

Pianola's, platen en patronen: de betekenis van art. 14 Aw in het kunstmatige tijdperk

IER 2025/44

1. Inleiding

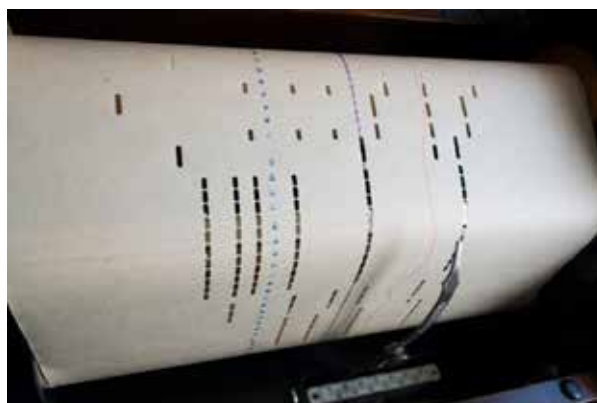
Generatieve AI roept fundamentele vragen op voor het auteursrecht. Aan de inbreukzijde is de aandacht daarbij vooral gericht geweest op het trainen van AI en dat wat eruit komt: de output. Het AI-model, dat deze twee fasen verbindt, is tot op heden minder geanalyseerd ondanks het feit dat dit een cruciale rol speelt in de productieketen en auteursrechtelijke analyse van GenAI. Deze bijdrage staat daarom stil bij de vraag of een AI-model verveelvoudigingen kan bevatten als gevolg van *machine learning* op de trainingsdataset. Centraal daarbij staat art. 14 Auteurswet en de betekenis hiervan voor het verveelvoudigingsrecht in relatie tot 'neurale netwerken' (de AI-modellen). Nadat de juridische betekenis en de historische achtergrond van dit artikel is belicht in paragraaf 2, worden in paragraaf 3 de argumenten van AI-bedrijven, waarom modellen geen verveelvoudigingen zouden bevatten, besproken en weerlegd. In de paragrafen die hierop volgen wordt stilgestaan bij de terminologische (par. 4) en praktische implicaties (par. 5) van een eventuele verveelvoudiging in het AI-model.

2. Artikel 14 en de verborgen reproductie

De wettelijke invalshoek van deze bijdrage is artikel 14, het artikel dat stelt dat iedere vastlegging van een werk – of deel daarvan – op enig voorwerp dat ertoe dient om het werk ten gehore te brengen of te vertonen,² als een verveelvoudiging is aan te merken. Om dit artikel goed te begrijpen is het nodig om de historische context te schetsen waarin deze tot stand kwam, de periode vlak voor het jaar waarin Nederlandse eerste moderne auteurswet van kracht werd, 1912 dus. Deze auteurswet breidde de typen beschermde werken aanzienlijk uit, waardoor niet langer slechts geschriften of in boeken gedrukte afbeeldingen waren beschermd maar ook beeldende kunst en muziek (los van de gedrukte bladmuziek).

Belangrijk is dat in de voorafgaande decennia de zogeheten mechanische reproducties hun intrede hadden gedaan. Muzikale werken werden vastgelegd op rollen en platen waaruit de werken slechts waarneembaar konden worden gemaakt door middel van een mechanisch hulpmiddel zoals een fonograaf.³ Verborgen reproducties dus, waarvan de populairste rond 1912 wel de pianolarol was. Niet voor niets

sprak artikel 14 Aw in de oorspronkelijke tekst van 1912 over deze (papieren) 'rollen': deze dragers bevatten specifieke patronen van perforaties (de draagtaal) die de pianola – een automatisch speelapparaat geplaatst in een piano – instrueerden welke snaren op welke momenten bespeeld dienden te worden. De mens hoefde na het plaatsen van de rol in de pianola slechts het pneumatische systeem met een voetpedaal op gang te brengen waarna de piano muziek afspelde zonder dat iemand de toetsen aanraakte.⁴



Een pianolarol (Thomas Quine, Wikicommons)

Omdat de pianola, maar ook andere '*machines parlantes ou chantantes*'⁵ zoals de fonograaf, ervoor zorgden dat mensen in steeds groteren getale steeds grotere muziekstukken konden afspelen, werd bij de herzieningsconferentie van 1908 de volgende bepaling toegevoegd aan de Berner Conventie: "Les auteurs d'oeuvres musicales ont le droit exclusif d'autoriser: 1. l'adaptation de ces œuvres a des instruments servant a les reproduire mécaniquement."⁶ Het was dit eerste lid van art. 13 uit de BC dat de basis vormde voor het Nederlandse art. 14 Aw (in het ontwerp art. 13 Aw). De aanleiding voor dit artikel uit de Berner Conventie – en daarmee 14 Aw – was dat grote fabrikanten van geluidsdragers in het buitenland, waaronder Duitsland en Amerika, met succes auteursrecht wisten te ontduiken voor hun geluidsdragers.⁷ Dit deden zij door de rechter en wetgever ervan te overtuigen dat het geestelijke werk niet aanwezig was in de pianolarol en andere geluidsdragers

1 Mr. drs. M.A. (Michiel) Smit is promovendus bij de sectie IE van de Universiteit Leiden. De auteur is prof. D.W.F. Verkade en prof. D.J.G. Visser dankbaar voor hun commentaar op een eerdere versie van dit artikel.

2 Hierna wordt dit gegeneraliseerd tot 'waarneembaar maken'.

3 De fonograaf was een voorloper van de grammofoon.

4 Was het gebruikelijke alternatief een pianist die – mede door het notenschrift te vertalen – het werk waarneembaar maakte, dat kon nu ook door een pianola gedaan worden, die op mechanische wijze het perforatieschrift uit de rol vertaalde naar klanken.

5 Actes BC 1908, p. 259.

6 Zie hierover ook: S. Ricketson & J. Ginsburg, 'The Subsequent Development of the Berne Convention, 1886–1971', *International Copyright and Neighbouring Rights: The Berne Convention and Beyond*, 3rd edn (Oxford, 2022), par. 3.17.

7 M. de Cock Buning, *Auteursrecht en Informatietechnologie* (diss.), Amsterdam: Otto Cramwinckel, 1998, p. 109-110 & 115-116.

zoals platen.⁸ Je kon het werk immers niet rechtstreeks met menselijke zintuigen waarnemen: het was slechts een rol met perforaties die in tegenstelling tot bladmuziek op zichzelf onbegrijpelijk was voor de mens.⁹ Het gevolg van deze succesvolle lobby was bijvoorbeeld dat componisten in Duitsland negen jaar alleen konden optreden tegen ouderwetse verveelvoudigingen van hun werk in de vorm van bladmuziek, maar machteloos stonden tegenover de zeer populaire *mechanische* reproducties.¹⁰

De juridische argumenten van deze fabrikanten waren op de keper beschouwd – ook destijds – nogal zwak.¹¹ De rechters en wetgevers zwichtten in landen als Duitsland en de VS echter onder de grote druk van deze invloedrijke fabrikanten.¹² Tegen deze achtergrond schreef De Beaufort in 1909 over de pianola-industrie:

