

Emma Stevens en
Lars Groeneveld*

De ontwikkeling van generatieve artificiële intelligentie (GenAI) is niet onopgemerkt gebleven. ChatGPT groeide van een ‘klein onderzoeksproject’ van OpenAI uit tot een household name met inmiddels meer dan 180 miljoen actieve gebruikers per maand. Ook andere GenAI-toepassingen die in staat zijn afbeeldingen en andere vormen van content te genereren, zijn steeds toegankelijker geworden en daarmee populairder geworden. Net als bij de opkomst van de internetzoekmachine roept het gebruik van GenAI door het grote publieke verschillende juridische vragen op. Vaak staat daarbij de balans tussen innovatie en het waarborgen van fundamentele rechten en vrijheden centraal.¹ In dit artikel bespreken wij GenAI als volgende stap na de traditionele zoekmachine. Wij onderscheiden daarbij de zogenaamde in- en outputfasen als afzonderlijke verwerkingsmomenten bij GenAI-toepassingen (par. 1). Wij zien in de wetgeving en jurisprudentie argumenten die ruimte bieden aan de ontwikkeling van GenAI (inputfase) (par. 2) en tegelijkertijd een uitbreiding van de reikwijdte van het verbod op geautomatiseerde besluitvorming (outputfase) (par. 3). Wij vatten onze bevindingen samen in enkele afsluitende opmerkingen (par. 4).

1. De inputfase en outputfase op hoofdlijnen

Er kunnen bij de ontwikkeling en het gebruik van GenAI *grosso modo* twee fasen worden onderscheiden: de *inputfase* en de *outputfase*. Natuurlijk valt het daadwerkelijke ontwikkelings- en operationele proces van GenAI-modellen uiteen in meerdere deelstappen met allen hun eigen juridische complexiteiten die wij hier onbesproken laten.²

De *inputfase* laat zich kenmerken door het verzamelen van gegevens voor het aanleggen van een trainingsdataset, en het ontwikkelen van het AI-model. Bij een operationele GenAI-toepassing vormen de ingevoerde gebruikersopdrachten (ook wel *prompts*) ook *input* waar het model op wordt getraind. Dergelijke *input* laten wij hier buiten beschouwing, omdat wij ons in dit onderdeel richten op de ontwikkelingsfase van GenAI.

De gegevens die worden gebruikt om een trainingsdataset aan te leggen verschillen naar gelang het type AI-model en het beoogde gebruik.³ Voor een tekstgenererend model, zoals ChatGPT (*Generative Pre-trained Transformer*), bestaat de *input*

meestal uit een reeks woorden, zinnen of een tekstfragment dat het model gebruikt om verder op voort te bouwen en een coherente *output* te genereren.⁴

Voordat de *input* daadwerkelijk aan het model wordt doorgegeven, ondergaat deze vaak een voorbewerking. Dit is een essentiële stap waarin de gegevens worden omgezet naar een vorm die het model effectief kan verwerken. Bij tekst kan dit bijvoorbeeld het normaliseren van de tekst inhouden, zoals het verwijderen van speciale tekens, of het tokeniseren van de tekst, waarbij deze wordt opgedeeld in kleinere eenheden zoals woorden of subwoorden.⁵ Voor beeldgenererende modellen kan de *input* bestaan uit afbeeldingen die eerst worden aangepast aan het juiste formaat of kleurenschema.

Daarnaast speelt de context van de verzamelde gegevens een belangrijke rol in de *inputfase*. Op basis van de context van de verzamelde gegevens kunnen bepaalde parameters worden ingesteld om de manier waarop het model *output* genereert te sturen, zo kan de voorspelbaarheid en creativiteit van de *output* worden beïnvloed.⁶

* Mr. E.W.V. Stevens en mr. L. Groeneveld zijn beiden advocaat te Den Haag (Pels Rijcken).

1 Zie ook: CBP, Richtsnoeren publicatie van persoonsgegevens op internet, december 2007.

2 D. De Silva, 'An artificial intelligence life cycle: From conception to production', *Patterns*, 2022/3, afl. 6.

3 A. Dermawan, 'Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese "nonenjoyment" purposes?', *The Journal of World Intellectual Property* 2023/27 afl. 1.

4 M. Ruby, How ChatGPT Works: The Model Behind The Bot, A brief introduction to the intuition and methodology behind the chat bot you can't stop hearing about, 30 januari 2023 beschikbaar via towardsdatascience.com.

5 De Silva 2022 en S. Feng, C.Y. Park, Y. Liu, Y. Tsvetkov, 'From pretraining data to language models to downstream tasks: tracking the trails of political biases leading to unfair NLP Models', arXiv:2305.08283.

6 T.W. Edgar, D.O. Manz, *Research Methods for Cyber Security*, Cambridge (Mass.): Elsevier 2017, p. 153-173.

Na het verwerken van de *input* en het ontwikkelen van het betreffende AI-model kan *output* worden gegenereerd die afwijkt van de gegevens die zijn ingevoerd tijdens de *input*fase. Voor tekstmodellen betekent *output* het produceren van zinnen, paragrafen of langere teksten die relevant en consistent zijn met de gegeven gebruikersopdracht of zogenaamde *prompt*. Voor andere vormen van GenAI, zoals beeldgeneratie, bestaat de *output* uit het creëren van afbeeldingen of video's op basis van een specifieke gebruikersopdracht.⁷

Als laatste stap in het verwerkingsproces wordt de gegenereerde *output* vaak geëvalueerd om te bepalen of deze voldoet aan de door de GenAI-aanbieder gestelde kwaliteitsnormen. Deze evaluatie kan geautomatiseerd plaatsvinden, bijvoorbeeld via algoritmen die de coherentie of originaliteit van de *output* meten, of met menselijke tussenkomst via *reinforced learning*.⁸

Reinforcement Learning from Human Feedback (RLHF) is een methode die wordt gebruikt om GenAI-modellen te verbeteren door menselijke feedback te integreren in de ontwikkeling van de GenAI-toepassing. In plaats van alleen te vertrouwen op geautomatiseerde evaluaties, wordt bij RLHF de *output* van GenAI-modellen beoordeeld door menselijke annotatoren. Deze annotatoren rangschikken de resultaten op basis van verschillende criteria, zoals nauwkeurigheid, relevantie en geschiktheid voor het beoogde gebruik.

