



## UvA-DARE (Digital Academic Repository)

### Artificiële Intelligentie: waar is de werkelijkheid gebleven?

Dommering, E.

**Publication date**

2023

**Document Version**

Final published version

**Published in**

Computerrecht

[Link to publication](#)

**Citation for published version (APA):**

Dommering, E. (2023). Artificiële Intelligentie: waar is de werkelijkheid gebleven? *Computerrecht*, 2023(6), 476-483. Article 258. <https://www.ivir.nl/publicaties/download/AI-Computerrecht-2023.pdf>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# Artificiële Intelligentie: waar is de werkelijkheid gebleven?

*Computerrecht 2023/258*

**Er is veel ophef ontstaan over de (te) snelle toepassing van AI in de samenleving. Dit artikel onderzoekt wat AI (in het bijzonder ChatGPT) is. Vervolgens laat het zien waar de invoering van AI al direct wringt in de gebieden van het auteursrecht, de privacy, vrijheid van meningsuiting, openbare besluitvorming en mededingingsrecht. Daarna wordt stilgestaan bij de vraag of de AI-verordening van de EU daar het antwoord op zal zijn. De conclusie is dat dat maar zeer ten dele zo is. Bescherming zal dus moeten komen van normen uit de deelgebieden. Het artikel formuleert tot slot vier beginselen die in ieder deelgebied een AI 'metakader' kunnen vormen waarmee een AI-product moet worden beoordeeld.**

## 1. Inleiding

Op dinsdag 13 juni 2023 werd aan de Tweede Kamer een petitie aangeboden en toegelicht in *de Volkskrant* van die datum, door initiatiefnemer Mark Thiessen (van de denktank 'De Nieuwe Vrije Eeuw') en Kees Verhoeven (oud-D66-kamerlid en oprichter van 'Bureau Digitale Zaken') onder de wanhopige kop: 'Politiek, breng kunstmatige intelligentie (AI) *nog dit jaar* onder menselijke controle'. De petitie werd al direct ondertekend door een kleine 600 'bekende' Nederlanders. Ook uit de VN (de Internationale Telecommunicatie Unie) kwam er een alarmsignaal.<sup>2</sup>

Wat is er aan de hand, wat is de nieuwe 'AI' die *nog dit jaar* onder controle gebracht moet worden? In het navolgende onderzoek ik wat de ontwikkeling in AI is waar nu acuut nieuwe regulering voor moet komen. Het is een groot onderwerp dat feitelijk en normatief sterk in beweging is, en daarom is het een eerste verkenning.

Ik ga eerst in op wat AI is en in het bijzonder de chat-bot AI waar zoveel ophef over is ontstaan (paragraaf 2). Vervolgens kijk ik hoe het bestaande recht daarop reageert (paragraaf 3). Daarna sta ik stil bij de Europese AI-verordening (paragraaf 4). Ik rond af met de vraag wat een integrale juridische benadering van deze revolutionaire AI-ontwikkeling zou moeten zijn (paragraaf 5). Ik pleit voor een effectief kader van beginselen waarmee AI-toepassingen in ieder rechtsgebied kunnen worden benaderd.

## 2. Wat is AI?

Artificiële intelligentie (hierna: AI) is zo oud als de computer. Dat is een volgens algoritmen (computerprogramma's) data-verwerkende machine. Door die combinatie 'leert' hij van de eigenschappen van de gebruiker (selectie van algoritmen) en zijn omgeving (data) en past hij zich daaraan aan. Maar het is meer omdat er door een combinatie van ingevoerde grote hoeveelheden specifieke data en geprogrammeerde leerprocessen de machine zelf problemen op specifieke gebieden kan oplossen en daardoor 'intelligent' gedrag gaat vertonen. Een bekend voorbeeld uit het nabije verleden is de schaakcomputer, die een schaakgrootmeester kan verslaan, omdat hij sneller meer data kan verwerken en meer spelsituaties kan overzien. Bij iedere nieuwe AI-toepassing wordt de rode vlag gehezen om de samenleving te redden van een onheil dat zich als een virus verspreidt. Daarbij speelt een bijgeloof een rol: de denkende mens staat in onze natuur boven alles maar dreigt te worden overgenomen door een robot. Er vinden vanaf het begin van de AI-ontwikkeling filosofische discussies plaats of de computer die 'dom' instructies uitvoert wel 'zelfbewustzijn' heeft en 'intelligent' is. Bekend is het 'Chinese Room' gedachte-experiment van de Amerikaanse filosoof John Searle: een computer die een Chinese tekst vertaalt, voert 'dom' (foutloos) instructies uit, produceert een Chinese tekst, maar 'kent' geen Chinees. Het gedachte-experiment dateert van 1980 toen de 'cognitive science' in opmars was; maar het is nooit meer uit de discussie weggeweest.<sup>3</sup>

Er zijn verschillende vormen van AI-toepassingen en technieken. Ik sta stil bij de nieuwste ontwikkeling die de oorzaak is van de publieke opwindings.

### 2.1 Generatieve taal AI ('Large language models')<sup>4</sup>

ChatGPT is een toepassing van een lerend AI-taalsysteem dat het mogelijk maakt dat een computer op basis van ingevoerde taaldata en een gecontroleerd 'leerproces' (algoritmen) vragen die aan hem worden gesteld, herkent en

<sup>3</sup> John Preston & Mark Bishop (eds.), *Views into the Chinese Room. New Essays on Searle and Artificial Intelligence*, Oxford: Clarendon Press 2002. De discussie gaat terug op de uitvinding van de computer door Turing in 1936 die deze karakteriseerde als een 'symbolen verwerkende' machine, maar meende dat de output van dit proces niet van tevoren kon worden voorspeld en daarom als 'intelligent' kon worden gekarakteriseerd. Dit heet in de literatuur de 'Turing machine' (John Preston in de inleiding a.w., p. 7 e.v.). Voor een historisch overzicht van de ontwikkeling van wat is gaan heten 'cognitive science', zie Howard Gardner, *The New Science, A History of the Cognitive Revolution*, New York, Basic Books 1985.

<sup>4</sup> Ik beperk mij hierna vooral tot dit recente fenomeen. Voor een helder overzicht van de wording van AI en haar toepassingen verwijs ik naar: Yorrick Wilks, *Artificial Intelligence, Modern Magic or Dangerous Future*, Londen: Icon Books 2023.

<sup>1</sup> Egbert Dommering is Emeritus-hoogleraar Informatierecht aan de Universiteit van Amsterdam en voormalig redacteur van *Computerrecht*.

<sup>2</sup> *United Nation Activities on Artificial Intelligence (AI) 2022*, ITU Publications 2023.

kan beantwoorden. Alle AI-paniek richt zich op de 'output' van AI, dat wat uit de computer komt: machtsmisbruik (verborgen camera's die aan 'gezichtsherkenning' doen) en discriminatoire beslissingen. Van even groot, zo niet groter, belang is de 'input', dat wat er in de computer gaat: data en instructies (algoritmen). Ik sta daarom eerst stil bij de input.