“In de Vereenigde Staten b.v. heeft zich eenige jaren geleden een syndicaat gevormd van de bij deze industrie betrokken fabrikanten, welke tezamen over een kapitaal van meer dan honderd miljoen dollars hadden te beschikken. Doch juist het feit, dat bij deze exploitatie zulke aanzienlijke geldelijke belangen zijn gemoeid, heeft in sommige landen de toepassing van de juiste beginselen van het auteursrecht op dit punt eenigermate tegengehouden.”¹³

Om dergelijke kunstmatige inperkingen van het auteursrecht te voorkomen werd in Nederland dus het artikel 13 uit de BC in licht gewijzigde vorm overgenomen in de eerste auteurswet. De wetgever breidde het daarbij wel uit

“tot alle werken, die door middel van het gehoor waargenomen kunnen worden, zoodat ook letterkundige werken onder de bepaling vallen”.¹⁴ Immers was er “geen goede grond om den letterkundige een recht te onthouden dan men den musicus toekent”.¹⁵

Het resulteerde in Nederland in het volgende artikel 14 Aw:

“Onder het verveelvoudigen van een werk, dat door middel van het gehoor kan worden waargenomen, wordt mede verstaan het vervaardigen van rollen, platen en andere voorwerpen, bestemd om het geheel of een gedeelte van het werk langs mechanischen weg ten gehore te brengen.”

Dat het werk pas voor de mens waarneembaar is nadat een bepaald hulpmiddel wordt ingezet deed niet (langer) ter zake. Het reproductierecht van 1912 was in zoverre dus techniekneutraal. Naarmate in de loop der jaren ook mechanische *visuele* reproductietechnieken algemener werden, moest het artikel worden aangepast om te verduidelijken dat het op alle typen werken zag. Artikel 14 – dat de techniekneutraliteit van het reproductierecht wilde onderstrepen – was namelijk nog niet neutraal genoeg geformuleerd; het sprak over aan het gehoorzintuig appellerende ‘rollen’, ‘platen’, en ‘andere mechanische voorwerpen’. In 1972 (overigens aan de late kant dus) werd hierom het artikel gewijzigd en kreeg het de huidige vorm, waarin is verduidelijkt dat het ook om de reproductie gaat van werken die door middel van een hulpmiddel *visueel* zijn waar te nemen.¹⁶ Bovendien behoefde de vastlegging of het waarneembaar *maken* niet langer ‘mechanisch’ te zijn: het artikel sloot geen enkele techniek of methode meer uit.¹⁷

Artikel 14 luidt sinds 1972:

“Onder het verveelvoudigen van een werk van letterkunde, wetenschap of kunst wordt mede verstaan het vastleggen van dat werk of een gedeelte daarvan op enig voorwerp dat bestemd is om een werk ten gehore te brengen of te vertonen.”

Hierbij dient te worden opgemerkt dat het woord ‘bestemd’ in art. 14 Aw geen bewustheid impliceert. Dit is een in mijn ogen wat ongelukkige vertaling van “l’adaptation de ces œuvres à des instruments *servant* à les reproduire mécaniquement” (mijn cursivering), de frase uit de BC waarop het is gebaseerd. Accurater is het om ‘servant’ te vertalen als *dienend* in plaats van bestemd. Een voorwerp dat *dient* (fungeert) om een werk waarneembaar te maken is ook een verveelvoudiging. Dit sluit aan bij de feitelijke en *functionele* benadering van verveelvoudigingen en hulpmiddelen: het gaat om het resultaat (treedt er re-

8 In Duitsland werd een wet aangenomen die in beginsel alle mechanische geluidsdragers buiten het verveelvoudigingsbegrip plaatste. In de VS oordeelde de hoogste rechter in *White-Smith* dat muziekrollen geen reproducties konden zijn: *White-Smith Music Pub. Co. v. Apollo Co.*, 209 U.S. 1 (1908), “A ‘copy’ of a musical composition within the meaning of the copyright statute is a written or printed record of it in intelligible notation, and this does not include perforated rolls which, when duly applied and properly operated in connection with musical instruments to which they are adapted, produce the same musical tones as are represented by the signs and figures on the copy in staff notation of the composition filed by the composer or copyright.” K. Raustiala en C.J. Sprigman, “How a terrible Supreme Court decision about player pianos made the cover song what it is today”, <https://slate.com/technology/2014/05/White-Smith Music case: A terrible 1908 Supreme Court decision on player pianos.>, *De Cock Buning* 1998, p. 115-117.

9 *White-Smith Music Pub. Co. v. Apollo Co.*, 209 U.S. 1 (1908).

10 *De Cock Buning* 1998, p. 116-117.

11 Het ging duidelijk om doelredeneringen waarbij het nationale economische belang voorop stond. Dit was voor deze tijd overigens heel gewoon: Nederland deed dit bijvoorbeeld ook wat betreft het vertalingsrecht. De vertaling zou het oorspronkelijke werk – van de eerste auteur – niet meer bevatten dankzij de creatieve arbeid van de vertaler. Het lucratieve gevolg voor de Nederlandse boekenbranche was dat zij zonder enige royaltybetaling buitenlandse boeken vertaalde en uitgaf in eigen land. Dit is de primaire verklaring voor Nederlands late toetreding bij de Berner Conventie. Door het werk in een andere menselijke taal om te werken werd het oorspronkelijke werk dus ontkend. Nederland deed zo tot begin twintigste eeuw op ouderwetse manier in wezen hetzelfde als de pianola-industrie: zij werkten het werk alleen om in een andere menselijke taal in plaats van een mechanische taal. Het argument een effect was echter soortgelijk: het oorspronkelijke geestelijke werk was niet meer aanwezig en dus hoefden er geen royalties betaald te worden.

12 *De Cock Buning* 1998, p. 115-116.

13 H.L. de Beaufort, *Het auteursrecht in het Nederlandsche en internationale recht* (diss.), Utrecht 1909, p. 242.

14 MvT 1912, p. 10.

15 MvT 1912, art. 13.

16 Herziening van de Auteurswet, *Zitting 1971-1972*, 7877 (Ingezonden 4 februari 1972) Nr. 16.

17 Dit is zoals te lezen in paragraaf 3 van groot belang voor het AI-vraagstuk.

productie op).¹⁸ Zelfs als er per ongeluk kopieën of bewerkingen worden vastgelegd, zijn deze als verveelvoudigingen aan te merken.