De feedback van deze menselijke beoordelaars wordt vervolgens gebruikt om het model bij te stellen via een proces dat lijkt op *reinforced learning*, waarbij het model leert om keuzes te maken die dichter bij de voorkeuren van de menselijke beoordelaars liggen. Dit proces helpt niet alleen om de nauwkeurigheid van het model te verbeteren, maar ook om ongewenste of schadelijke *output* te vermijden, zoals het genereren van bevooroordeelde of ongepaste reacties. RLHF is vooral nuttig in complexe of gevoelige domeinen waar menselijke nuances en ethische overwegingen een belangrijke rol spelen, en wordt toegepast door bedrijven zoals OpenAI en Google.⁹

2. De inputfase (scraping)

Terug naar de *input*fase. Er wordt bij het ontwikkelen van GenAI-modellen vaak gebruikt gemaakt van *scraping* voor het aanleggen van de benodigde trainingsdataset. Scraping is het op geautomatiseerde en systematische wijze doorzoeken van webpagina's met behulp van softwareprogramma's (zgn. *crawlers*), waarbij de inhoud van webpagina's wordt gedownload en relevante informatie wordt geëxtraheerd. Afhankelijk van het soort trainingsdataset dat wordt samengesteld, kan het daarbij gaan om tekst, afbeeldingen, video's, of andere gegevens. De verzamelde gegevens worden vervolgens gebruikt om een trainingsdataset aan te leggen, en GenAI-modellen te trainen, waarbij het te trainen model patronen

leert herkennen in de trainingsdataset die het later kan gebruiken om nieuwe content te genereren.¹⁰

Het belang van scraping voor de ontwikkeling van GenAI ligt in het vermogen toegang te bieden tot omvangrijke en diverse datasets die anders moeilijk te verkrijgen zouden zijn. Door gegevens vanuit verschillende bronnen, waaronder nieuwsartikelen, wetenschappelijke publicaties en sociale media te combineren, kunnen ontwikkelaars GenAI-modellen trainen die in staat zijn complexere en meer contextueel relevante content te genereren.

Daarnaast is scraping niet alleen belangrijk voor het vergaren van initiële trainingsdata, maar ook voor het *up-to-date* houden van GenAI-modellen. In een constant veranderende wereld met ontelbare datastromen en ontelbare datapunten is het een voor ontwikkelaars een uitdaging trainingsdata relevant en actueel te houden. Scraping biedt de mogelijkheid modellen voortdurend te voeden met relevante informatie, waardoor ze beter in staat zijn actuele trends en ontwikkelingen te weerspiegelen.

Scraping vertoont overeenkomsten met web-indexeren, de techniek waarop zoekmachines zijn gestoeld. Indexeren is het proces waarbij, net als bij scraping, een zogenaamde *crawler* het web doorzoekt en een database (index) aanlegt. De index zorgt ervoor dat een zoekmachineaanbieder sneller en *up-to-date* zoekresultaten kan aanbieden, zonder bij elke zoekopdracht het hele internet te doorzoeken. Beide technieken – scraping en indexeren – zien dus op het verzamelen van grote hoeveelheden (in beginsel) vrij toegankelijke gegevens, waaronder persoonsgegevens, op het internet.

De overeenkomsten tussen scraping en indexeren zijn relevant, omdat het Hof van Justitie van de Europese Unie (hierna HvJ EU of het Hof) in 2019 oordeelde over de rechtmatigheid van indexeren ten behoeve van zoekmachines, in casu die van Google.¹¹

In het *GC e.a./CNIL*-arrest buigt het Hof zich allereerst over de vraag of de AVG van toepassing is op het indexeren dat plaatsvindt voor de Google zoekmachine. In lijn met het *Google/Costeja*-arrest uit 2014 oordeelt het Hof dat de activiteiten van een zoekmachineaanbieder die bestaan uit, het vinden van door derden op het internet gepubliceerde of opgeslagen informatie, het vervolgens automatisch indexeren, het tijdelijk op te slaan, en uiteindelijk het in een bepaalde volgorde ter beschikking stellen aan internetgebruikers, kwalificeren als verwerkingen van persoonsgegevens, waarvoor de zoekmachine verwerkingsverantwoordelijk is, en waar de AVG (of vóór 25 mei 2018: de Richtlijn 95/46/EG ("de richtlijn")) op van toepassing is.¹²

Vervolgens beantwoordt het Hof de vraag of het verbod op de verwerking van bijzondere gegevens en strafrechtelijke gegevens (art. 8, eerste en vijfde lid, van de richtlijn resp. art. 9,

7 Denk aan: Dall-E, Midjourney, en Stable Diffusion.

8 Reinforcement learning (RL) is een vorm van machine learning waarin een AI-model leert door te proberen en fouten te maken. Het model maakt zelfstandig keuzes en krijgt beloningen of straffen, afhankelijk van hoe goed die acties zijn. Het doel is om te leren welke acties de meeste beloningen opleveren, zonder dat de juiste uitkomst wordt voorgeprogrammeerd. Zie ook: R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction*, Cambridge (Mass.): MIT Press 2018.

9 OpenAI, 'GPT-4 Technical Report: Using RLHF to Train Models'. Beschikbaar via: <https://openai.com/research/gpt-4https://openai.com/research/gpt-4> geraad-

pleegd op 5 september 2024, en DeepMind, 'Learning through Human Feedback: AI Safety and Ethics', beschikbaar via: <https://www.deepmind.com/publications/geraadpleegd> op 5 september 2024.

10 ChatGPT analyseert de ingevoerde tekst, voorspelt en genereert passende antwoorden. Het model berekent waarschijnlijkheden voor alle mogelijke volgende woorden en kiest het woord met de hoogste waarschijnlijkheid, rekening houdend met de context van de hele conversatie.

11 HvJ EU 24 september 2019, C-507/17, ECLI:EU:C:2019:772, (*Google/CNIL*) nr. 41; HvJ EU 24 september 2019, C-136/17, ECLI:EU:C:2019:773, (*GC e.a./CNIL*).

12 HvJ EU 13 mei 2014, C-131/12, ECLI:EU:C:2014:317 (*Google/Costeja*)

eerste lid en art. 10 AVG) (hierna: 'de verwerkingsverboden') ook van toepassing is op aanbieders van zoekmachines. Het Hof neemt tot uitgangspunt dat er geen reden is om, zoals Google had betoogd, ervan uit te gaan dat de verwerkingsverboden niet van toepassing zijn op verwerkingen die de aanbieder van een zoekmachine verricht vanwege de 'specifieke kenmerken' daarvan. De specifieke kenmerken van de verwerkingen die de aanbieder van een zoekmachine verricht, kunnen volgens het Hof wel van invloed zijn op de *omvang* van diens verantwoordelijkheid en de concrete verplichtingen in het licht van de verwerkingsverboden.¹³

In dat kader acht het Hof het relevant dat de verantwoordelijkheid van de zoekmachine zich niet uitstrekt tot hetgeen wordt geput uit websites van derden.¹⁴ De verantwoordelijkheid van de partij die ter exploitatie van een zoekmachine gegevens indexeert, wordt volgens het Hof pas geactiveerd nadat (gevoelige) gegevens in de vorm van zoekresultaten online zijn geplaatst. Pas dan, zo lijkt te volgen uit de redenering van het Hof, kunnen het recht van de betrokkene op eerbiediging van het privéleven en op bescherming van persoonsgegevens aanzienlijk worden aangetast.¹⁵ Het Hof concludeert vervolgens dat de verwerkingsverboden, 'gelet op de verantwoordelijkheden, de bevoegdheden en de mogelijkheden' van de aanbieder van een zoekmachine, pas kunnen gelden voor die aanbieder naar aanleiding van een door de betrokkene ingediend verwijderingsverzoek (artikel 17 AVG).¹⁶