## 2.1.1 *De input (1): Alles wordt 'data'*

De opwindning over AI is in 2023 ontstaan doordat ChatGPT door het bedrijf OpenAI eind 2022 op de markt werd gebracht.<sup>5</sup> De website van het bedrijf informeert ons dat het een AI and Research Company is en dat het zijn missie is 'to ensure that artificieel general intelligence benefits all of humanity'.<sup>6</sup> Het bedrijf streeft dit doel na met ChatGPT die getraind is om met de gebruiker in gesprek te gaan. Deze dialoog maakt het mogelijk 'to answer follow up questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests'. In deze omschrijving valt al meteen iets op: In tegenstelling tot de hondse zeden en gewoonten van de gebruikers van de sociale media, presenteert ChatGPT zich als een bijzonder beleefd wezen, net zoals vroeger de djinn die door op een lamp te wrijven werd opgeroepen om problemen op te lossen in het Duizend-en-één-nacht sprookje 'Aladin en de wonderlamp'. Zo vreemd is die beleefdheid niet. Machines zijn als bedienden: zij volgen de instructies van de bestuurder op. Als een machine deze gehoorzame vriendelijkheid niet vertoont, krijgt zij meteen een monteur op haar dak.

Maar let even op: de Chat-machine beantwoordt vragen; dat is iets anders dan een bewijsvoering of iets waar of onwaar is. "Waarheid is een bijproduct."<sup>7</sup>

De kennis en intelligentie van ChatGPT is het product van de turbulente geschiedenis van het internet, en dan met name het World Wide Web. Deze toepassing heeft immers het 18<sup>e</sup> eeuwse encyclopedie-ideaal verwezenlijkt door alle op de aarde gepubliceerde kennis (webpagina's), uit welke bron dan ook, toegankelijk te maken en met elkaar te verbinden.<sup>8</sup> De steeds 'intelligenter' zoekmachine (een voorloper van de chatbot) deed de rest.

Om het kenniswonder ChatGPT mogelijk te maken is allereerst nodig dat het 'taalmodel' (dat teruggaat op het model uit 1948 van Claude Shannon, de man die de 'bits' heeft bedacht) waaruit deze put zo groot mogelijk is. De kans dat hij bij een open vraag het antwoord herkent, wordt daardoor immers groter. De ontwikkeling van ChatGPT is in 2018 begonnen, en in vier jaar tijd is het taalmodel waarmee wordt gewerkt duizenden malen groter geworden.

Deze enorme berg taal (en bijbehorend beeld) moet worden bewerkt, vervolgens in nullen en enen opgeslagen in het computergeheugen, dan afgebroken in kleinere eenheden en dan door de computer zo te trainen dat deze bij een vraag, door vergelijking met de ingevoerde data vrij nauwkeurige voorspellingen over het juiste antwoord kan geven. Dat bewerkingsproces heeft de laatste twee jaar – door de enorme schaalvergroting – grote stappen gemaakt.

Heeft de Geest nu de Werkelijkheid overwonnen dankzij het internet? Helaas is dat niet zo. We zijn veeleer getuige van een moorddadig datificeringsproces, zoals Kate Crawford dat in haar boek *Atlas of AI* onder woorden brengt.<sup>9</sup> De werkelijkheid wordt omgezet in een massa neutrale bits:

"The data are not seen as carrying meaning or ethical weight as images of individual people. The personal, the social and the political meanings are all neutralised. It represents the shift from image to infrastructure, where the meaning or care that might be given to the image of an individual person or context behind a scene, is presumed to be erased at the moment it becomes part of an aggregate mass that will drive a broader system."

Crawford wijst er terecht op dat AI niet alleen een wereldomspannende database is, maar ook een verdeling van markten en infrastructuren tussen wereldbedrijven zoals Microsoft (de grootste aandeelhouder van Open AI), Meta, Adobe, Google, en Open AI (allen met eigen generatieve systemen onder andere merknamen). China (en TikTok) rukt op met een eigen data-infrastructuur.

Dit is een begin van een nauwelijks gereguleerde innovatie in een bestaande technologie. De sprong die binnen Google met het vervolmaken van het vertaalsysteem is gemaakt is een essentieel moment in deze ontwikkeling geweest.<sup>10</sup>

Kijken we naar de geschiedenis van het internet, dan zien we een begin van kleinschalige diensten georganiseerd rond een vrijheidsfilosofie, die is overgenomen door de grote commercieel gerunde sociale media. AI begint vanuit de commerciële kant heel grootschalig. Met de groei van de technologie zullen, in een optimistisch scenario, kleinschaliger en gespecialiserdere toepassingen doorzetten, maar, in een pessimistisch scenario, de machtsconcentraties niet meer zijn te verslaan.<sup>11</sup> Gespecialiseerde generatieve taal AI zal in het optimistische scenario een soortgelijke functie als de snelle rekenmachines van de huidige computertechnologie kunnen gaan vervullen. Op deelgebieden zijn al spectaculaire resultaten geboekt, zo-

5 29 november 2022 in San Francisco. Het werd meteen een hype. De winst van de onderneming wordt dit jaar geschat op \$ 200 miljoen; bron *Financial Times* 17 juni 2023.

6 <https://openai.com/blog/chatgpt>, geraadpleegd juli 2023.

7 Citaat uit de door mij bijgewoonde lezing van 12 juni 2023 van Jelle Zuidema, specialist Neuro linguïstisch programmeren aan de Faculteit Wis- en Natuurkunde van de UvA.

8 Ik verwijs voor de geschiedenis naar Egbert Dommering, *De Europese Informatierechtsorde*, Amsterdam: deLex 2019, hoofdstuk II, paragraaf 1 en 2.

9 Kate Crawford, *Atlas of AI*, New Haven and London: Yale University Press 2021, p. 18-21.

10 'Transformers: the Google scientists who pioneered an AI revolution', *Financial Times* 23 juli 2023.

11 Ruchir Sharma, voorzitter van Rockefeller International, 'What's wrong with the giants riding the AI wave?', *Financial Times* 31 juli 2013.

als de enorme kwalitatieve verbetering die vertaalmachines in de laatste twee jaar hebben gemaakt laat zien.