Tot slot: in dit artikel is gekozen voor de insteek van artikel 14 Aw omdat dit vanuit rechtshistorisch oogpunt verhelderend is. Het historische perspectief helpt om het hedendaagse reproductierecht beter te begrijpen en om de in wezen onveranderde belangenbotsingen en daarbij gebruikte argumenten te analyseren (zie par. 3). Voor de volledigheid zij echter opgemerkt dat het reproductierecht Europees is geharmoniseerd en dat het Unierecht de hiervoor belichte beginselen inmiddels heeft geabsorbeerd. De Auteursrechtlijn 2001 stelt dat een reproductie niet direct waarneembaar hoeft te zijn en dat de wijze waarop een reproductie plaats vindt niet ter zake doet,¹⁹ zoals bevestigd in de rechtspraak.²⁰

3. De verveelvoudiging in statistische vorm: kan het model kopieën bevatten?

3.1 Inleiding

Nu de achtergrond van art. 14 Aw is geschetst, zal in deze paragraaf de vraag centraal staan of een werk ook kan zijn vastgelegd in een *neuraal netwerk*: een computermodel dat is geïnspireerd op het menselijk brein (hierna kortweg: het '(AI-)model'). Belangrijk is hierbij om te beseffen dat het AI-model onderdeel uitmaakt van een productieketen.²¹ Zeer summier komt dit erop neer dat allereerst menselijke werken in gedigitaliseerde vorm worden verzameld in datasets. Het AI-model wordt vervolgens herhaaldelijk blootgesteld aan deze data gedurende het trainingsproces. De reproducties tijdens dit trainingsproces zijn in beginsel tijdelijk.²² Als gevolg van deze training heeft het model statistische patronen en correlaties uit de trainingsdata (werken) afgeleid en opslagen in de 'parameters'²³ van het model. Dit eenmaal getrainde model en zijn parameters zijn niet tijdelijk maar in beginsel stabiel en permanent, in elk geval zolang het niet opnieuw wordt

getraind.²⁴ Omdat in dit artikel de vraag centraal staat of het *getrainde model* verveelvoudigingen bevat, zal geen verdere aandacht worden besteed aan 13a Aw en tijdelijke reproducties. Wanneer iets in het model is opgeslagen is dit in beginsel immers niet tijdelijk.²⁵

Het model is op zichzelf evenwel onbruikbaar: het moet eerst worden verwerkt in een *AI-systeem*.²⁶ GPT is bijvoorbeeld het onderliggende model van zowel ChatGPT als Co-pilot. De laatsten zijn de softwaresystemen waarin het model is verwerkt. Het model is vergelijkbaar met een motor en het systeem is de totale auto: aan de motor zijn wielen, een chassis, carrosserie en een besturingssysteem (*interface*) enz. toegevoegd zodat het model ook bruikbaar is voor de prompteur.²⁷ Wanneer dit systeem eenmaal wordt aangeboden, kan er output worden gegenereerd. Pas in dit laatste stadium van de keten komt de prompteur om de hoek kijken: deze kan instructies (prompts) aan het systeem geven die vervolgens door het onderliggende model worden verwerkt om zo output te genereren.²⁸

De cruciale onderdelen zijn dus de getrainde parameters binnen het model (ook wel het 'parametrische geheugen').²⁹ Deze hebben statistische patronen uit de trainingsdata opgeslagen en bepalen zodoende welke output het model voortbrengt – berekent – nadat het een prompt is gegeven. De grote vraag die echter resteert is *wat* er is opgeslagen in deze patronen.

3.2 Narratief AI-bedrijven: modellen slaan geen werken op, herinneren slechts patronen

Omdat AI-modellen alleen statistische patronen en correlaties kunnen opslaan, hebben verschillende AI-bedrijven de boodschap verspreid dat een model per definitie geen werken kan opslaan. Zo stelde OpenAI:

“Despite a common and unfortunate misperception of the technology, the models do not store copies of the information that they learn from. Instead, models are made up of large strings of numbers (called ‘weights’ or ‘parameters’), which software code interprets and executes. The most powerful models consist of billions

18 Dat blijkt ook uit de MvT die spreekt van de ‘capaciteit van het toestel’ om iets waarneembaar te maken, MvT 1912 bij art. 12 en 13.

19 Richtlijn 2001/29/EG art. 2: “in het uitsluitende recht, de directe of indirecte, tijdelijke of duurzame, volledige of gedeeltelijke reproductie van dit materiaal, met welke middelen en in welke vorm ook, toe te staan of te verbieden.”

20 HvJ EU 4 oktober 2011, C-403/08 en C-429/08, ECLI:EU:C:2011:631 (*Premier League*), 159; HvJ EU 16 juli 2009, C-5/08, ECLI:EU:C:2009:465 (*Infopaq*), 40-43; Rb. Amsterdam ECLI:NL:RBAMS:2011:BU6271 (*Brein/NSE*), ro 4.3.

21 Lee, Cooper & Grimmelman, ‘Talkin’ Bout AI Generation: Copyright and the Generative-AI Supply Chain’, *ArXiv* 2023, p. 5-6.

22 In beginsel. Zoals later uitgebreider betoogd, wordt sommige data wel degelijk permanent opgeslagen.

23 Deze ‘parameters’/‘gewichten’ bepalen de sterkte van de verbindingen tussen de ‘neuronen’ (rekeneenheden) in het neurale netwerk. Als het neurale netwerk eenmaal is getraind en het systeem ontvangt een prompt, dan wordt deze prompt (in afgeleide numerieke vorm) herhaaldelijk – in verschillende ‘lagen’ – verrekend door deze neuron met behulp van de getrainde parameters om zo uiteindelijk de output te genereren. A. Cooper & J. Grimmelman, ‘The Files are in the Computer: On Copyright, Memorization, and Generative AI’, *ArXiv* 2025 (<https://arxiv.org/abs/2404.12590v6>), p. 9-10; M.A. Smit, ‘Leiding en toezicht over de kunstmatige helper’, *Auteursrecht* 2025/3, voetnoot 21.

24 Een eenmaal aangeboden model wordt niet vaak opnieuw getraind. Zie hierover ook par. 4.3.

25 T.W. Dornis & S. Stober, ‘Urheberrecht und Training generativer KI-modelle’, *Nomos*: 2024, p. 54-55.

26 Lee e.a. 2023, p. 45-48. Dat neemt niet weg dat sommige modellen ook als zodanig worden aangeboden op het internet, zoals bijv. Llama en Stable Diffusion; de modellen en de daarin getrainde parameters zijn zo te downloaden.

27 Net als bij auto’s wordt de motor en het besturingssysteem niet altijd door dezelfde producent vervaardigd.

28 Een eventuele laatste stap is zogeheten *alignment*. De aanbieder van het systeem kan maatregelen en beperkingen inbouwen die voor of na het genereren worden geactiveerd. Auteursrechtelijk is met name interessant dat een aanbieder zo bijvoorbeeld namen van kunstenaars kan weigeren of wijzigen als prompt of dat inbreukmakende output uiteindelijk niet aan de gebruiker wordt medegedeeld. Lee e.a. 2023, p. 6.