Door de reikwijdte van AVG-verplichtingen afhankelijk te maken van de omvang en reikwijdte van de verantwoordelijkheden, bevoegdheden, en mogelijkheden van de zoekmachineaanbieder, kunnen verboden en beperkingen enkel worden gehandhaafd op grond van gepubliceerde zoekresultaten en bijgevolg via een controle *a posteriori* (in de vorm van een verwijderingsverzoek).¹⁷ In die zin geldt bij het indexeren van gegevens door zoekmachineaanbieders het verbod op verwerking van bijzondere persoonsgegevens niet in absolute vorm.¹⁸

De opkomst van GenAI, in het bijzonder ChatGPT, werpt de vraag op of de conclusies van het Hof in *GC e.a./CNIL* ook kunnen worden toegepast op scraping ten behoeve van de ontwikkeling van GenAI. Het *GC e.a./CNIL*-arrest maakt duidelijk dat de omvang van de verplichtingen van zoekmachines onder de AVG kunnen worden beperkt naar gelang hun verantwoordelijkheden, bevoegdheden en mogelijkheden. Voor zoekmachines betekent dit dat zij niet verplicht zijn reeds tijdens het doorzoeken en indexeren van websites aan alle AVG-verplichtingen te voldoen; pas tijdens de outputfase waarin zoekresultaten worden getoond, ontstaat een te effectueren verantwoordelijkheid. Voor het scrapen van gegevens voor de ontwikkeling van GenAI-systemen suggereert dit dat dergelijke systemen mogelijk ook niet onderworpen zijn aan alle verplichtingen uit de AVG op het moment van gegevensverzameling. Immers, ook de ontwikkelaar van GenAI kan in gelijke zin niet verantwoordelijk worden gehouden voor de gegevens op de webpagina's die worden gescrept.¹⁹

In lijn met hetgeen het Hof in *GC e.a./CNIL* overweegt over de reikwijdte van de verantwoordelijkheden van de zoekmachine-

aanbieder, zouden de verplichtingen uit de AVG dan in een later stadium kunnen gelden, bijvoorbeeld op het moment dat *output* wordt gegenereerd. Of in de woorden van het Hof in het *GC e.a./CNIL*-arrest "in het kader van zijn verantwoordelijkheden, zijn bevoegdheden en zijn mogelijkheden als verantwoordelijke voor de verwerking die tijdens de activiteit van deze machine wordt verricht."²⁰

Een uitleg waarbij ook voor scraping het verbod op verwerking van bijzondere persoonsgegevens niet in absolute vorm geldt, zou ook in lijn zijn met het oordeel van het Hof over de zorgplicht van zoekmachines bij de publicatie van hun zoekresultaten. In *GC e.a./CNIL* oordeelde het Hof dat geen sprake is van *mere conduit* bij het aanbieden van zoekresultaten, door zoekmachineaanbieders ervoor verantwoordelijk te houden dat zoekresultaten een actueel beeld van de betrokkene schetsen.²¹ Er ligt bij de zoekmachineaanbieder dus een zekere redactionele verantwoordelijkheid.

In de context van GenAI draagt scraping niet alleen bij aan de ontwikkeling van AI-modellen, maar maakt scraping het ook mogelijk AI-modellen te voeden met een actuele representatie van de werkelijkheid. Het kunnen samenstellen van een actuele trainingsdataset is essentieel voor het genereren van *output* die eveneens actueel en waarheidsgetrouw is. Wanneer geen ruimte wordt geboden voor dergelijke technieken, wordt dus niet alleen de ontwikkeling van GenAI bemoeilijkt, maar wordt het voor bestaande modellen ook moeilijker om actuele, waarheidsgetrouwe en kwalitatieve output te leveren. Dit lijkt haaks te staan op de uiteindelijke redactionele eisen die het Hof aan het genereren en publiceren van zoekresultaten stelt.

Samenvattend is de scheiding van verantwoordelijkheden bij het indexeren van gegevens en het genereren van zoekresultaten, zoals in het *GC e.a./CNIL*-arrest, voor de ontwikkeling van GenAI op basis van gescrepte gegevens zeer relevant. Hoewel het *GC e.a./CNIL*-arrest zich richt op zoekmachines, zouden de overwegingen goed toepasbaar kunnen zijn op scraping voor GenAI systemen. Net zoals bij zoekmachines, kan er een onderscheid worden gemaakt tussen de fase van gegevensverzameling (scraping) en de fase waarin verantwoordelijkheden uitvoerbaar zijn (genereren van *output*). Toepassing van *GC e.a./CNIL* op scraping kan betekenen dat bij het samenstellen van de trainingsdataset geen sprake is van een absoluut verwerkingsverbod, maar dat de verdere verwerking en *output* aan strengere (redactionele) eisen moeten voldoen.

2.1 De AP

In haar onlangs gepubliceerde handreiking over scraping, onder de kop "AP: scraping bijna altijd illegaal", neigt de Autoriteit persoonsgegevens (AP)²² naar een andere conclusie. De toezichthouder erkent daarin dat er overeenkomsten zijn tussen "zoekmachines en scrapers".²³ Echter, zij lijkt meer waarde te hechten aan de veronderstelde verschillen dan aan de overeenkomsten. Zo laat de AP meewegen dat, volgens haar, geen extra verwerkingen plaatsvinden op de verzamelde informatie nadat de zoekmachine deze eenmaal heeft

13 *GC e.a./CNIL*, nr. 46.

14 *GC e.a./CNIL*, nr. 46.

15 *GC e.a./CNIL*, nr. 46.

16 *GC e.a./CNIL*, nr. 47.

17 Conclusie A-G Szpunar 10 januari 2019, C-136/17, ECLI:EU:C:2019:14.

18 Zie ook de annotatie E.J. Dommering bij *GC e.a./CNIL* in *NJ* 2019/435, nr. 9.

19 *GC e.a./CNIL*, nr. 46.

20 *GC e.a./CNIL*, nr. 48.

21 *GC e.a./CNIL*, nr. 78.

22 Voor de aanduiding van de toezichthouder sluiten wij aan bij de spellingswijze waarvan gebruik wordt gemaakt in de Uitvoeringswet AVG (*Stb.* 2018, 44). Dit impliceert dat we de toezichthouder aanduiden met een hoofdletter en een kleine letter: Autoriteit persoonsgegevens, afgekort AP.

23 AP, 'Handreiking Scraping door particulieren en private organisaties', mei 2024.

geïndexeerd. En dat nog onzeker is of scraping op dezelfde voet als zoekmachines kan worden geacht de vrijheid van informatie te dienen.²⁴ Volgens de AP zijn ontwikkelaars van GenAI-modellen die gebruik maken van scraping dus al tijdens de *input*-fase gebonden aan het verwerkingsverbod voor bijzondere en strafrechtelijke gegevens.