Een ander positief scenario is de impact op de arbeidsprocessen. Koplopers zijn de zakelijke dienstverlening (juridisch, fiscaal, accountancy) en de programmering van software. Een algemeen effect is verbetering van arbeidsomstandigheden. Dat loopt van door AI ondersteunde intelligente menselijke arbeid tot vervanging van monotone menselijke arbeid. Bij dat laatste kun je denken aan vervanging van het digitale lopendebandwerk (het internetproletariaat dat allerlei slechtbetaalde monotone arbeid verricht om de diensten die op of door internet aangeboden worden draaiende te houden).<sup>12</sup>

### 2.1.2 *De input (2): Categorisering*

Een onvermijdelijk proces bij het verzamelen van data is dat ze in een menselijk systeem moeten passen. In de wetenschap is dat een conceptuele ordening gebaseerd op een voldoende getoetste en aanvaarde theorie. Maatschappelijk gezien ordenen bedrijven en openbare instellingen de kenmerken van hun 'klanten' om producten en diensten aan te passen aan de vraag: waar wonen ze (postcode), wat is hun opleidingsniveau, wat is hun leeftijd en geslacht, wat is hun nationaliteit, wat zijn hun voorkeuren? De overheid is aangesteld om goederen en diensten te verdelen en gedrag van burgers in positieve (subsidies) of negatieve (criminaliteit) zin te beïnvloeden; zij ontwerpt daarom voortdurend categorieën om behoeften en eigenschappen van de onderdanen vast te stellen en de toepassing van de categorieën met gezag af te dwingen. Dit is de maatschappelijke en historische structuur van de samenleving zoals Geoffrey Bowker en Susan Leigh Star in hun studie uit 2000 laten zien:

“Formal classification systems are, in part, an attempt to regularize the movement of information from one context to another; to provide a means of access to information across time and space (...) to move information across the globe, over decades, and across multiple conflicting belief and practice systems.”<sup>13</sup>

Maar dit is ook het gebied dat wordt aangeduid met de 'algoritmes' die verkeerd of discriminatoir kunnen uitpakken, met name als het gaat om data met conflicterende waarden in categoriseringssystemen. Waar het bij de eerste input gaat om alle data van hun positieve of negatieve maatschappelijke en culturele waarde te ontdoen, gaat het bij categorisering juist om toekenning van positieve of negatieve waarde aan die data aan de hand van categoriseringssystemen.

Een recente formulering van dit aspect van ordening door algoritmes is die van Louise Amoore:

<sup>12</sup> Zie mijn column 'Internet proletariërs aller landen verenigt u', *Mediaforum* 2023, nr. 3.

<sup>13</sup> Geoffrey Bowker & Susan Leigh Star, *Sorting things out: Classification and Its Consequences*, Cambridge/Massachusetts: The MIT Press 2000, hoofdstuk 9.

“The very essence of algorithms is that they afford greater degrees of recognition and value to some features of a scene than they do to others. In so doing, algorithms generate themselves as ‘ethicopolitical’ beings in the world. (...) ‘I have studied the algorithm not as a finite series of programmable steps but as perennially adjustable and modifiable in relation to a target output.’”<sup>14</sup>

En met die laatste zinsnede raken we dus de kern: het algoritme is in deze benadering niet een instructie om de machine beter te laten werken (een betere machinecode), maar een instrument om door middel van een machine een menselijke doelstelling te verwezenlijken (een betere maatschappelijke code).

## 3. **Regulering van AI: wat is de stand van het recht?**

### 3.1 *Het algemene recht*

Er zijn rechtsgebieden waarin deze vorm van AI al direct 'neerdaalt'. Ik beperk mij tot enkele hoofdlijnen. Het raakt zowel de input als de output.

#### 3.1.1 *Auteursrecht*

We zien op het gebied van het auteursrecht al een beweging ontstaan die we bij de zoekmachines hebben gezien. Auteursrechthebbenden bestreden daar dat de informatie die de zoekmachines genereren (bijvoorbeeld kranten) vrij zou zijn en zij stelden dat voor dit hergebruik moet worden betaald. Datzelfde standpunt nemen de grote media (zoals News Corp, Axel Springer, The New York Times, The Guardian) nu ook in ten aanzien van AI-gegenereerde informatie, met name tegenover Open AI en Google. Dat gaat niet alleen over de invoer van de door dezen geproduceerde data zelf, maar ook over het gebruik van nieuwsberichten nodig voor de AI-herkenningsprocestraining. Zoals een van de woordvoerders voor de informatie-industrie opmerkte: Het gebruik van deze informatie is essentieel in het trainingsproces, omdat: ‘Chatbots won't be credible tools if they are literally trained primarily on the sewers of misogyny and racism that make up most of open accessible text.’<sup>15</sup>

De producenten van chatbots trachten het dilemma van het vinden van kwalitatief goede informatie voor training en output te vermijden door daarvoor informatie te gebruiken waarop zij het auteursrecht hebben of waarop het auteursrecht al is verlopen. Maar ze zullen ook deals moeten sluiten. Er is een voorzichtig optimisme dat er deals kunnen worden gemaakt omdat ook de uitgevers begrijp-

<sup>14</sup> Louise Amoore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*, Durham: Duke University Press 2020, hoofdstuk 1. Oudere bronnen zijn James C. Scott, *Seeing Like a State*, Yale: Yale University 1998 en John Cheney-Lippold, *We are Data, Algorithms and the Making of our digital selves*, New York: New York University Press 2017.

<sup>15</sup> *Financial Times* 17 juni 2023, 'AI and media companies negotiate landmark deal over news content'.

pen dat zij veel geld aan de groei van de verspreiding van kwalitatief goede informatie via chatbots kunnen verdienen. Maar of individuele auteurs en uitvoerende kunstenaars daarmee akkoord zullen gaan, is maar de vraag. Toepassing van AI vormde een belangrijk onderdeel van de staking die in de zomer van 2023 in Hollywood onder acteurs en auteurs uitbrak.<sup>16</sup>

Zoals wel vaker bij de ontwikkeling van het auteursrecht als het geconfronteerd werd met massaal gebruik van data (denk aan de kopieermachine en het 'reprorecht'), wordt er ook gespeculeerd over het invoeren van AI-heffingssystemen.<sup>17</sup>

### 3.1.2 Privacy

In het voorgaande spelen ook portretrechten een grote rol. Ten aanzien van het gebruik van andere persoonsgegevens<sup>18</sup> is het beeld minder duidelijk. Het lijkt erop dat de partijen en toezichthouders zich hier op de output (het gebruik van algoritmes bij besluitvorming) concentreren. Maar dat kan veranderen, omdat er toch ook al allerlei vragen zijn gerezen rond de massale opslag van persoonsgegevens (en daar gaat het bij de input over) in het kader van bescherming van de staatsveiligheid en bestrijding van financiële fraude. Ik verwijs naar mijn commentaar in de *NJ* bij verschillende uitspraken van Europese rechters op dit punt.<sup>19</sup> Er is in de jurisprudentie een tendens te bespeuren dat massale opslag van persoonsgegevens voor dit soort doeleinden wel is toegestaan, maar strikte regels voor toegang en gebruik bij de output vergt. De jurisprudentie van het HvJ EU in het kader van het 'vergeetrecht' van de AVG (artikel 17) verbiedt bovendien het produceren van persoonsgegevens bij gevoelige output van een zoekresultaat van een zoekmachine. Met name op dit gebied zagen we dat de rechter zich moest uitspreken over een conflict tussen privacy en vrijheid van meningsuiting, omdat de zoekmachines de vrijheid om een zoekresultaat te publiceren stellen tegenover de privacy van de persoon die in het zoekresultaat (in een bepaalde context) wordt getoond.<sup>20</sup>