29 U.S. Copyright Office, *Copyright and Artificial Intelligence: Part 3: Generative AI Training (pre-publication version)* (mei 2025), voetnoot 109: “Since knowledge is implicitly stored in the model’s parameters – or weights – it is sometimes referred to as ‘parametric’ memory.” Zie hierover ook: Patrick Lewis e.a., ‘Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks’, *Arxiv* 2021, <https://arxiv.org/abs/2005.11401>, p. 1.

of weights. Each weight roughly reflects the statistical relationship between different words in different scenarios. [...] It does not copy its response from its pre-training data, or access it via a database. Much like a person who has read a book and sets it down, our models do not have access to training information after they have learned from it.”³⁰

Het model zou kortom alleen statistische relaties bevatten in de vorm van getallen.³¹ En er is ná het trainen geen toegang meer tot de trainingsdata.³² Als er iets wordt opgeslagen in de statistische patronen, dan zijn dat onbeschermd elementen zoals feiten, abstracties en semantische relaties. De bewering dat een model geen kopieën kan bevatten rust dus op de premisse dat er een ‘schoon’ trainingsproces plaatsvindt; uit de trainingsdata – de werken – worden slechts onbeschermd semantiek, abstracties en feiten gedestilleerd. De beschermd expressie (syntaxis) – en het geestelijke werk – zou worden gescheiden van de onbeschermd elementen, en niet worden opgeslagen in het model. Gevolg daarvan zou zijn dat het model per definitie onaantastbaar is vanuit auteursrechtelijk oogpunt.

3.3 *Verveelvoudigingen in statistische patronen*

Dit standpunt is echter onvolledig, zoals het best is beschreven door Cooper en Grimmelman. Het klopt dat modellen informatie alleen opslaan in statistische patronen; maar de gelijkstelling van statistische patronen met louter onbeschermd elementen is onjuist.³³ Deze patronen kunnen behalve onbeschermd abstracties ook beschermd en concrete, zelfs gedetailleerde, expressie bevatten en permanent bewaren:

“The problem is that the ‘patterns’ learned by a model can be highly abstract, highly specific, or anywhere in between. It is a ‘pattern’ that every frame of 4K UHD video is 3840 pixels wide; a multimodal model that learns to generate images of the correct resolution when the prompt contains ‘4k uhd’ has not memorized any protectable expression. But it is also a ‘pattern’ that in the first sentence of The Restaurant at the End of the

Universe, the word ‘widely’ is followed by the word ‘regarded,’ and a model that learns enough such patterns can memorize all of Douglas Adams’s oeuvre. So even to the extent that the results of model training are made up of ‘patterns’, so is all memorization, because memorization is part of what happens during training.”³⁴

Het feit dat het model uit statistische ‘patronen’ bestaat, betekent kortom allerm minst dat het model alleen onbeschermd materie bevat. Dornis, Stober³⁵ en Sesing-Wagenpfeil³⁶ benadrukken eveneens dat modellen de concrete beschermd vorm van een werk kunnen bevatten of beschrijven. De ‘Syntaxis’ zoals dit in de Duitse literatuur veelal wordt genoemd, en dus niet slechts de onbeschermd ‘Semantik’.³⁷ Sesing-Wagenpfeil vergelijkt de verveelvoudiging in het model met codering en met een ‘pad’ dat door de training is ingesleten in het model. In het geval van ‘memoriseren’ zijn bepaalde waarschijnlijkheidsverdelingen of verbindingen met expressieve elementen zo sterk aangelegd, dat het werk in de output van het model kan worden gereconstrueerd:³⁸ “Solch ein Pfad beschreibt letztlich eine werkspezifische Syntax und kann daher – in Verbindung mit den im Modell hinterlegten Fragmenten – prinzipiell als Repräsentation eines Werkes und damit als Vervielfältigungsstück angesehen werden.”³⁹

De patronen in het model bevatten kortom de *beschermd vorm* (syntaxis) van het werk, in elk geval in die mate dat reconstructie – waarneembaarmaking – hiervan in de output mogelijk is.⁴⁰

3.4 *Fragmentatie, bouwpakketten en compressie*

Een dergelijk statistisch patroon dat de beschermd vorm van het werk beschrijft, kan overigens worden ‘geactiveerd’ door enkel de ‘juiste’ prompt in te voeren zoals de naam van een tekenaar, muzikant of fotograaf.⁴¹ Het is dus niet zo dat de prompteur het werk zelf nog moet assembleren. Dit brengt ons bij een ander mogelijk tegenargument – en de weerlegging daarvan – dat samenhangt met het hiervoor behandelde: de modellen zouden alleen *fragmenten* van werken; dat wil zeggen onbeschermd delen bevatten.

30 Commentaar OpenAI aan USCO, 1 november 2023, p. 6.

(<https://www.regulations.gov/comment/COLC-2023-0006-8906>).

31 Zie ook Googles verklaring: “The interconnections between these neurons, also referred to as nodes, are numerical weights that essentially represent the importance of the contribution of that neuron to the final output. It is also relevant to note that there is no copy of the training data – whether text, images, or other formats – present in the model itself. Deep neural networks themselves determine the attributes of the data that they use to recognize patterns, as opposed to a human coder setting those attributes manually. [...] When an LLM is pre-trained, training material is analyzed to examine and extract statistical relationships among the individual words and sentences, e.g., their frequency, importance, and semantic relationship to each other. The AI ‘model’ is simply the encapsulation of those statistical facts in numbers”. Commentaar van Google aan USCO, 1 november 2023, p. 3-4. & 5 (<https://www.regulations.gov/comment/COLC-2023-0006-9003>).

32 Commentaar OpenAI aan USCO, 1 november 2023, p. 7.

33 A.F. Cooper & J. Grimmelman, ‘The Files are in the Computer: On Copyright, Memorization and Generative AI’, 2025, p. 41-42.

34 Cooper & Grimmelman 2025, p. 41.

35 T.W. Dornis & S. Stober, *Urheberrecht und Training generativer KI-Modelle*, Nomos 2024, p. 86.

36 A. Sesing-Wagenpfeil, ‘Trainierte KI-Modelle als Vervielfältigungsstücke im Sinne des Urheberrechts’, *ZGE*2024 (16), p. 212-268, aldaar, p. 240-241.

37 Gervais maakt in andere woorden hetzelfde punt. D. Gervais, ‘The Heart Of The Matter: Copyright, AI Training, And LLMs’, *SSRN* 2024, p. 12.

38 Sesing-Wagenpfeil 2024, p. 240.

39 Sesing-Wagenpfeil 2024, p. 240-241.

40 Sesing-Wagenpfeil 2024, p. 243. Zie ook: USCO, *Part 3: Generative AI Training* (pre-publication), p. 28-29.

41 Dit is wel afhankelijk van het systeem en eventuele ‘alignement’: sommige aanbieders staan het niet toe om namen van kunstenaars te gebruiken. Anderen wel. Hoe dan ook verandert dit niets aan wat er in het model zit. Sterker nog: het feit dat dergelijke *alignement* nodig is, kan al een indicatie zijn dat het model waarschijnlijk auteursrechtelijk beschermd trekken bevat in gecomprimeerde vorm. Als die *alignement* honderd procent effectief is – wat op dit moment een illusie lijkt – is natuurlijk wel de vraag welke schade er wordt geleden.