De afweging van de AP heeft tot gevolg dat bijzondere en strafrechtelijke persoonsgegevens zullen moeten worden geïdentificeerd voordat zij worden verwerkt met een scraping-tool, hetgeen in de praktijk ondoenlijk is, net als bij het indexeren van webpagina's.

In zijn conclusie bij *Google/Costeja* wees A-G Jääskinen op de naar zijn mening 'absurde conclusie' dat, waar het gaat om de verwerking van bijzondere gegevens, de verwerkingen in het kader van zoekmachines in strijd zouden zijn met het Unierecht.

*“De tegengestelde opvatting zou meebrengen dat internetzoekmachines in strijd met het Unierecht zijn, wat mijns inziens een absurde conclusie is. Met name zouden de activiteiten van aanbidders van internetzoekmachines, wanneer zij werden beschouwd als verantwoordelijken voor de verwerking van de persoonsgegevens op bronpagina's van derden die ergens „bijzondere” gegevens in de zin van artikel 8 van de richtlijn zouden bevatten (namelijk persoonsgegevens waaruit politieke opvattingen, godsdienstige overtuigingen, of gegevens die de gezondheid of het seksuele leven betreffen), automatisch illegaal worden zodra niet is voldaan aan de strenge voorwaarden die dat artikel stelt aan de verwerking van zulke gegevens.”*²⁵

De AP stuurt met haar handreiking aan op, volgens ons, een eveneens bijzondere conclusie dat de verwerkingen in het kader van het ontwikkelen van GenAI-modellen “bijna altijd” in strijd zijn met het Unierecht. De AP ziet de consequentie van haar standpunt in, en zegt daarover het volgende:

*Als we kijken naar de hierboven besproken uitzonderingen op het verbod op het verwerken van bijzondere persoonsgegevens, dan wordt duidelijk dat het bij het scrapen van informatie van internet vaak lastig of onmogelijk zal zijn om onderscheid te maken tussen gewone persoonsgegevens en bijzondere persoonsgegevens. En dat u dus al snel (ook) bijzondere persoonsgegevens scrapet, wat (behoudens uitzonderingen) verboden is. Dit kan tot gevolg hebben dat een verwerking in zijn geheel niet is toegestaan vanwege de verwerking van bijzondere persoonsgegevens.*²⁶

Wij achten de overeenkomsten tussen indexeren (en dus niet zoekmachines in zijn geheel, zoals de AP stelt) en scrapen belangrijker dan de, al dan niet, daadwerkelijke verschillen waarop de AP haar conclusies over de 'illegaliteit' van scraping baseert.

Allereerst is het een misvatting dat zoekmachines geen extra verwerkingen toepassen op de verzamelde informatie nadat deze eenmaal zijn geïndexeerd. Zoekmachines zoals Google of Bing voeren diverse complexe verwerkingen uit op de geïndexeerde data om de relevantie, context en kwaliteit van zoekresultaten te optimaliseren, zoals snippets, thumbnails,

maar ook de volgorde waarin zoekresultaten worden gepresenteerd. Deze processen omvatten algoritmische analyses, zoals het beoordelen van de inhoudskwaliteit, het herkennen van patronen, en het toepassen van semantische zoektechnieken die helpen bij het beter begrijpen van de zoekopdrachten en de onderliggende intenties van gebruikers. Bovendien gebruiken zoekmachines machine learning-algoritmen om de zoekresultaten voortdurend te verfijnen en personaliseren op basis van gebruikersgedrag en feedback.²⁷ Ook voor de ontwikkeling van GenAI wordt de gescrapete data bewerkt om de kwaliteit of relevantie van de gegevens aan te passen naar gelang dit de ontwikkeling van het AI-model ten goede komt.

Daarbij komt dat zowel bij scraping als bij indexeren de partij die de gegevens verzamelt in de praktijk geen invloed kan uitoefenen op de gegevens die het binnenhaalt. En, juist om die reden, de feitelijke onmogelijkheid controle uit te oefenen, beperkt het Hof de reikwijdte van de verplichting uit de AVG naar gelang de verantwoordelijkheden, bevoegdheden en mogelijkheden van de zoekmachineaanbieder.

Een tweede overeenkomst tussen scraping en indexeren is de noodzaak grote hoeveelheden gegevens te verzamelen om de betreffende dienst aan het grote publiek te kunnen aanbieden. Waar indexeren het voor zoekmachines mogelijk maakt grote hoeveelheden informatie doorzoekbaar te maken en te ontsluiten voor het grote publiek, vervult scraping eenzelfde onmisbare functie bij de steeds belangrijker rol van GenAI bij het ontsluiten van informatie. Zo wordt ChatGPT in veel gevallen al ingezet als alternatieve zoekmachine, en wordt de functionaliteit ingebouwd in reeds bestaande zoekapplicaties.²⁸ In die zin vervult GenAI reeds een vergelijkbare rol voor de vrijheid van informatie als de meer traditionele zoekmachines, waarbij scraping net als indexeren een onmisbare schakel vormt.

Door zich te concentreren op de ogenschijnlijke verschillen tussen “zoekmachines en scrapen” lijkt de AP de meer relevante overeenkomsten over het hoofd te zien. En zij laat met een *de facto* verbod op scraping de balans tussen innovatie en het waarborgen van fundamentele rechten en vrijheden geheel anders uitslaan dan het Hof in *GC e.a./CNIL*. Dit terwijl de *output*-fase van GenAI voldoende ruimte biedt om de zorgplicht van de aanbieder te effectueren en de negatieve impact op betrokkenen te beperken, zonder direct te raken aan de ontwikkeling van GenAI-modellen.

Over de regulering van de *output*-fase het volgende.

3. De *output*-fase

Al sinds zijn voorganger in Richtlijn 95/46/EC is artikel 22 AVG een artikel dat zich in het bijzonder toespitst op de *output*-fase van systemen. Zo is artikel 22 AVG erop gericht individuen te beschermen tegen de potentiële risico's en negatieve gevolgen van geautomatiseerde besluitvorming, die zonder menselijke tussenkomst plaatsvindt.

²⁴ Idem, p. 18.

²⁵ Conclusie A-G Jääskinen van 25 juni 2013, C-131/12, ECLI:EU:C:2013:424 nr. 90.

²⁶ AP, Handreiking Scraping door particulieren en private organisaties, mei 2024, p. 21.

²⁷ Zie ook Conclusie A-G Jääskinen van 25 juni 2013, C-131/12, ECLI:EU:C:2013:424 nr. 91.

²⁸ Microsoft heeft bijvoorbeeld ChatGPT geïntegreerd in de Bing-zoekmachine, zie <https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/>.