Het is dus heel goed denkbaar dat er op grond van de privacy (zoals uitgewerkt in de AVG) op den duur allerlei eisen aan de input (opslag, toegang en gebruik) en de output (formulering van het antwoord op de vraag) zullen moeten worden gesteld omdat er hoge eisen aan de non-discriminatie van deze systemen moeten worden gesteld of andere rechtstatelijke belangen moeten worden gewaarborgd.<sup>21</sup> Daarbij is een probleem (dat zich nu al voordoet) de Amerikaanse herkomst van de grote taalsystemen, omdat er op dit moment door de uitspraken van het HvJ EU in de Schrems-zaken een nog steeds geldend verbod van toepassing is op de uitwisseling van persoonsgegevens tussen Europa en de VS.<sup>22</sup> Daarvoor is een nieuwe overeenkomst tussen de EU en de VS nodig, waarin de VS zich verplicht op het eigen grondgebied een privacybescherming gelijkwaardig aan de AVG te realiseren. Die overeenkomst is er op dit moment nog steeds niet.

### 3.1.3 Vrijheid van meningsuiting

Wij hebben tot dusver nog geen op deze generatieve AI gebaseerde toepassingen gezien. Wel vragen al geruime tijd de onderwerpen 'desinformatie', 'fake nieuws' en 'deep fake' de aandacht. De op dit gebied verrichte wetenschappelijke onderzoeken hebben niet tot betekenisvolle actie van wetgeving of bestuur geleid. Het standpunt is dat het bestaande (straf)recht toereikende middelen biedt. Ik verwijs naar de discussie over 'desinformatie'.<sup>23</sup> Een soortgelijk afhoudend standpunt zagen we ten aanzien van 'deep fake'.<sup>24</sup> Ik acht dit betreurenswaardig, te meer daar hoe langer hoe duidelijker wordt dat de sociale- en informatietechnologische infrastructuur, die de directeur van het Oxford Internet Institute Philip Howard, 'Lie machines' heeft genoemd,<sup>25</sup> tot een toenemende schade aan de publieke sfeer heeft geleid. Daardoor worden burgers steeds meer misleid en afgeschrikt zich in de publieke sfeer te uiten of deel te nemen aan de publieke besluitvorming. De generatieve taal AI-systemen verhogen mijns inziens aanzienlijk het risico op verspreiding van niet-geverifieerde, maar 'gezaghebbende' antwoorden op vragen over de sociale en natuurkundige werkelijkheid. Ze kunnen ook ingezet worden voor politieke manipulatie. Howard:

16 Zie bijv. de *New Scientist* 28 juli 2023. Er loopt ook een class action over deze kwestie, zie Stef van Gompel, 'Kunstmatige intelligentie: de opmars voorbij', *Auteursrecht* 2023/2, p. 53-54. Zie ook Dirk Visser, 'Robot kunst en auteursrecht', *NJB* 2023/7. In Europa spelen voor wetenschappelijk AI-onderzoek al jaren de artikelen 3 en 4 van de Auteursrechtlijn (DSM-richtlijn 2019/790/EU) een rol, zie daarover P.B. Hugenholtz, 'Artikelen 3 en 4 DSM-richtlijn: tekst en datamining', *Auteursrecht* 2019/5, p. 167-171.

17 Martin Senftleben, 'Generatieve AI – Is de tijd rijp voor een AI-heffing die inkomsten oplevert voor menselijke makers', *Auteursrecht* 2023, afl. 3, p. 103-114.

18 'Portretten zijn 'persoonsgegevens'. De juridische implicaties daarvan worden om historische redenen (de auteursrechten van de fotografen) in het auteursrecht behandeld, maar daar hebben ze niets mee te maken.

19 HvJ EU 22 november 2022, C-37/20 en C-601/20, *NJ* 2023/62, m.nt. E.J. Dommering (*UBO register*); EHRM 25 mei 2021, *NJ* 2021/361 (*Big Brother Watches You*), en HvJ EU 6 oktober 2021, C-511/18, C-512/18, C-520/18, *NJ* 2021/362, m.nt. E.J. Dommering (*La Quadrature du Net*).

20 HvJ EU 13 mei 2014, C-131/12, *NJ* 2014/385 (*Google I*); HvJ EU 24 september 2019, C-136/17, *NJ* 2019/434 (*Google/Frankrijk I*); HvJ EU 24 september 2019, C-507/17, *NJ* 2019/435 (*Google/Frankrijk II*), beide met noot E.J. Dommering.

21 Zie M. van Bekkum & F. Zuiderveen Borgesius, 'De spanning tussen het non-discriminatiebeginsel en het gegevensbeschermingsrecht', *NJB* 2023/1957, afl. 27, p. 2271-2280, en in dit nummer: Charlotte van Oirsouw & Jurgen Goossens, 'Naar een *by design* benadering van algoritmische, bestuursrechtelijke besluitvorming' (*Computerrecht* 2023/255).

22 HvJ EU 16 juli 2020, C-311/18, *NJ* 2021/24, m.nt. E.J. Dommering (*Schrems II*), waarin ook Schrems I wordt besproken.

23 Brief Minister van Binnenlandse Zaken aan de Tweede Kamer van 23 december 2022, kenmerk 2022-0000708642.

24 Zie: B. van der Sloot, Y. Wagenveld & B.J. Koops, *Deep fakes: De juridische uitdagingen van een synthetische samenleving*, WODC-rapport 3137, Tilburg University november 2021 en de kabinetsreactie aan de Tweede Kamer daarop van 16 juni 2023 (4362199).

25 Philip Howard, *Lie machines. How to save democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives*, New Haven/London: Yale University Press 2020.

“More and more political parties, government agencies and data-mining firms are playing with machine learning tools, and such tools will be very useful for helping to manage public life where independent journalism is marginalized.”<sup>26</sup>

Dit dwingt tot het ontwikkelen van effectieve normen en toezicht die nu (en na een ongewijzigde toestandkoming van de AI-verordening) blijven ontbreken. Ik kom er par. 5 van dit artikel op terug.

