de in de EP-tekst opgenomen algemene verplichting om verantwoord te handelen op grond van de beginselen die van toepassing zijn op alle AI-systemen, voor de niet-HRAIS alleen handen en voeten krijgt als er sprake is van een zelf opgelegde gedragscode.

Governancevereisten

Verder valt op dat de governancevereisten voor aanbieders van HRAIS en voor aanbieders van foundationmodellen verschillend zijn. Het blijft een beetje gissen naar de achterliggende gedachte hiervan. Het kan te maken hebben met de haast waarmee deze nieuwe regels zijn opgesteld, maar dat hoeft niet. Het voorstel bevat daarnaast geen verwijzingen naar governancepraktijken die nu steeds meer gebruikelijk worden, zoals het instellen van een ethische of technologische commissie als losstaande commissie of als onderdeel van de raad van commissarissen of bijvoorbeeld het aanstellen van een bestuurslid met specifieke kennis op het gebied van technologie. Ook mis ik een link tussen de rol die bedrijfscultuur kan spelen in het komen tot ethische en betrouwbare en het liefst ook duurzame AI. Het implementeren van deze governance-aspecten in de onderneming zou onder omstandigheden gezien kunnen worden als een belangrijk onderdeel van een adequaat kwaliteitsmanagementsysteem. Het ontbreken hiervan in het Voorstel AI Act komt waarschijnlijk doordat deze wetgeving gericht is op productveiligheid en in het voorstel alleen de activiteiten in het kader van risicomanagement zijn uitge-

183 Artikel 28 (dit artikel betreft de verantwoordelijkheden in de waardeketen en is van toepassing op HRAIS en niet op niet-HRAIS) en artikel 52 (dit artikel bevat transparantieverplichtingen voor AI-systemen met een beperkt risico (chatbots, deepfakes etc.) hebben waarschijnlijk betrekking op de andere beginselen en niet op het beginsel van social and environmental well-being.

184 Paragraaf 5.2.7 Voorstel AI Act, p. 18.

185 EDPB (2021), *supra* noot 48, p. 2 en 13.

werkt, en niet zozeer de structuur waarin deze activiteiten plaatsvinden. Daarnaast zullen er ook nog standaarden worden opgesteld waarin dit mogelijk wel een plek zou kunnen krijgen.[186] Toch hadden deze governance-aspecten mijns inziens in de overwegingen van de verordening zelf genoemd kunnen worden en daarnaast in het artikel dat gaat over zelfregulering, waarin bijvoorbeeld wel de diversiteit van ontwikkelteams wordt genoemd. Maar dat is niet het geval. Bij de duurzaamheidswetgeving komt de koppeling tussen het bedrijfsmodel en de strategie en het reduceren van impact veel sterker terug. Niet alleen dient hierover gerapporteerd te worden onder de Corporate Sustainability Reporting Directive ('CSRD'),[187] in de voorgestelde wetgeving voor de Corporate Sustainability Due Diligence Directive ('CSDDD')[188] is ook een verplichting voor lidstaten opgenomen te zorgen dat bestuurders stappen ondernemen om de bedrijfsstrategie aan te passen teneinde rekening te houden met de feitelijke en potentiële negatieve effecten die in het impactassessment zijn geïdentificeerd.[189] Ondernemingen dienen passende zorgvuldigheid te integreren in hun beleid en feitelijke of potentiële negatieve effecten te identificeren, te reduceren, te voorkomen, te beëindigen en te beperken en in het kader daarvan maatregelen te nemen.[190] En daarnaast dienen ondernemingen een transitieplan op te stellen dat verenigbaar is met de overgang naar een duurzame economie en de beperking van de opwarming van de aarde tot 1,5 °C. Deze voorgestelde wetgeving is echter slechts op ondernemingen van een bepaalde grootte van toepassing en met name Europese ondernemingen.[191] In het kader van AI-systemen kan de onderneming klein zijn maar de impact (zeer) groot, onder andere omdat het bereik van technologie via de netwerksamenleving snel wijdverbreid kan zijn. De CSRD en CSDDD bespreek ik in het laatste deel van dit drieluik.

186 In het kader van de DPIA onder de AVG is in de richtsnoeren bijvoorbeeld ook opgenomen wat de rol van de CSIO (chief security information officer) kan inhouden; Groep Gegevensbescherming artikel 29 (2017), *supra* noot 64, p. 19.

187 Corporate Sustainability Reporting Directive (Richtlijn (EU) 2022/2464, *PbEU* L 322/15, 16 december 2022 (hierna: 'CSRD').

188 European Commission (2022), 'Proposal for a Directive of the European Parliament and of the Council on Corporate Sustainability Due Diligence and amending Directive (EU) 2019/937', COM(2022) 71 final, 23 februari 2022 (hierna: 'Voorstel CSDDD').

189 Artikel 26 Voorstel CSDDD. Dit artikel is echter in de EP-tekst CSDDD weer verwijderd, waarschijnlijk omdat dit specifiek gericht was op bestuurders en in artikel 4 e.v. in het kader hiervan allerlei verplichtingen opgenomen zijn die rusten op ondernemingen.