De onderliggende gedachte is dat juist geautomatiseerde systemen risico's meebrengen die zich met name manifesteren in de uiteindelijke *output* van een dergelijk systeem, zoals onjuiste of discriminerende uitkomsten.²⁹ Tot voor kort werd de voorganger van artikel 22 AVG in Richtlijn 95/46/EC een 'second-class data protection right that remained largely dormant' genoemd, en was de relevantie van het artikel ook in de rechtspraak zeer beperkt.³⁰

Met de verspreiding van GenAI lijkt artikel 22 AVG een grotere rol van betekenis te gaan spelen. Temeer nu uit het recente *Schufa*-arrest zou kunnen worden opgemaakt dat de *output* van een GenAI-toepassing reeds op zichzelf een vorm van uitsluitend geautomatiseerde besluitvorming kan zijn die onder het verbod van artikel 22 AVG valt.³¹ We lichten dit toe.

3.1 Uitsluitend geautomatiseerde besluitvorming

Uitsluitend geautomatiseerde besluitvorming zoals bedoeld in artikel 22 AVG ziet op het nemen van individuele besluiten met technische middelen, zonder menselijke tussenkomst. De EDPB heeft in zijn richtsnoeren toegelicht dat de term 'volledige geautomatiseerde besluitvorming' inhoudt dat er geen zinvolle menselijke tussenkomst plaatsvindt in het besluitvormingsproces. Hierbij mag de tussenkomst niet slechts symbolisch van aard zijn, maar moet deze uitgevoerd worden door iemand die bevoegd en bekwaam is om het besluit te veranderen.³²

Het HvJ EU legt uit dat artikel 22 AVG slechts van toepassing is, indien is voldaan aan drie cumulatieve voorwaarden:

1) er moet sprake zijn van een 'besluit'; 2) dit besluit dient "uitsluitend op geautomatiseerde verwerking, waaronder profilering, te zijn gebaseerd", en; 3) aan het besluit moeten "voor [de betrokkene] rechtsgevolgen" zijn verbonden of het besluit moet hem "anderszins in aanmerkelijke mate treffen".³³

Anders dan de tekst wellicht doet vermoeden, is het begrip 'besluit' niet beperkt tot het bestuursrechtelijke besluitbegrip. Elk ander besluit met rechtsgevolgen voor de betrokkene kan een besluit in de zin van artikel 22 AVG zijn. Voorbeelden van besluiten met rechtsgevolgen in de zin van artikel 22 AVG zijn de beëindiging van een overeenkomst, het beoordelen van sollicitaties, en de weigering van een bepaalde wettelijke toegekende sociale uitkering zoals kinderbijslag of huurtoeslag, de weigering van toelating tot een land.³⁴

Niet alleen een besluit met rechtsgevolgen, maar ook een besluit dat de betrokkene anderszins in aanmerkelijke mate treft, kan onder het verbod van artikel 22 AVG vallen.

Een gegevensverwerking treft iemand volgens de WP29 in aanmerkelijke mate wanneer de effecten van de verwerking groot of belangrijk genoeg zijn om aandacht te verdienen, dat

wil zeggen als het besluit het potentieel heeft om de omstandigheden, het gedrag of de keuzen van de betrokken personen in aanmerkelijke mate te treffen; een langdurig of blijvend effect op de betrokkene te hebben; of in het uiterste geval, tot uitsluiting of discriminatie van personen te leiden.³⁵

In de considerans bij de AVG worden als voorbeelden van besluiten die de betrokkene in aanmerkelijke mate treffen genoemd het geautomatiseerd afwijzen van een sollicitatie en het geautomatiseerd afwijzen van een online ingediende kredietaanvraag.³⁶

In de Nederlandse rechtspraak is een aantal uitspraken gewezen over hoe het gebruik van een GenAI-toepassing door taxi-platforms Uber en Ola betrokkenen in aanmerkelijke mate kan treffen. Zo oordeelt het Gerechtshof Amsterdam dat het besluit van Uber om het account van Uberchauffeurs zonder menselijke tussenkomst te deactiveren op basis van een door software gegenereerd fraudesignaal als uitsluitend geautomatiseerde besluitvorming in de zin van artikel 22 AVG kwalificeert.³⁷ De Uberchauffeurs worden door dit deactiveringsbesluit volgens het Hof in aanmerkelijke mate getroffen omdat zij niet meer hun werkzaamheden kunnen uitoefenen waardoor zij inkomsten missen en de door hen gedane investeringen niet meer kunnen terugverdienen.³⁸

Ook de ritverdeling onder de Uberchauffeurs op basis van een *batched matching systeem* is volgens het Gerechtshof Amsterdam een geautomatiseerde verwerking die Uberchauffeurs in aanmerkelijke mate treft. Op basis van dit systeem worden chauffeurs op geautomatiseerde wijze gekoppeld aan passagiers. In het koppelingsproces worden onder meer de door chauffeur doorgevoerde persoonlijke voorkeuren meegewogen. Het systeem houdt tevens rekening met lage waarderingen door passagiers, waardoor die specifieke passagier niet meer wordt gekoppeld aan de betreffende chauffeur. Nu dit systeem ertoe kan leiden dat een chauffeur geen of minder vaak een aanbod krijgt, en daarmee bepalend is voor de inkomsten van de chauffeur, treft het besluitvormingsproces op basis van het *batched matching systeem* de Uberchauffeurs in aanmerkelijke mate.³⁹

3.2 De output van AI als 'uitsluitend geautomatiseerd besluit'

De keten van besluitvorming die uiteindelijk als *uitsluitend geautomatiseerd* wordt aangemerkt, strekt in de *Uber*-zaken van het gebruik van het algoritme tot aan een *besluit* met praktische gevolgen: in de ene zaak is dat uiteindelijk besluit het deactiveren van een account en in de andere zaak het maken van een ritverdeling. Uit de *Ola*-zaak lijkt echter te volgen dat de *output* van een GenAI-toepassing reeds op zichzelf een uitsluitend geautomatiseerd besluit in de zin van artikel 22 AVG kan vormen, ook als een eventueel (nadelig) besluit met praktische gevolgen uitblijft.

29 HvJ EU 7 december 2023, C-634/21, ECLI:EU:C:2023:957.

30 L.A. Bygave, 'Minding the Machine v2.0: The EU General Data Protection Regulation and Automated Decision-Making' in K. Yeung en M. Lodge (red.) *Algorithmic Regulation*, Oxford: Oxford University Press 2019.

31 HvJ EU 7 december 2023, C-634/21, ECLI:EU:C:2023:957, nr. 50.

32 WP29, Richtsnoeren inzake geautomatiseerde individuele besluitvorming en profilering voor de toepassing van Verordening (EU) 2016/679, (WP251 rev. 01), 3 oktober 2017, laatstelijk gewijzigd 6 februari 2018; overweging 71 bij de AVG.