### 3.1.4 Mededingingsrecht

De generatieve taalsystemen worden via applicaties, zoals ChatGPT, in de markt gezet door partijen met zeer aanzienlijke marktmacht. Zij presenteren data als gebruikersvriendelijke systemen die via open-source software zijn te controleren. Maar hoe open is die software? Een recent onderzoek van een groep onderzoekers aan de Radboud Universiteit plaatst daar zeer kritische kanttekeningen bij, ik citeer:

“We find that while there is a fast-growing list of projects billing themselves as ‘open source’, many inherit undocumented data of dubious legality, few share the all-important instruction-tuning (a key site where human annotation labor is involved), and careful scientific documentations are exceedingly rare.”<sup>27</sup>

Tezelfdertijd publiceerden drie deskundigen in het Tijdschrift *Nature Machine Intelligence* een commentaar dat de software van deze chatbots niet transparant is en onvoldoende informatie verschaft aan het publiek en toezichthouders. Ik citeer:

“Two key problems arise from this opacity: first, foundation models such as Pal. Mare are trained on unknown collections of datasets using unknown training procedures; second the models behind the chat functionality are rendered inaccessible to public scrutiny by their proprietors, who protect their training procedures as intellectual property.”<sup>28</sup>

En hoe objectief zal de ‘output’ zijn? De EU-commissie is in 2022 een mededingingsonderzoek tegen Meta en Google gestart omdat zij eigen advertentiediensten (in zoekresultaten) bevoordelen boven die van concurrenten.<sup>29</sup> Ontwik-

kelen die vriendelijke en behulpzame ‘chatbots’ zich tot geraffineerde influencers die de diensten van het platform aan de man zullen gaan brengen? Het is niet onwaarschijnlijk dat de exploitanten van deze door grote platformen met door deze technieken gestuurde commerciële en sociale aandachtsmarkten ‘neutrale’ chatbots zullen inzetten.<sup>30</sup>

### 3.1.5 Publieke en private ‘automatische’ besluitvorming

Nederland heeft in de afgelopen periode affaires gekend die kenmerken vormen van ‘automatische’ besluitvorming door en over mensen. Eén ervan leidde in 2020 tot een beslissing door de Rechtbank Den Haag.<sup>31</sup> Het ging over de zogenaamde SyRI-toepassing (Systeem Risico Indicatie). Dat was (?)<sup>32</sup> een wettelijk instrument dat de overheid gebruikte ter voorkoming en bestrijding van fraude op het terrein van de sociale zekerheid en inkomensafhankelijke regelingen, de belasting- en premieheffing en de arbeidswetten. Het ging om technische infrastructuur en de bijbehorende procedures waarmee in een beveiligde omgeving anoniem data kunnen worden gekoppeld en geanalyseerd, zodat risicomeldingen kunnen worden gegenereerd. Het systeem bestond uit een samenwerking tussen de minister van Sociale Zaken, gemeentes en een groot aantal betrokken overheidsdiensten (zoals de Belastingdienst, immigratie en sociale diensten), waarbij persoonsgegevens over de verdachte personen konden worden uitgewisseld. In de zaak die door de Rechtbank Den Haag werd berecht ging het in het bijzonder om naleving van de Wet werk en bijstand (Wwb). Sinds 2004 is het op grond van deze wettelijke regeling verplicht aan het Uitkeringsinstituut Werknemersverzekeringen (het UWV) inlichtingen te verstrekken die voor de uitvoering van die wet noodzakelijk zijn. Dat is het project ‘Waterproof’ geworden. In dit project werden bestanden van 65 gemeenten gekoppeld met betrekking tot verbruikscijfers van waterbedrijven, woongegevens (m.n. van ontvangers van Wwb-uitkeringen). De wetgeving en het daarop gebaseerde uitvoeringsbesluit die dit project mogelijk maakten vormde de inzet van deze procedure, namelijk om deze wetgeving te toetsen aan de privacy-artikelen in het EU Handvest, het EVRM en de Algemene Verordening Gegevensbescherming (AVG). De Rechtbank verklaarde het systeem in strijd met de AVG, artikel 8 EVRM en artikel 8 van het EU Handvest omdat het intransparant was en er te veel persoonsgegevens zonder duidelijke doelbinding in waren opgeslagen. Kregen we bij de SyRI-affaire al de indruk dat het misschien toch meer mens dan machine was, de toeslagenaffaire (die

26 Howard a.w. noot 29, p. 144 e.v.

27 Andreas Liesenfeld, Alianda Lopez & Mark Dingemans ‘Opening up ChatGPT’, <https://arxiv.org/abs/2307.05532> (geraadpleegd 25 juli 2023).

28 Fabian Ferrari, José van Dijk & Antal van Bosch, ‘Foundation models and the privatization of public knowledge’, in: *Nature Machine Intelligence* 23 juli 2023, <https://www.nature.com/articles/s42256-023-00695-5> (geraadpleegd augustus 2023).

29 Nieuwsartikel, ‘Concurrentie: Europese Commissie start onderzoek naar deal tussen Meta en Google over online advertentiediensten’, 11 maart 2022. Beschikbaar op: [https://netherlands.representation.ec.europa.eu/nieuws/concurrentie-europese-commissie-start-onderzoek-naar-deal-tussen-meta-en-google-over-online-2022-03-11\\_nl](https://netherlands.representation.ec.europa.eu/nieuws/concurrentie-europese-commissie-start-onderzoek-naar-deal-tussen-meta-en-google-over-online-2022-03-11_nl).

30 Ik wijs op de analyse van deze platformmarkten van Paddy Leerssen, *Studies in the regulation of transparency for social media recommender systems*, diss. Universiteit van Amsterdam 12 april 2023.

31 Rb. Den Haag 5 februari 2020, ECLI:NL:RBDHA:2020:865), NJ 2020/386, m.nt. E.J. Dommering (SyRI).

32 Is het verleden tijd? David Davidson en Sebastiaan Brommersma van *Follow the Money* berichtten op 15 juni 2023 dat het systeem na de uitspraak van de Rechtbank Den Haag formeel gestopt is maar materieel als methode van aanpak is blijven bestaan. Zie David Davidson & Sebastiaan Brommersma, ‘Onder de motorkap van een fraudeopsporingsstelsel: wat niet ‘normaal’ is, is verdacht’, FTM.nl, 13 juni 2023.

ook ging over het opsporen van fraude bij uitkeringen) stond geheel in het teken van menselijke (bestuurlijke en rechterlijke) besluitvorming. Het proefschrift van Aviva de Groot<sup>33</sup> over AI gaat uiteindelijk vooral over het gebrek aan transparantie en uitleg van menselijke bestuurlijke en rechterlijke beslissingen en een analyse daarvan aan de hand van deze affaire. Zij noemt het een legitimiteitscrisis of nog sterker: 'het gebrek aan betrouwbaarheid en menselijkheid in de relaties tussen burger en overheid'.