Dit argument doet denken aan de merkwaardige zaak uit 1989 tussen V&D en Lift Verkaufsgeräte over de vraag of een CD-rek-*bouwpakket* een bewerking was in de zin van art. 13 Aw.⁴² Lift verkocht een opbergrek voor CD's in onder meer Nederland en V&D zou op het daar op rustende auteursrecht inbreuk maken door een bouwpakket voor CD-opbergrekjes te verkopen. Dit pakket bestond uit losse onderdelen waarmee het opbergrek kon worden geassembleerd naar het voorbeeld van de bijgesloten instructie. Een van de verweren van V&D was dan ook dat het verkochte product slechts een bouwpakket was 'bestaand uit ongeordende materialen' en dat deze materialen zelf niet als werk kwalificeerden. Daarvoor zou het immers eerst nog in elkaar moeten worden gezet door de consument. De rechtbank verwierp dit verweer: "Van belang is dus niet dat de losse onderdelen in het bouwpakket nog geen 'werk' vormen, maar het antwoord op de vraag of het met die onderdelen in elkaar te zetten rek – de cd-storage – in de bewoording van art. 13 Aw is te beschouwen als een "gehele of gedeeltelijke" bewerking of nabootsing in gewijzigde vorm, welke niet als een nieuw, oorspronkelijke werk moet worden aangemerkt."⁴³

Over deze 'bouwpakketten-kwestie' constateerde Verkade in zijn noot dat het verveelvoudigingsbegrip te zeer werd opgerekt door de "op zich niet inbreuk makende losse onderdelen/bouwstenen" als verveelvoudiging aan te merken "omdat daarmee het beschermde werk kán (resp. vermoedelijk zal) worden nagemaakt".

En verder: "[...] als bij deze bouwdoos een *verveelvoudiging* wordt aangenomen, dan moet dat ook gelden voor een Lego-bouwdoos met net zo veel steentjes dat je daarvan bijvoorbeeld het Centraal Station in Rotterdam kunt nabouwen (+ gebruiksaanwijzing)." [...] "Het lijkt verkieslijker bouwpakketten niet te beoordelen via een naar de rechtspraak reeds aanvechtbare toepassing van art. 13 Aw, maar via een (al dan niet) aanvullende werking van art. 1401 BW."⁴⁴

In AI-context treffen we in wezen hetzelfde 'lego-argument' aan. AI-modellen zouden slechts onbeschermd materie bevatten, namelijk losse onbeschermd grond- of bouwstoffen. Wat betreft deze elementen is het argument van Verkade ook toepasbaar op deze nieuwe technologie: het feit dat uit die op zichzelf onbeschermd materie een inbreukmakend geheel kan worden gevormd, betekent nog niet dat er een verveelvoudiging in het model zit. De enkele onbeschermd en losse bouwstoffen zijn vrij.

Zoals echter uitgelegd bevat het model niet alleen onbeschermd en 'losse' bouwstenen; het kan ook de concrete expressie van een bepaald werk latent hebben opgeslagen. In het geval van deze *memorization* – een specifieke eigen-

schap van het model – zit er wel degelijk een verveelvoudiging in het model en zijn dit niet slechts 'losse bouwonderdelen'. De prompteur kan immers met een enkele term – zoals de naam van een auteur – het werk opvragen waarna deze kant-en-klaar aan de prompteur wordt gepresenteerd. De betere vergelijking is in dit geval een gecomprimeerde databank waaruit het beschermde werk kan worden opgevraagd met de juiste *query*. In deze situaties is het dus niet juist om de verveelvoudiging in het model met terugwerkende kracht op te knippen tot losse onbeschermd elementen (bouwstoffen).⁴⁵ Dit is een argument dat AI-bedrijven gebruiken omdat zij met deze molecuulisering van het werk hun aansprakelijkheid kunnen 'wegalgoritmiseren'.⁴⁶

Een en ander wordt ook onderstreept in de Duitse literatuur.⁴⁷ Op dezelfde wijze zou anders aansprakelijkheid kunnen worden ontlopen voor bijvoorbeeld verveelvoudigingen in *clouds* of via *streaming*, aldus Schulze.⁴⁸ Ook kan worden gedacht aan peer-to-peer-openbaarmakingen waarbij door inbreukmakers wel (vergeefs) is betoogd dat er geen verveelvoudiging en openbaarmaking was omdat het slechts om het uploaden van kleine en losse onbeschermd pakketjes ging.⁴⁹

42 Rb. Amsterdam 29 november 1989, *AMI* 1990, p. 141 (*Discit cd-rek*) met noot Verkade.

43 Rb. Amsterdam 29 november 1989, *AMI* 1990, p. 141 (*Discit cd-rek*) met noot Verkade. Overigens verloor Lift het geding omdat het (eenmaal vervaardigde) opbergrek van V&D kennelijk voldoende verschilde om als een 'nieuw, oorspronkelijk werk' te worden aangemerkt.

44 Rb. Amsterdam 29 november 1989, *AMI* 1990, p. 142 (*Discit cd-rek*) met noot Verkade.

45 Zie voor een vergelijkbaar argument ook: Cooper & Grimmelman 2025, p. 27.

46 De reden dat het toch aanslaat is omdat het in andere gevallen van succesvolle 'generering' – zonder het ophoesten van beschermde trekken (generalization dus in plaats van memorization: zie par. 4) – wel valide is en omdat deze verschillende gedragingen van GAI-modellen vaak worden verhaspeld.

47 "Wenn ein Rezipient die sukzessive Übermittlung und Speicherung eines Werks (oder schutzfähiger Teile) als einheitliches Ganzes wahrnimmt, sollte es nicht darauf ankommen, wie groß die Speicherpakete zu einem bestimmten Zeitpunkt sind. Man muss den Vorgang funktional bewerten. Andernfalls ließe sich eine Nutzungshandlung allein dadurch umgehen, dass die Bruchteile des Ganzen in sehr kleine Pakete aufgeteilt werden." Dreier en Schulze, *Kommentar zum Urhebergesetz* 2025, art. 16-23.

48 Schulze, 'Werkgenuss und Werknutzung in Zeiten des Internets', *NJW* 2014, p. 721-726, aldaar 722.

49 Er zijn hier overigens parallellen te zien met de prejudiciële vragen in *Like Company* waar een van de vragen is of het voor de openbaarmakingskwalificatie uitmaakt dat de output deels AI gegenereerd is. Hier lijkt mij een zelfde functionele en techniek neutrale benadering als bij het reproductierecht gewenst. Rechtbank Budapest 10 maart 2025 (*Like Company/Google Ireland*), vragen aan HvJ EU Zaak C-250/25.



Prompt van M.A. Smit tegen Sora: 'Draw me Fred Flintstone'. 18-7-2025. AI-output: links. Een verveelvoudiging van zowel het figuur (character) als de specifieke oorspronkelijke tekening (rechts). Getekend door William Hanna en Joseph Barbera (gestorven in 2001 en 2006). Ik heb deze tekening slechts opgevraagd (besteld) en niet zelf getekend of geassembleerd. De beschermde vorm zit in het model en de output is hier bewijs van. De enige informatie die ik immers heb verschaft was de tekst: 'Fred Flintstone', zonder enige beschrijving van het uiterlijk (expressie).