33 HvJ EU 7 december 2023, C-634/21, ECLI:EU:C:2023:957 (*Schufa*), punt 42.

34 Zie WP29, Richtsnoeren inzake geautomatiseerde individuele besluitvorming en profilering voor de toepassing van Verordening (EU) 2016/679, (WP251 rev. 01), 3 oktober 2017, laatstelijk gewijzigd 6 februari 2018, p. 25

35 Zie WP29, Richtsnoeren inzake geautomatiseerde individuele besluitvorming en profilering voor de toepassing van Verordening (EU) 2016/679, (WP251 rev. 01), 3 oktober 2017, laatstelijk gewijzigd 6 februari 2018, p. 25-26; Rb. Den Haag 5 februari 2020, ECLI:NL:RBDHA:2020:865 (*SyRI*), r.o. 6.36.

36 Overweging 71 bij de AVG.

37 Hof Amsterdam 4 april 2023, ECLI:NL:GHAMS:2023:793.

38 Het gerechtshof noemt als aanmerkelijk gevolg verder dat de aantijgingen (al dan niet strafrechtelijke) repercussies kunnen hebben, met name voor toekomstige werkzaamheden. Zie Gerechtshof Amsterdam 4 april 2023, ECLI:NL:GHAMS:2023:793, r.o. 3.19.

39 Hof Amsterdam 4 april 2023, ECLI:NL:GHAMS:2023:796, r.o. 3.33.

Het Gerechtshof buigt zich in het *Ola*-arrest over de vraag of bij het genereren van de *fraud probability score* sprake is van uitsluitend geautomatiseerde besluitvorming. De *fraud probability score* is een geautomatiseerd systeem dat de kans op fraude door een *Ola*-chauffeur kan inschatten. Het Hof concludeert dat de besluitvorming *in de vorm van de score* de taxichauffeurs in aanmerkelijke mate treft omdat van deze score afhangt of *Ola* al dan niet overgaat tot nadere maatregelen tegen de chauffeur. Het gerechtshof benoemt daarnaast als aanmerkelijk gevolg dat de score, als die negatief is, een ernstige aantijging impliceert jegens de chauffeur, die zware repercussies kan hebben. Dat de score positief is uitgevallen bij één van de chauffeurs, doet volgens het gerechtshof niet af aan het belang van de score.⁴⁰ Het gerechtshof concludeert dan ook dat de score, nu deze geautomatiseerd tot stand komt, een vorm is van uitsluitend geautomatiseerde besluitvorming. Of per individu daadwerkelijk wordt overgegaan tot het treffen van nadere maatregelen lijkt voor de vraag of sprake is van geautomatiseerde besluitvorming niet van belang.

Het Gerechtshof Amsterdam lijkt in zijn uitspraak te zijn geïnspireerd door de conclusie van A-G Pikamaë in de *Schufa*-zaak⁴¹ die ruim twee weken was gepubliceerd voor het moment dat het Gerechtshof uitspraak deed. Dat de *output* van een GenAI-toepassing reeds op zichzelf een vorm van *uitsluitend geautomatiseerde besluitvorming* kan zijn, volgt namelijk ook de conclusie van A-G Pikamaë.

Het bedrijf *Schufa* ontwikkelt zogenoemde kredietcores die een voorspelling doen over het waarschijnlijke gedrag van consumenten. Kredietverstrekkers gebruiken de diensten van *Schufa* om kredietaanvragen te beoordelen. In het *Schufa*-arrest is door het Hof van Justitie aangenomen dat de geautomatiseerd gegenereerde kredietcore ten aanzien van de betrokkene, op basis waarvan een derde beslist of een krediet wel of niet wordt verstrekt aan de betrokkene, moet worden gezien als besluit in de zin van artikel 22 AVG.⁴² In lijn met de overweging van het Gerechtshof Amsterdam oordeelt het HvJ EU dat niet enkel de beslissing om een lening al dan niet toe te kennen, maar ook de vaststelling van de score op zichzelf een vorm van geautomatiseerde besluitvorming is:

“Hieruit volgt dat in omstandigheden als deze van het hoofdgeding – waarin de door een kredietinformatiebureau vastgestelde waarschijnlijkheidswaarde die aan een bank is doorgegeven een beslissende rol speelt bij het verlenen van een krediet – de vaststelling van die waarde derhalve als zodanig moet worden aangemerkt als een besluit „waaraan voor [de betrokkene] rechtsgevolgen zijn verbonden of dat hem anderszins in aanmerkelijke mate treft”, in de zin van artikel 22, lid 1, AVG.”⁴³

(onderstreeping toegevoegd)

Gevolg van de conclusie dat de kredietcore op zichzelf reeds een vorm van uitsluitend geautomatiseerde besluitvorming is, is dat de ontwikkelaar van de score, naast degene die acteert op basis van de score, te maken krijgt met het verbod van artikel 22 AVG.

Gelet op het voorgaande kan de mogelijke ruimte die *GC e.a./CNIL* biedt voor scraping tijdens de *input*fase, worden gecompenseerd door de ruime uitleg van artikel 22 AVG in de arresten *Ola* en *Schufa*.

Daarbij moet worden opgemerkt dat het bieden van ruimte voor scraping invloed heeft op (de ontwikkeling van) GenAI in algemene zin, en *Ola* en *Schufa* slechts duidelijke handvatten bieden voor een beperkt soort GenAI-toepassingen. Niet in alle gevallen is de *output* van een GenAI op zichzelf namelijk al een vorm van geautomatiseerde besluitvorming. Op basis van de *Ola*-uitspraak van het Gerechtshof Amsterdam en het *Schufa*-arrest van het Hof van Justitie geldt in ieder geval dat de op een individu betrekking hebbende *output* van een GenAI-toepassing op zichzelf onder het verbod van geautomatiseerde besluitvorming kan vallen indien i) van die *output* afhangt of al dan niet maatregelen worden getroffen tegen een werknemer (*Ola*), of ii) van die *output* afhangt of (door een derde) krediet wordt verstrekt aan een natuurlijke persoon (*Schufa*).⁴⁴

Van Kolfshoeten signaleert dat ook andere *output* van GenAI-toepassingen onder het verbod van geautomatiseerde besluitvorming kan vallen. Bijvoorbeeld *output* van GenAI-toepassingen die worden gebruikt voor medische besluitvorming hetgeen resulteert in een medische beslissing met betrekking tot een individuele patiënt.⁴⁵ Van Kolfshoeten acht het op basis van het *Schufa*-arrest mogelijk dat de *output* van medische AI-systemen die bijvoorbeeld de geschiktheid van een patiënt voor deelname aan een klinische studie beoordelen een vorm van *uitsluitend geautomatiseerde besluitvorming* in de zin van artikel 22 AVG is, zelfs als een zorgprofessional zelf de uiteindelijke beslissing neemt over de selectie van deelnemers aan het klinisch onderzoek.⁴⁶

Ongeacht de specifieke toepassing, is het sinds het *Schufa*-arrest van belang om te bepalen in hoeverre de *output* van het betreffende GenAI-model doorslaggevend is bij uiteindelijke besluitvorming.