Afgezien van de algemene bepalingen over verwerken van persoonsgegevens trekt in het bijzonder de aandacht artikel 22 AVG dat verbiedt een 'uitsluitend op geautomatiseerde verwerking, waaronder profilering, gebaseerd besluit waaraan voor de betrokkene rechtsgevolgen zijn verbonden of dat hem anderszins in aanmerkelijke mate treft.' Het Hof Amsterdam gaf op 4 april 2023 een vergaande beslissing in een geschil tussen Uber en haar chauffeurs in een geschil waar een negatieve automatische beslissing de rechtspositie van de chauffeurs had getroffen.<sup>34</sup> Er is ook bij het HvJEU een zaak aanhangig waarin is geconcludeerd door de advocaat-generaal.<sup>35</sup> Het betrof een automatisch besluit in het kader van een solvabiliteitsbeoordeling door een kredietmaatschappij. We zien bij deze voorbeelden dat de toepassing van 'gevaarlijke' algoritmes bij de besluitvorming heel dicht ligt tegen menselijke vooroordelen en onbehoorlijk bestuur in de 'real world'. Dat heb ik hiervoor behandeld bij de 'categorisering' (paragraaf 2.1.2).

## 4. Het EU recht: de AI-verordening

### 4.1 Een groot diffuus EU-domein

Alle aandacht richt zich op het voorstel van de AI-verordening (COM/2021/206 final) waarvan de bedoeling is dat zij eind 2023 definitief wordt vastgesteld.<sup>36</sup> Deze verordening een onderdeel is van een geheel van een 'Europese datastrategie', waarvan de Dataverordening die nog in voorbereiding is het sluitstuk gaat vormen.<sup>37</sup>

#### 4.1.1 De regels van de verordening

Artikel 1 definieert een AI-systeem op basis van een technische omschrijving die bij de verordening is gevoegd. De AI-verordening hanteert dus een specifieke definitie van AI en volgt daarin een risico-gerelateerde aanpak. Er wordt een onderscheid gemaakt tussen verboden systemen met onaanvaardbaar risico, toegelaten systemen met

verhoogde risico's en lage of minimale risico's. Voor de systemen met verhoogde risico's worden speciale regels gegeven, al beveelt de verordening een gedragscode aan om deze voor alle systemen toe te passen. Al met al is dit nog een zeer ruime definitie, zeker in de geamendeerde versie van het Europees Parlement.

De verboden systemen worden gedefinieerd aan de hand van de middelen die worden ingezet: subliminaal, gericht op kwetsbare personen, beoordelingscriteria van mensen gebaseerd op sociaal gedrag en biometrische identificatie van potentiële daders van strafbare feiten.

Het 'hoge risico' wordt in Bijlage III gedefinieerd aan de hand van mogelijke schending van grondrechten en dat gebeurt ook aan de hand van de *doelstelling* van de systemen die weer zijn ondergebracht in categorieën, bijvoorbeeld 'biometrische identificatie op afstand' in de rubriek 'beheer en exploitatie van kritieke infrastructuur'. De verordening bevat ten aanzien van de aanbieders van 'hoge risico'-systemen transparantie- en controlevoorschriften. Zij krijgen in hoofdstuk 2 van Titel III specifieke administratieve verplichtingen opgelegd die betrekking hebben op beheer, beperken van risico's en toegankelijkheid van de gebruikte software. Ook legt zij bijzondere zorgvuldigheidsverplichtingen op aan de gebruikers van AI-systemen met bijzondere risico's. Artikel 14 schrijft permanent menselijk toezicht voor.

De eerste vraag die opkomt na lezing van dit hele stuk is of ChatGPT en soortgelijke 'bot's' onder de verordening vallen. Dat was in het ontwerp van de verordening niet het geval, omdat zij geen specifieke doelstelling hebben aan de hand waarvan kan worden bepaald of het een verboden of hoog risicosysteem is. De resultaten van de stormachtige ontwikkeling in de large language-AI kwamen naar buiten op het moment dat de EU-commissie het voorstel voor de AI-verordening al bij het EU-parlement had ingediend. Je kunt je daarom zelfs afvragen of de aanpak van op basis van doelstellingen benoemde risico's die kan leiden tot een verbod of de kwalificatie van een 'verhoogd risico' niet al meteen achterhaald was. Je moet veeleer kijken naar de algemene risico's die aan dit soort systemen zijn verbonden en daar je normen op afstemmen.<sup>38</sup> Maar er is ook een zware lobby van de eigenaren van deze algemene systemen om buiten de AI-verordening te vallen. Het Amerikaanse weekblad *Times* bracht in juni een lobbypaper van OpenAI uit waaruit dat blijkt.<sup>39</sup> Inmiddels zijn de algemene systemen er in het intensieve amenderingstraject in het Europese parlement ingevoegd.<sup>40</sup>

De zwaar geamendeerde versie van de verordening heeft er een bijzonder ingewikkelde en moeilijk toe te passen regeling van gemaakt. Daarom moeten we allereerst terugvallen op de algemene rechtsgebieden die ik hiervoor heb aangestipt waarin AI is 'neergedaald'. Verder is de steeds centralere rol die de AVG speelt een aandachtspunt.

33 Aviva de Groot, *Care to explain? A critical epistemic injusticebased analysis of legal explanation obligations and ideals for "AI"-infused times*, diss. Universiteit van Tilburg, mei 2023, hoofdstuk 4 en 6.

34 Hof Amsterdam 4 april 2023, ECLI:NL:GHAMS:2023:793, 796 en 804.

35 Zaak C-634/21 (*Land Hessen/Schufa Holding*).

36 Zie voor een handzaam overzicht:

[https://www.europarl.europa.eu/news/nl/headlines/society/20230601STO93804/ai-verordening-eerste-regels-voor-artificiele-intelligentie,geraadpleegd 4 oktober 2023.](https://www.europarl.europa.eu/news/nl/headlines/society/20230601STO93804/ai-verordening-eerste-regels-voor-artificiele-intelligentie,geraadpleegd%204%20oktober%2023)

37 Zie hierover Marjolein Geus & Feyo Sickinghe, 'Voorstel voor de Europese Dataverordening: Sluitstuk van een tussentijdse reguleringsrevolutie', *Mediaforum* 2023-1, p. 2-14.

38 Aldus Natalie Helberger & Nicolas Diakopoulos, 'Chat GPT and the AI Act', *Internet Policy Review* 2023, Volume 12 (1).

39 'OpenAI White Paper on the European's Artificial Intelligence Act'.

40 [https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236\\_NL.html](https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_NL.html), geraadpleegd oktober 2023.