3.5 Comprimeren van informatie

Verder is van belang dat AI het vermogen heeft om informatie zeer gecomprimeerd op te slaan zonder dat daarbij de auteursrechtelijk beschermde trekken echt verloren gaan.⁵⁰ Deze blijven dan latent en gecomprimeerd aanwezig waardoor we nog steeds van een verveelvoudiging kunnen spreken.⁵¹ Compressietechnieken kennen we in de digitale wereld al veel langer – denk aan JPEG, PDF en andere digitale omzettingen⁵² – maar zijn nu sterk verbeterd door onder meer de zogeheten *manifold hypothesis*.⁵³ Het feit dat AI-modellen informatie *gefragmenteerd* en *gecomprimeerd* kunnen opslaan betekent kortom niet dat geen sprake kan zijn van verveelvoudigingen in het model. Het betekent dat opslagtechnieken steeds efficiënter worden.

50 In elk geval dusdanig dat deze ook weer gereconstrueerd kan worden in de output. DeepAI, 'What is the Manifold Hypothesis?', <https://deepai.org/machine-learning-glossary-and-terms/manifold-hypothesis>.

51 Zolang de auteursrechtelijk beschermde trekken niet verloren zijn gegaan kan ondanks inkrimping gewoon van een reproductie worden gesproken. Zo ging het auteursrecht van Karel Appel niet verloren door de 'enkele verkleining' van zijn werken tot postzegelformaat. Spoor, Verkade, Visser 2025/4.17. En verder: Vzr. Rb. Utrecht 25 maart 2009, ECLI:NL:RBUTR:2009:BH7617, IER 2009/39 m.nt. MdCB, r.o. 4.4. Zie ook: Hof Amsterdam 27 juli 2010, ECLI:NL:GHAMS:2010:BN3065 (*Schuld/Maroc-nl*), r.o. 4.4.

52 Cooper & Grimmelman 2025, p. 42-43.

53 De theorie dat alle natuurlijke data in feite schuilt in laagdimensionale *manifolds* (oppervlakken) die besloten liggen binnen een hoogdimensionale ruimte. F. Chollet, *Deep Learning with Python*, 2021, p. 129-130; https://en.wikipedia.org/wiki/Manifold_hypothesis.

4. Terminologische implicaties: het 'neurale netwerk' als drager

4.1 'Machine learning', 'memoriseren' en 'generaliseren'

In het voorafgaande is vanuit het perspectief van het Nederlandse (en Europese) recht betoogd dat we van verveelvoudigingen *in* het model kunnen spreken. Daarmee wordt bepaald niet gesteld dat AI-modellen als geheel een kopie vormen of alleen maar uit verveelvoudigingen bestaan. Anderzijds wordt ook de visie van GAI-bedrijven verworpen: modellen kunnen – en zullen – wel degelijk verveelvoudigingen in statistische vorm bevatten. Hiermee wordt aangesloten bij Cooper en Grimmelman die stellen dat de waarheid in het midden ligt en dat het trainingsproces zowel tot vastlegging van onbeschermde elementen als beschermde elementen en werken leidt.⁵⁴ Het *US Copyright Office* lijkt ook deze tussenweg te volgen en stelt dat de parameters in modellen werken uit de trainingsdata kunnen bevatten.⁵⁵ Vanuit auteursrechtelijk oogpunt betekent dit dat *machine learning* een proces is waarbij informatie wordt gedestilleerd uit trainingsdata die vervolgens als statistische patronen wordt opgeslagen in het computermodel. De gewonnen en bewaarde informatie kan zowel abstract en onbeschermd zijn als gedetailleerd en beschermd. Het eerste fenomeen hangt samen met de term 'generaliseren'⁵⁶ en het tweede fenomeen met de term 'memoriseren': specifieke werken worden in het model opgeslagen ('gememoriseerd').

In dit artikel zijn beide termen echter bewust weinig gebruikt. Deze zijn onder meer verwarrend omdat deze niet precies het tegenovergestelde betekenen (al wordt dit soms wel beweerd): generaliseren gaat namelijk ook gepaard met het opslaan ('memoriseren') van (onbeschermd) informatie.⁵⁷ Het zijn verder nadrukkelijk geen juridische termen of precieze equivalenten van auteursrechtelijke begrippen zoals de reproductie. Het is misschien verleidelijk om termen als memoriseren uit te wisselen met kopiëren of verveelvoudigen maar dit zou volgens mij tot misverstanden leiden. De term 'memorization' is overgewaaid uit de computerwetenschappelijke literatuur: daar hebben onderzoekers – m.b.v. computers – gespeurd naar (vrijwel) exacte overeenstemming tussen trainingsdata enerzijds en GenAI-output anderzijds. Daarbij hebben zij echter nadrukkelijk niet met een auteursrechtelijke bril trainingswerk en output met elkaar vergeleken.⁵⁸ Reproducties die niet 100 procent identiek zijn opgeslagen, zijn daardoor

54 Cooper & Grimmelman 2025, p. 41-42 en 59.

55 USCO, *Part 3: Generative AI Training* (pre-publication), p. 19-21 en 28-29.

56 Generalisatie is het vermogen van een AI-model om goed te reageren op externe data – waarop het niet is getraind. *Glossary*, GenLaw <https://blog.genlaw.org/glossary.html#generalization>.

57 Die informatie is dan echter eerder abstract en onbeschermd. Als een model een paard moet kunnen genereren dankzij 'generalisatievermogen' moet het 'herinneren' hoe de abstractie van een paard eruit ziet. Het tegenovergestelde van generalisering is niet memorisering maar optimalisering: het fenomeen waarbij het model perfect getraind is om enkel en alleen op bekende trainingsdata te reageren. Chollet 2021, p. 122.

58 Cooper & Grimmelman 2025, p. 51-54.

mogelijk niet meegerekend. Dat geldt al helemaal voor verder afwijkende verveelvoudigingen die de jurist niettemin als bewerkingen in de zin van art. 13 zou kwalificeren.⁵⁹ Of auteursrechtelijk beschermde trekken zijn overgenomen in een auteursrechtelijk relevante mate – qua totaalindrukken – is op dit moment nog een rechterlijk en dus menselijk oordeel, niet een dat door computers kan worden gemaakt, in elk geval niet wat betreft bewerkingen.

4.2 Antropomorfismen en metaforen: herinneren of opslaan?

De gelijkenschakeling van memoriseren met verveelvoudigen is dus niet accuraat. Daarnaast zijn termen als *herinneren* – en daarmee *memoriseren* – tevens misleidend vanwege hun antropomorfe ondertoon; het zijn metaforen die verhullen dat informatie (en werken) in feite wordt opgeslagen.⁶⁰ Dergelijk taalgebruik kan dan ook doelbewust en strategisch zijn, zeker als het wordt ingezet door AI-bedrijven of politici als Donald Trump.⁶¹ Dat dergelijke metaforen vonnissen over AI en auteursrecht vergaand kunnen beïnvloeden is reeds gebleken uit de uitspraak in *Anthropic/Bartz*.⁶² Een ‘neuraal netwerk’ dat slechts zou leren, denken, weten en herinneren bevindt zich immers net als het *menselijke* brein buiten het bereik van auteursrecht. Door GenAI te vermensenlijken poneert de AI-industrie zich daarnaast als vooruitstrevend en educatief betrokken. *Leren* en *herinneren* – in plaats van opslaan en kopiëren – appelleert aan positieve en menselijke vaardigheden die we niet moeten belemmeren door krampachtig aan het auteursrecht vast te houden.⁶³