Waarover het volgende.

3.3 Menselijke tussenkomst bij het gebruik van GenAI-toepassingen

In *Schufa* stelt het Hof van Justitie voor de toepasselijkheid van artikel 22 AVG als voorwaarde dat de handelswijze van de kredietverstrekker hoofdzakelijk door de score wordt bepaald.⁴⁷ De score kan dus reeds op zichzelf een vorm van *uitsluitend geautomatiseerde besluitvorming* zijn als de hierop volgende handeling met aanmerkelijke gevolgen (in het geval van *Schufa*: het verlenen van krediet) *hoofdzakelijk* op die score wordt gebaseerd. Het Hof van Justitie verbindt de constatering dat de handelswijze van de bank *hoofdzakelijk* door de risicoscore wordt bepaald aan de feitelijke vaststellingen van de verwijzende rechter dat een ontoereikende waarschijnlijkheidswaarde, in geval van een kredietaanvraag, er in bijna alle gevallen toe leidt dat de bank weigert het gevraagde krediet te verlenen.⁴⁸

40 Gerechtshof Amsterdam 4 april 2023, ECLI:NL:GHAMS:2023:804, r.o. 3.36-3.37.

41 Conclusie A-G Pikamaë 16 maart 2023, C-634/21, ECLI:EU:C:2023:220.

42 HvJ EU 7 december 2023, C-634/21, ECLI:EU:C:2023:957 (*Schufa*).

43 HvJ EU 7 december 2023, C-634/21, ECLI:EU:C:2023:957 (*Schufa*), nr. 50.

44 Vgl. Conclusie A-G Pikamaë waarom hij stelt: “Mijns inziens dient per geval te worden onderzocht of een ten aanzien van de betrokkene ingenomen standpunt als een „besluit” kan worden aangemerkt, en dient hierbij rekening te worden gehouden met de specifieke omstandigheden en de ernst van de gevol-

gen voor de juridische, economische en sociale positie van de betrokkene.”, zie Conclusie A-G Pikamaë 16 maart 2023, C-634/21, ECLI:EU:C:2023:220, nr. 39.

45 H.B. van Kolfshoeten, ‘A Health-conformant reading of the GDPR’s right not to be subject to automated decision-making’, *Medical Law Review* 2024, p. 11.

46 H.B. van Kolfshoeten, ‘A Health-conformant reading of the GDPR’s right not to be subject to automated decision-making’, *Medical Law Review* 2024/0, p. 11.

47 HvJ EU 7 december 2023, C-634/21, ECLI:EU:C:2023:957, nr. 62 en 73.

48 HvJ EU 7 december 2023, C-634/21, ECLI:EU:C:2023:957, nr. 48.

De vraag hoe kan worden gewaarborgd dat een uiteindelijk besluit niet *hoofdzakelijk* wordt gebaseerd op de *output* van generatieve AI is open en is afhankelijk van de omstandigheden van het geval. Zinnvolle menselijke tussenkomst in de vorm van een ingebouwd handmatig onderzoek voorafgaand aan het uiteindelijke besluit lijkt een noodzakelijke waarborg te zijn om ervoor te zorgen dat het uiteindelijke besluit niet *hoofdzakelijk* wordt gebaseerd op de *output* van een GenAI-toepassing.⁴⁹ Bij dat handmatige onderzoek moet dan worden voorkomen dat de mens niet ‘automatisch’ de uitkomst van de GenAI-toepassing overneemt. Daartoe is van belang dat de mens die het handmatige onderzoek verricht niet wordt beïnvloed door zogenoemde ‘*automation-bias*’: de neiging van mensen om voorkeur te geven aan suggesties van geautomatiseerde besluitvormingssystemen en tegenstrijdige informatie te negeren. Een maatregel die kan worden getroffen om een dergelijke *automation bias* te voorkomen, is door de mens die het handmatige onderzoek verricht een beoordeling te laten verrichten zonder dat hij op de hoogte is van de *output* van de GenAI-toepassing.

Het gebruik van op een individu betrekking hebbende GenAI-*output* lijkt aldus volgens de lijn van het HvJ EU in *Schufa* niet per definitie in strijd met artikel 22 AVG, zolang de *output* slechts ondersteunend en in ieder geval niet doorslaggevend is voor een hierop volgend besluit. Dit kan worden gewaarborgd door zeker te stellen dat zinnvolle menselijke tussenkomst plaatsvindt voorafgaand aan het besluit met aanmerkelijke gevolgen.

3.4 Of, is menselijke tussenkomst niet mogelijk?

Er zijn gevallen denkbaar waarbij menselijke tussenkomst niet mogelijk lijkt te zijn. Als immers moet worden aangenomen dat de *output* van een GenAI-toepassing de betrokkene reeds in aanmerkelijke mate treft doordat op basis van die *output* nader onderzoek kan worden ingesteld (zoals in de *Ola*-zaak het geval lijkt te zijn), is menselijke tussenkomst bij een dergelijk gebruik van een GenAI-toepassing onmogelijk. Met andere woorden: als eenmaal is bepaald dat de besluitketen reeds compleet is met de *output* van de GenAI-toepassing, kan geen mens daar meer tussen komen en is sprake van *uitsluitend geautomatiseerde besluitvorming*. In overweging 3.43 in de uitspraak *Ola* overweegt het Gerechtshof Amsterdam dat het bij de vraag of sprake is van een *uitsluitend geautomatiseerde verwerking* niet gaat om de vraag ‘of er op basis van de score geautomatiseerde besluitvorming plaatsvindt’, maar om de vraag ‘of deze score zélf tot stand komt op basis van uitsluitend geautomatiseerde verwerking’.⁵⁰ Het antwoord op die vraag zal in het geval van een door AI ontwikkelde score onvermijdelijk ‘ja’ zijn, waarmee menselijke tussenkomst bij de ontwikkeling van de score lijkt te zijn uitgesloten. Ook als een mens op basis van de *fraud probability score* een handmatig onderzoek start en vervolgens besluit de werknemer buiten dienst te stellen, lijkt, volgens de redenering van het Gerechtshof Amsterdam, het feit dat de score op zichzelf reeds een zware aantij-

ging vormt op grond waarvan een onderzoek kan worden ingesteld al problematisch in het licht van artikel 22 AVG.⁵¹

De *output* van een ‘toezichthoudende’ GenAI-toepassing zal naar analogie met de redenering van het gerechtshof om twee redenen al snel een aanmerkelijk gevolg hebben, i) omdat op basis van de *output* kan worden besloten nadere maatregelen te treffen, ii) omdat de *output* op zichzelf reeds een zware aantijging vormt.⁵² Het is de vraag of bij het gebruik van dergelijke toezichthoudende GenAI-toepassing menselijke tussenkomst kan plaatsvinden op een manier dat géén sprake is van *uitsluitend geautomatiseerde besluitvorming*.