#### 4.1.2 *Het toezicht*

De AI-verordening voorziet in een certificatiesysteem dat we uit de EU-productwetgeving kennen. Nu de verordening kiest voor een systeem van door de aanbieder van AI na te leven risicoregels, betekent dit dat het zwaartepunt van de rechtsbescherming komt te liggen op preventief en repressief bestuurlijk toezicht. In Nederland is al eerder een 'algemeen algoritme-toezicht' van start gegaan. Op basis van het coalitieakkoord van 15 december 2021 moeten 'algoritmen gecontroleerd worden op transparantie, discriminatie en willekeur'. De brief aan de Kamer over dit onderwerp draagt duidelijk de sporen van de 'toeslagenaffaire'. De focus ligt op 'discriminatie, uitsluiting en transparantie'.<sup>41</sup> De AP heeft daarvoor in januari 2023 een nieuwe directie Coördinatie Algoritmes ingesteld. Deze is bedoeld als coördinerende instantie voor alle specifieke toezichthouders die met dit probleem te maken hebben.<sup>42</sup> Het ligt voor de hand dat de AP ook de centrale coördinator op grond van de AI-verordening zal worden (artikel 59 lid 8 van de verordening), maar definitieve beslissingen zijn daar op dit moment nog niet over genomen. De wolk van data en instructies die een AI-systeem vormt is per definitie moeilijk transparant te maken. Hier kan de AVG het voortouw nemen (er zullen altijd wel persoonsgegevens bij AI-toepassingen in het spel zijn), bij de invulling van het inzagerecht van de betrokkene op grond van artikel 15 AVG. Artikel 15 verlangt dat dat inzagerecht effectief is (in de Nederlandse tekst: de betrokkene heeft het recht *uitsluitend* te verkrijgen). Het HvJ EU heeft over de inhoud en effectiviteit van het inzagerecht een duidelijke beslissing gegeven.<sup>43</sup> De toepassing en uitleg van dit artikel van de AVG zal ongetwijfeld effect hebben op de invulling van de transparantieplichting van artikel 13 van de AI-verordening (de gebruikte software moet "voldoende transparant" zijn om gebruikers in staat te stellen de output van het systeem te interpreteren en op passende wijze te gebruiken').

### 5. **Conclusie: moet er nieuwe of andere 'AI-regulering' komen?**

#### 5.1 *Inleiding*

Uit het voorgaande komt naar voren dat de aanpak van de geamendeerde AI-verordening wel enige sturing biedt, maar tekortschiet in het stellen van algemene regels voor behoorlijk AI-gebruik en effectief toezicht daarop, zodat we op de materiële regels uit algemene rechtsgebieden moeten terugvallen die ik in paragraaf 4 de revue heb laten passeren.<sup>44</sup> We kunnen ons daarbij ook laten leiden door de *Blueprint for an AI Bill of Rights* die de Amerikaan-

se regering, ter voorbereiding op wetgeving, in oktober 2022 heeft gepubliceerd.<sup>45</sup>

Ik vat dit samen in vier overkoepelende 'controle' beginselen mede aan de hand waarvan bij de toepassing van de materiële regels in ieder rechtsgebied kan worden vastgesteld of een AI-systeem juridisch aanvaardbaar is.

#### 5.2 *Transparantiebeginsel*

Artikel 52 van de verordening formuleert het als een centraal beginsel voor AI-systemen die ontworpen zijn om op het algemene publiek te worden toegepast en artikel 69 beveelt codes of conduct aan die het regiem van de systemen met verhoogd risico aanvaarden voor alle aan het publiek aangeboden AI-systemen. Ik heb al eens bepleit dat dit een beginsel van behoorlijk bestuur zou moeten zijn om overheidsbesluiten te beoordelen.<sup>46</sup> In dat kader moet ook de 'algoritmische' basis van een bestuursbeleid worden beoordeeld, maar helaas is met de aanpassing van de Awb met dit aspect in het geheel geen rekening gehouden.<sup>47</sup> Maar het moet ook gelden voor andere rechtsgebieden, die steeds meer door automatische processen worden beheerst. Dat transparantiebeginsel houdt dus kort gezegd in, dat verantwoord moet (kunnen) worden hoe en met welke technische middelen beslissingen (die feitelijke en rechtsgevolgen voor individuen hebben) tot stand zijn gekomen. De eerste bevindingen zijn dat deze transparantie ontbreekt.<sup>48</sup>

#### 5.3 *Verificatiebeginsel*

Ik bepleit in aanvulling op het transparantiebeginsel een algemeen menselijk verificatiebeginsel.<sup>49</sup> Als zodanig zie ik in artikel 22 van de AVG (verbod op automatische beslissingen ten aanzien van zaken die iemands privacy raken) de uitdrukking van een algemener beginsel dat ook andere grondrechten regardeert, met name de waarheidsvinding die door het recht van vrije meningsuiting wordt beschermd. Een mens moet controleren of de feiten en de rechtsregels waar de beslissing op steunt wel kloppen.

41 Brief van de Staatssecretaris van Binnenlandse Zaken van 22 december 2022 aan de Tweede Kamer, kenmerk 2022-0000689652, p. 8.

42 Zie Directie Coördinatie Algoritmes (DCA) | Autoriteit Persoonsgegevens, geraadpleegd oktober 2023.

43 HvJ EU 4 mei 2023, C-203/22 (*Dun & Bradstreet/Magistrat der Stadt Wien*).

44 Over het tekortschietend toezicht, Marijen Hordijk & Thomas Lindsen, 'Kan de EU de black box openen?', in: C.J.H. Janssen, *Digitalisering en recht*, Nijmegen: Ars Aequi Libri 2023, p. 85-100.

45 <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>; Persoonlijk vind ik die beter dan de ethische code van de Raad van Europa:

<https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment>, beide geraadpleegd oktober 2023.

46 E.J. Dommering, 'Naar een algemeen transparantiebeginsel?', *NTB* 2022/5, p. 265-271.

47 Zie in dit nummer: Johan Wolswinkel 'De grote onbekende in de vergelijking: geautomatiseerde besluitvorming binnen de Wet versterking waarborgfunctie Awb' (*Computerrecht* 2023/256). In gelijke zin: Mirte Houweling & Lotte Kuijl, 'Algoritmische besluitvorming. Een leemte in de bestuursrechtelijke rechtsbescherming', in: C.J.H. Janssen (red), *Digitalisering en recht*, p. 117-130 en de door hen aangehaalde literatuur.

48 Andreas Liesenfeld, Alianda Lopez & Mark Dingemans 'Opening up ChatGPT', <https://arxiv.org/abs/2307.05532> (geraadpleegd 25 juli 2023). Fabian Ferrari, Josee van Dijk & Antal van Bosch, 'Foundation models and the privatization of public knowledge', in: Nature Machine Intelligence 23 juli 2023, <https://www.nature.com/articles/s42256-023-00695-5> (geraadpleegd augustus 2023).

49 Dit sluit aan op de algemene controleerbaarheid die Paddy Leerssen bepleit in hoofdstuk 6 van zijn proefschrift (zie noot 30) met wat hij noemt 'platform observability': de mogelijkheid om toegang te hebben tot de selectiemechanismen die een platform hanteert om de aandacht van de gebruikers te trekken.