Hoewel metaforen⁶⁴ in zijn algemeenheid niet te vermijden zijn⁶⁵ – en bij een complex fenomeen als GenAI des te minder⁶⁶ – is het dus belangrijk om taalkritisch te zijn ten

aanzien van AI.⁶⁷ Met name wat betreft de belangrijkste en hardnekkigste metafoor: ‘het AI-model is een *neuraal netwerk*’. Oftewel een computerprogramma als menselijk brein. Het AI-model als ‘neuraal netwerk’ is een *systematische* metafoor:⁶⁸ het weerspiegelt een bredere en structurele manier van (menselijk) denken en gaat gepaard met allerlei samenhangende antropomorfismen zoals neuronen, synapsen, hersenlagen, hallucineren, leren, denken en dus ook herinneren en memoriseren. Belangrijk voor de auteursrechtelijke (en technische) analyse is echter om te onthouden dat het in al deze gevallen om overdrachtelijk en niet letterlijk taalgebruik gaat. Het is niet uitgesloten dat de (metaforische) hypothese van het artificiële neurale netwerk ooit door (neuro)technologische ontwikkelingen wordt verwezenlijkt⁶⁹ zodat de genoemde termen niet meer overdrachtelijk maar accuraat zijn.⁷⁰ Maar op dit moment is daar beslist nog geen sprake van: een AI-model is geen brein en de wezenlijke verschillen hier tussen zijn ook auteursrechtelijk relevant.⁷¹

59 De literatuur over *memorization* is weliswaar belangrijk maar het is waarschijnlijk geen uitputtend auteursrechtelijk onderzoek naar de omvang van verveelvoudigingen aanwezig in het model.

60 In het geval van ‘memoriatie’ dus en op een niet direct waarneembare wijze.

61 Transcript: Donald Trump’s Address at ‘Winning the AI Race’ Event’, Tech Policy Press, 24 juli 2025. (<https://www.techpolicy.press/transcript-donald-trumps-address-at-winning-the-ai-race-event/>).

62 Voor deze rechter is het trainen van AI niet anders dan een schoolkind dat ‘leert’ door boeken te lezen, zo blijkt herhaaldelijk uit dit vonnis. Dit overdrachtelijke redeneren – zich uitend in de in het vonnis gebruikte metaforen en analogieën – lijkt een doorslaggevende rol te hebben gespeeld in de beslissing dat sprake was van *fair use*. Men zou ook wel van een ‘metafoordeel’ kunnen spreken. Rechtbank Noord-Californië 23 juni 2025 (*Bartz/Anthropic*), p. 12 en 28. <https://storage.courtlistener.com/recap/gov.uscourts.cand.434709/gov.uscourts.cand.434709.231.0.pdf>; Mr. D. Visser, ‘Trainen van AI is fair use’, <https://www.mr-online.nl/trainen-van-ai-is-fair-use/>.

63 In de Amerikaanse context betekent educatief gebruik bovendien dat eerder sprake kan zijn van transformatief gebruik (de eerste *fair use* factor). In *fair use*-rechtszaken zijn dergelijke antropomorfismen dus ook zeer nuttig.

64 Een metafoor is een overdrachtelijke uitdrukking die bovendien een overdrachtelijke denk- en redeneerwijze kan weerspiegelen. Lakoff en Johnson definiëren de essentie van de metafoor als: het beschrijven en ervaren van een fenomeen in termen die eigenlijk bij een ander fenomeen horen. G. Lakoff & M. Johnson, *Metaphors we live by*, Chicago 1980, p. 5.

65 Taal is in de kern overdrachtelijk; het is nauwelijks mogelijk om een zin te uiten zonder een ‘levende’ of ‘dode’ metafoor te gebruiken. Zie ook: Lakoff & Johnson 1980.

66 Zie: Smit, ‘De verboden metafoor: AI en antropomorfe taalgebruik’, *IEF* 22777.

67 De rechtscheppende discussie rond AI – is de output beschermd – is haast uitsluitend een academische discussie. Aan de inbreukmakende kant – waarover dit artikel gaat – is er echter duidelijk bemoedigen en grote invloed van de AI-industrie die ook hun eigen metaforen dusdanig systematisch gebruiken dat deze mogelijk niet meer als dusdanig – als overdrachtelijk – worden herkend. Dit kan grote gevolgen hebben voor de auteursrechtelijke analyse.

68 Zoals betoogd door Lakoff en Johnson kunnen metaforen niet alleen toevoeging van op zichzelf staand zijn maar ook systematisch en structureel. Het woord riviermond is een meer oppervlakkige metafoor; het maakt door-gaans immers geen deel uit van een systeem waarbij rivieren en mens uitvoerig worden vergeleken. Een AI-model wordt daarentegen structureel, haast systematisch, met de mens vergeleken en vice versa. Daarom bestaat het uit een systeem van samenhangende metaforen, zoals neuronen, enz. Zie over systematische metaforen: Lakoff & Johnson 1980, p. 7-10.

69 Met name *Organoïde Intelligentie* (OI) zou het (auteurs)recht voor grote dilemma’s plaatsen omdat in deze *levende* ‘computer’ de siliconenchip vervangen is door een uit menselijke stamcellen gekweekt organoïde ‘minibrein’. Op dit moment verkeert OI echter slechts in een experimentele fase. Over OI en ‘leren’, zie bijv.: ‘Effort aims to uncover the learning and reasoning potential of brain organoids’, UC Santa Cruz 20-10-2025 (<https://news.ucsc.edu/2025/10/learning-and-reasoning-potential-of-brain-organoids/>).

70 Immers is het neurale netwerk vanuit computerwetenschappelijk en neurologisch oogpunt natuurlijk ook een hypothese: in de jaren 1940 werd al de theorie uitgesproken dat een computer ontwikkeld zou kunnen worden die hetzelfde als de menselijke hersens werkt. Het zou overigens niet de eerste keer zijn dat een metafoor wordt ‘ingehaald’ door technologische ontwikkelingen en deze dus niet langer metaforisch maar letterlijk is: denk bijv. aan het woord computer dat vroeger werd gebruikt voor menselijke rekenaars. ‘The Vocabulary: What’s the root of the word computer?’, BBC (<https://www.bbc.com/news/blogs-magazine-monitor-35428300>). M. Cobb, ‘A Brief History of Wires in the Brain’, *Frontiers* 16-12-2021.

71 Het menselijk brein is een levend en plastisch orgaan dat continu leert en verandert. Daarmee samenhangend is het menselijke geheugen bijzonder feilbaar en onvergelijkbaar met de betrouwbare machinale opslag in een computer of met het fenomeen van artificiële ‘memoriatie’. De getallen in de parameters van een eenmaal getraind AI-model liggen vast en zijn stabiel en ‘bevroren’. In tegenstelling tot mensen en hun feilbare geheugens kunnen computer(modellen) waaronder GenAI werken daadwerkelijk *opslaan* en stabiel en nauwkeurig bewaren, of deze ‘memoriatie’ nu door de ontwikkelaar is bedoeld of niet. M.S. Bennett, *A Brief History of Intelligence*, Londen 2023, p. 132; S. Miller, ‘Brains are Made of Plastic’, 5 september 2013, (<https://www.brainfacts.org/brain-anatomy-and-function/anatomy/2013/brains-are-made-of-plastic/>); ‘Lifelong learning and the plastic brain’, 19 november 2024, <https://www.cam.ac.uk/research/features/lifelong-learning-and-the-plastic-brain>; USCO, *Part 3: Generative AI Training* (pre-publication), p. 29. Over het feilbare en veranderlijke menselijke geheugen: C. Greene & G. Murphy, *Memory Lane* (Princeton 2025), p. 21, 27-28.