Afgevraagd kan worden of de uitkomsten van een zoekopdracht via ChatGPT, of andere GenAI-toepassingen die ook als (alternatieve) zoekmachine worden gebruikt, kunnen worden aangemerkt als *output* die op zichzelf reeds een vorm van *uitsluitend geautomatiseerde besluitvorming* is. Dat niet geheel ondenkbaar is dat zware aantijgingen de *output* van dergelijke zoekopdrachten kunnen zijn, blijkt al uit de voorbeelden waarin ChatGPT journalisten als dader bestempelden van misdaden waarover zij veelvuldig publiceerden.⁵³ Net zo goed is denkbaar dat onder omstandigheden dergelijke *output* tot het treffen van maatregelen kan leiden.

Bovenstaande uitleg van artikel 22 AVG lijkt GenAI-aanbieders er dus onder omstandigheden toe te dwingen *output* die een zware aantijging vormt in ieder geval technisch onmogelijk te maken. Waar in *GC e.a./CNIL* werd geoordeeld dat voor zoekmachine-aanbieders een zorgplicht is weggelegd in de wijze waarop zij zoekresultaten aanbieden, volgt uit artikel 22 AVG ten aanzien van de *output* van GenAI onder omstandigheden dus een vergelijkbare redactionele zorgplicht.

4. Afsluitende opmerkingen

GenAI-toepassingen ontwikkelen en verspreiden zich in een razend tempo, en vormen nog een relatief nieuw fenomeen binnen de informatiesamenleving. Echter, de gegevensverzameling die ten grondslag ligt aan GenAI is lang niet altijd zo vernieuwend. Zo wordt gebruik gemaakt van reeds bestaande scrapingtechnieken voor het creëren van trainingsdatasets waarmee GenAI wordt ontwikkeld.

Scraping vertoont belangrijke overeenkomsten met indexeren dat plaatsvindt in het kader van het aanbieden van zoekmachinediensten. In het *GC e.a./CNIL*-arrest creëerde het Hof de ruimte voor het indexeren van webpagina’s die noodzakelijk is voor het aanbieden van zoekmachinediensten. Het Hof oordeelde dat zoekmachineaanbieders niet verplicht zijn om reeds bij het doorzoeken en indexeren van het internet te voldoen aan verplichtingen uit de AVG. Dit omdat de verantwoordelijkheid van de zoekmachine zich niet uitstrekt over hetgeen wordt verzameld van websites van derden. Pas wanneer zoekresultaten daadwerkelijk worden gepubliceerd ontstaat, volgens het Hof, een verantwoordelijkheid daarvoor.

49 Recentelijk oordeelde de Rechtbank Den Haag dat bij het gebruik van het zogenoemde Transactie Monitoringssysteem door de bank Bunq géén sprake is van uitsluitend geautomatiseerde besluitvorming vanwege zinnvolle menselijke tussenkomst. Bunq gebruikt dit systeem in het kader van cliëntenonderzoek dat zij verricht op grond van de Wwft. De rechtbank komt tot de conclusie dat sprake is van menselijke tussenkomst omdat een mens, en niet het systeem, besluit om op basis van een hit in het systeem een onderzoek te starten. Redengevend voor de conclusie dat sprake is van menselijke tussenkomst is daarnaast dat het onderzoek is uitgevoerd door medewerkers van Bunq zelf door

onder andere documenten op te vragen, zie Rb. Den Haag 9 september 2024, ECLI:NL:RBDHA:2024:14477, r.o. 4.10.

50 Hof Amsterdam 4 april 2023, ECLI:NL:GHAMS:2023:804, r.o. 3.43.

51 Hof Amsterdam 4 april 2023, ECLI:NL:GHAMS:2023:804, r.o. 3.36-3.37.

52 Hof Amsterdam 4 april 2023, ECLI:NL:GHAMS:2023:804, r.o. 3.36-3.37.

53 NOS, Kunstmatige intelligentie beschuldigt onschuldige journalist van kindermisbruik, 23 augustus 2024, vindplaats: <https://nos.nl/artikel/2534266-kunstmatige-intelligentie-beschuldigt-onschuldige-journalist-van-kindermisbruik>, geraadpleegd op 5 september 2024.

De scheiding van verantwoordelijkheden bij het indexeren van gegevens en het genereren van zoekresultaten, zoals in het *GC e.a./CNIL*-arrest, is voor de ontwikkeling van GenAI op basis van gescrapete gegevens zeer relevant. Toepassing van *GC e.a./CNIL* op scraping kan betekenen dat bij het samenstellen van trainingsdatasets voor GenAI geen sprake is van een absoluut verwerkingsverbod, maar dat de verdere verwerking en *output* aan strengere (redactionele) eisen moeten voldoen.

In haar onlangs gepubliceerde handreiking over scraping erkent de AP dat er overeenkomsten zijn tussen “zoekmachines en scrapers”, maar ziet zij geen reden op de conclusies uit het *GC e.a./CNIL* arrest ook toe te passen op scraping. Hierdoor stuurt de AP aan op de conclusie dat scraping in het kader van het ontwikkelen van GenAI-toepassingen “bijna altijd” in strijd is met het Unierecht.

Door te concentreren op de ogenschijnlijke verschillen tussen “zoekmachines en scrapen” lijkt de AP de meer relevante overeenkomsten tussen indexeren en scrapen over het hoofd te zien. En laat zij met een *de facto* verbod op scraping de balans tussen innovatie en het waarborgen van fundamentele rechten en vrijheden geheel anders uitslaan dan het Hof in *GC e.a./CNIL*.

Daarbij komt dat artikel 22 AVG sinds de arresten *Ola* en *Schufa* een belangrijkere rol kan gaan spelen bij het reguleren van *output* door GenAI-systemen. Uit de *Ola*-uitspraak en later het *Schufa*-arrest lijkt te volgen dat de *output* van een GenAI-toepassing onder omstandigheden reeds op zichzelf een uitsluitend geautomatiseerd besluit in de zin van artikel 22 AVG kan vormen, ook als een eventueel (nadelig) besluit met praktische gevolgen uitblijft. Dat kan bijvoorbeeld het geval zijn wanneer *output* bestaat uit zware aantijgingen aan het adres van een betrokkene. Artikel 22 AVG creëert in dat geval een (redactionele) plicht voor de GenAI-aanbieder die verlangt dat de betreffende *output* niet wordt gegenereerd.

Door de nadruk te leggen op de zorgplicht van de GenAI-aanbieder tijdens de *output*fase, die onder meer lijkt te volgen uit *GC e.a./CNIL* en de hernieuwde betekenis van artikel 22 AVG na de arresten *Ola* en *Schufa*, kan mogelijk een betere balans worden gevonden tussen innovatie en de bescherming van fundamentele vrijheden van betrokkenen. Met een *de facto* verbod op de ontwikkeling van GenAI lijkt die balans uit het oog te worden verloren.