Een voorbeeld van een geleidelijke toepassing van het verificatiebeginsel door de rechter, zou je de door het EU Hof ontwikkelde jurisprudentie rond de zoekmachine kunnen noemen, die meer verantwoordelijkheid bij de exploitant van de zoekmachine is gaan leggen voor het zoekresultaat in het kader van bescherming van het privéleven (jurisprudentie zie noot 20). Artikel 34 lid 1 onder b van de Digital Service Act sluit daarop aan (dat Helberger & Nicolas Diakopoulos,<sup>50</sup> al blijft het problematisch dat de DSA vasthoudt aan het verouderde systeem van de E-commerce-richtlijn dat ‘service providers’ in beginsel niet aansprakelijk zijn voor de inhoud van de informatie die zij doorgeven. De essentie van deze platforms is dat ze de informatie die zij verspreiden voortdurend modereren.<sup>51</sup> De aanbieders van de generatieve AI waar het hier om gaat, zullen in ieder geval als ‘grote’ platformen of zoekmachines gelden die onder de verificatieplicht vallen om ‘werkelijke en voorzienbare aantasting van grondrechten’, zoals de AI-verordening het uitdrukt, te voorkomen. Maar dit in mijns inziens niet voldoende, omdat AI-systemen vervangers lijken te worden van menselijke waarheidsvinding, terwijl ze dat niet zijn. Zij representeren immers niet de werkelijkheid, maar alleen het resultaat binnen het ‘grote taalsysteem’ waarin zij zijn getraind (zie hiervoor paragraaf 2). Ze zijn dus op dit punt helemaal niet ‘intelligent’ of ‘bovenmenselijk’ en ook geen zoekers van de waarheid (‘waarheid is een bijproduct’). Ze zijn ten opzichte van zoekmachines eigenlijk een stap terug. Die presenteren immers een zoekresultaat waarvan de feitelijke en logische uitgangspunten door de menselijke vrager geverifieerd kunnen worden, omdat bronnen worden getoond, althans door de vrager kunnen worden gevonden.<sup>52</sup> Dat valt weg bij de gediensstige chatbot, maar de pretentie is groter: het is geen *zoekmachine* maar een *antwoordmachine*. Om de machine te kunnen controleren moet je dus weten hoe de machine tot een antwoord is gekomen. In de medische wereld wordt al gewerkt aan AI-toepassingen die daarin kunnen voorzien. Jeroen van der Hoven, hoogleraar ethiek en technologie aan de TU-Delft: “Een cruciaal punt is dat duidelijk moet zijn hoe een taalmodel tot een bepaalde conclusie komt. Maar dat is lastig want deze modellen werken op basis van statistiek, niet causaliteit.”<sup>53</sup> Hear, hear!

En eenzelfde verhaal is bij ‘beeld’ en ‘geluid’ te houden. We kunnen aan met AI gefabriceerde digitale beeld of geluid

niet meer zien of horen of het ‘echt’ is. Aan dit aspect besteedt de AI-verordening wel aandacht bij de systemen die zijn gericht op ‘biometrische identificatie op afstand’, hiervoor geciteerd, maar het is een veel algemener probleem. Het behoeft geen betoog dat het transparantiebeginsel en het verificatiebeginsel nauw samenhangen (zie paragraaf 4 slot: er moet ‘voldoende’ transparantie zijn, zoals de AVG dat bij het inzagerecht formuleert), maar ze moeten m.i. toch van elkaar onderscheiden worden, omdat niet alleen de ‘black box’ moet worden geopend, maar ook de middelen aanwezig moeten zijn de juistheid van de inhoud daarvan langs menselijke waarneming te verifiëren. Dit kan in verschillende vormen gebeuren. De zwaarste vorm is dat er altijd (een weg naar) bronnen moet worden geopend. Een lichtere vorm is een waarschuwingsplicht en adequate voorziening op verzoek bronnen te ontsluiten. En er is bij een begane fout correctie en rectificatieplicht.

#### 5.4 Veiligheidsbeginsel

Dit is het beginsel aan de hand waarvan moet worden beoordeeld of het AI-systeem zo is ingericht dat het de waarden die er voor de gebruikers mee zijn gemoeid effectief beschermt.

#### 5.5 Beginsel van menselijke tussenkomst

Ik gaf hiervoor al aan dat dat de regel over automatische besluitvorming in artikel 22 AVG de uitdrukking is van een veel algemener beginsel dat degene die geraakt wordt in individuele rechten door een automatisch dataverwerkingsproces de mogelijkheid moet hebben beroep te doen op een menselijke (correctieve) beslissing. In de Amerikaanse ‘AI Bill of Rights’ neemt het beginsel aangeduid als ‘Human alternatives, Consideration and Fall Back’ een prominente plaats in.

De toepassing van de beginselen mondt uit in de vraag of het systeem voldoende veilig is ingericht. En zo ja, of er voldoende correctiemogelijkheden zijn door menselijke tussenkomst.

Maar ook de AI-verordening kan in deze gebieden doorwerken. Kan Bijlage III van de lijst van AI-systemen met een hoog risico niet een heuristisch instrument voor de rechter zijn om hoge zorgvuldigheidseisen te stellen aan voorbereiding en motivering van AI-toepassingen in deze risicogebieden? Dit zou de ‘reflexwerking’ van de AI-verordening op de algemene rechtspraak kunnen zijn.

## 6. Een positieve noot tot slot

In tal van technische en wetenschappelijke disciplines vervult AI hoe langer hoe meer een onmisbare functie. Een goed gecontroleerde inzet van AI kan dus naar een betere en efficiëntere wetenschappelijke en maatschappelijke praktijk in tal van sectoren leiden.

50 Aldus Natalie Helberger & Nicolas Diakopoulos, ‘Chat GPT and the AI Act’, *Internet Policy Review* 2023, Volume 12 (1).

51 Zie ook Lisa Janssen & Lisanne Kooijman, ‘Trekt Europa de teugels van TikTok aan?’, in: C.J.H. Janssen (red.) *Digitalisering en recht*, p. 29–49. Een algemene analyse van deze ontwikkeling in Tarleton Gillespie, *Custodians of the Internet. Platforms, content moderation, and the hidden decisions that shape the social media*, New Haven & London: Yale University Press 2018.

52 Daar is natuurlijk ook nog het nodige over te zeggen, toch blijft dit een belangrijk na te streven uitgangspunt. Er is een interessante analyse van de rol van Wikipedia bij het behoud van ‘feiten’ op het World Wide Web, Heather Ford, *Writing the revolution. Wikipedia and the Survival of Facts in the Digital Age*, Cambridge/Massachusetts: MIT Press 2022.

53 Citaat uit het artikel in *het Financieele Dagblad* 3 augustus 2023, p. 11 ‘Gaan Chatbots de druk op de zorg verlichten’.