190 Artikel 4 t/m 8 Voorstel CSDDD.

191 Momenteel bestaat er discussie over de reikwijdte van de CSDDD. In het Voorstel CSDDD van de Europese Commissie gaat het met name om zeer grote Europese bedrijven (> 500 werknemers en een wereldwijde nettoomzet van meer dan 150 miljoen euro) of bepaalde bedrijven die in sectoren met een bijzonder hoge impact opereren (> 250 werknemers en een wereldwijde netto-omzet van meer dan 40 miljoen euro, waarvan minimaal 50% in een high-impactsector werd behaald); artikel 2(1)(a) en (b) Voorstel CSDDD. In de EP-tekst is de reikwijdte vergroot door de drempels te verlagen naar Europese ondernemingen met meer dan 250 werknemers en een wereldwijde netto-omzet van meer dan 40 miljoen euro. Het voorgestelde toepassingsbereik van de CSDDD is echter groter bijvoorbeeld bepaalde non-EU-ondernemingen en groepsholdings, maar dat laat ik hier buiten beschouwing.

3. Tot slot

Technologie kan een vriend en tegelijkertijd een vijand zijn. De uitdaging is om de vriendschap met technologie te verstevigen en de mate van vijandigheid te elimineren of zoveel als mogelijk te reduceren om de ons dierbare aspecten van het menselijk bestaan te beschermen. Dit is in een globaliserende, polariserende en individualistische maatschappij geen sinecure, doordat opvattingen over wat beschermingswaardig is alsook over de manier waarop bescherming vormgegeven dient te worden, van elkaar verschillen. Daar komt bij dat de verschuivende macht van staten naar een handjevol (Amerikaanse) ondernemingen die de technologische infrastructuur en de (generatieve-) AI-modellen in handen hebben niet bijdraagt aan een gelegitimeerde invulling van wat de mensen in Europa belangrijk vinden. Dat is zorgelijk. Europese digitaliseringswet- en regelgeving zou een reflectie moeten zijn van de in Europa heersende opvatting over hoe negatieve impact van technologie aan banden moet worden gelegd en de wijze waarop innovatie kan worden ondersteund. Beide elementen zijn verbonden met ESG. Technologie kan zowel een (potentiële) negatieve impact als een (potentiële) positieve impact hebben op zowel de 'E'- als de 'S'-pijler.

De hoeksteen-wetgeving op het gebied van data, de AVG, lijkt echter niet goed bestand te zijn tegen technologische vooruitgang. De huidige regels maken het toenemende gebruik van data voor ESG-doelstellingen ingewikkeld of zijn in sommige gevallen zelfs prohibitief. Het Voorstel voor de AI Act probeert dat voor een deel op te lossen door AI-systemen die een positieve bijdrage (kunnen) leveren aan de klimaat- en milieuproblematiek een helpende hand te bieden door middel van *regulatory sandboxes*. De regels die nu voor deze testomgevingen worden voorgesteld, matchen niet goed met de AVG, en het is niet duidelijk of de ondersteunde regels ook kunnen worden gebruikt na de testperiode. Initiatieven op het gebied van digitaliseringswet- en regelgeving hebben gelukkig steeds meer aandacht voor klimaat en milieu. Dit wordt onder andere duidelijk in de Europese verklaring over digitale rechten en plichten, en sijpelt langzaam door in voorgestelde wetgeving, zoals die ten aanzien van de Europese AI Act. Daar waar het Voorstel AI Act nog met name toegespitst is op bescherming van mensen, tracht de EP-tekst het brede palet van ESG, dus ook klimaat en milieu, beter in te bedden. Dat kan ook niet anders nu de Europese Commissie, het Europees Parlement en de Raad geconfronteerd worden met de rasse schreden waarmee de ontwikkelingen doordenderen. Niet alleen krijgen klimaat en milieu in toenemende mate een prominentere rol in het raamwerk van mensenrechten, de technologische ontwikkelingen doen de vraag naar energie, water en grondstoffen alleen maar toenemen, en dat is onwenselijk. Het op grote schaal inzetten van LLM's gaat gepaard met een zeer groot beroep op beschikbare (energie)bronnen. Het opnemen van regels gericht op ESG in de digitaliseringswetgeving zelf, zoals

het impactassessment ten aanzien van grondrechten, de verplichte transparantie over energieverbruik in de verschillende stadia van de AI-levenscyclus, de verplichting om bij het ontwerp en de ontwikkeling rekening te houden met de footprint en de introductie van testomgevingen met het oog op groene AI-innovaties, zijn belangrijke stappen in het proces om tot een goede AI Act te komen. De vormgeving van de regels laat mijns inziens echter nog wel te wensen over. Een belangrijk deel van AI-toepassingen blijft buiten schot en duurzaamheidsinitiatieven uit het verleden laten duidelijk zien dat zelfregulering niet het gewenste resultaat heeft opgeleverd. Ook zou er meer oog moeten zijn voor betekenisvol stakeholderengagement in de AI Act. In Deel III bespreek ik of recente duurzaamheidswetgeving dit mogelijk opvangt.