4.3 *Het AI-model als slimme drager*

Samenvattend, is een AI-model in tegenstelling tot het menselijk brein een levenloos en stabiel voorwerp waarin, zoals hier betoogd, werken permanent kunnen zijn vastgelegd.⁷² In dit opzicht kunnen we het ook een drager noemen. De term drager voelt niettemin wat onnatuurlijk omdat ouderwetse dragers 'dom' en 'deterministisch' zijn. *Dom* in die zin dat ouderwetse gegevensdragers slechts één functie hebben, namelijk het dragen van al dan niet beschermde informatie. Een vinylplaat, cd-rom of harde schijf kan niet zelf voortbrengselen genereren. Deze voorwerpen hebben niet de 'intelligente' capaciteiten die GenAI heeft, namelijk om 'zelf' op voorspellende wijze een nieuw voortbrengsel te doen ontstaan op verzoek van de prompteur.

Een AI-model kan dus als een drager fungeren van een interne statistische kopie die met behulp van een prompt waarneembaar kan worden gemaakt én het model kan zelf nieuwe content genereren. In de eerste situatie is het slechts een drager en in de tweede situatie is het een synthetische 'maker'.

Daar komt bij dat een AI-systeem – waarin het model is verwerkt – minder deterministisch is dan ouderwetse dragers en bijbehorende hulpmiddelen als gevolg van de stochastische aard van het genereringsproces.⁷³ Eén en dezelfde prompt kan tot verschillende uitkomsten leiden.⁷⁴ Dit stochastische effect heeft echter vooral met het genereringsproces van het 'hulpmiddel' (AI-softwarestelsel) te maken en minder met het onderliggende model (de 'drager'). Wanneer bijvoorbeeld de *seed* in het gebruikersstelsel wordt gefixeerd zal het model in principe telkens dezelfde output voortbrengen.⁷⁵

Bovendien: het feit dat dezelfde prompt mogelijk ook tot *niet*-inbreukmakende output leidt – wanneer een *seed* niet is gefixeerd – rechtvaardigt de andere inbreukmakende output allerminst.⁷⁶ Verdedigbaar is juist dat de verveelvoudigingen in het model extra schadelijk zijn doordat deze niet alleen gebruikt kunnen worden om exacte

kopieën te 'genereren' in de output maar ook talloze variërende bewerkingen of stijlimitaties.⁷⁷

Concluderend moge het duidelijk zijn dat generatieve AI het auteursrechtelijke begrippenkader uitdaagt en dat AI-modellen in diverse opzichten aanzienlijk verschillen van ouderwetse dragers. Desalniettemin zijn deze getrainde modellen op dit moment gewoon stabiele en levenloze voorwerpen die – onder meer – als drager kunnen dienen van auteursrechtelijk beschermde werken.⁷⁸ Dat betekent ook dat een werk hierin kan zijn vastgelegd zoals bedoeld in art. 14 Aw en er dus sprake is van reproductie.⁷⁹

5. **Praktische implicaties: het belang van de verveelvoudiging in statistische vorm**

5.1 *Hoe vaak worden werken uit de trainingsdataset opgeslagen in het model?*

In dit artikel is tot dusver betoogd dat werken in een statistische vorm kunnen zijn vastgelegd in AI-modellen hetgeen kwalificeert als verveelvoudiging. De vraag die echter resteert is wat nu het praktische belang is van deze verveelvoudigingskwalificatie *in het model*. Om dit uit te leggen zal ik eerst kort reflecteren op de kwantitatieve omvang hiervan en vervolgens reageren op Torenbosch' artikel over 'memorization' en auteursrecht.⁸⁰

Op dit moment wordt naar schatting tussen de 0,1 en 10 procent van de gebruikte trainingsdata in GenAI-modellen opgeslagen,⁸¹ afhankelijk van het model en de grootte daarvan.⁸² Daarbij moet direct worden opgemerkt dat deze onderzoeken dit percentage doorgaans inschatten op basis van letterlijke kopieën en dat het subjectieve auteursrecht breder is dan dat.⁸³ Het lijkt er dus op dat in elk geval bewerkingen veelal niet zijn verdisconteerd in deze onderzoeken. Ten tweede is het zo dat ook een laag percentage nog altijd in een massale inbreuk kan resulteren omdat de trainingdatasets extreem groot zijn.⁸⁴ Een theoretische illustratie: dataset LAION-5b, waarop verschillende diffusiemodellen zijn getraind, bestaat uit verwijzingen

72 AI-modellen mogen wellicht ontworpen zijn met andere doeleinden, maar feit is dat deze computermodellen wel degelijk als opslagvoorwerp kunnen fungeren. USCO, *Part 3: Generative AI Training* (pre-publication), p. 28.

73 Het genereringsproces moet niet worden verward met het onderliggende model dat immers onveranderlijk is. In het genereringsproces wordt bewust willekeur geïntroduceerd door middel van instellingen zoals *seed* en temperatuur. Het stochastische schuilt dus meer in het hulpmiddel – het systeem en de genereringswijze – dan in de drager, het model zelf (dat immers stabiel is). Zie hierover ook: Smit, 'Prompteursrecht', *AA 2025/1*; Cooper & Grimmelman 2025, p. 47-50.

74 De variëteit in de uitkomst wil dus niet zeggen dat binnen het model geen werken duurzaam en meer deterministisch zijn vastgelegd. In tegendeel: de getallen in een model liggen vast en veranderen niet, in elk geval zolang het niet opnieuw wordt getraind. Zoals Sasing-Wagenpfeil stelt kan dergelijke statistische opslag in het geheugen worden gelijkgesteld met codering. De specifieke parameters bevatten kortom de verveelvoudiging en deze parameters maken weer deel uit van het model. Smit, 'Prompteursrecht', *AA 2025/1*; Sasing-Wagenpfeil 2024, p. 240-241.

75 Dit hangt wel af van het gebruikte systeem. Alle eventuele andere stochasten en variabelen moeten ook eender zijn. Smit, 'Prompteursrecht', *AA 2025/1*, p. 42-43; Cooper & Grimmelman 2025, p. 49.

76 De prompt 'Fred Flintstone' of 'Uderzo' kan bijvoorbeeld in 80% van de gevallen tot een verveelvoudiging leiden en in 20% tot een 'nieuw, oorspronkelijk werk' in de zin van art. 13 Aw. In 80% van de gevallen is dat niettemin 100% inbreuk. Cooper & Grimmelman 2025, p. 50-52.

77 AI-systemen worden in de regel bewust aangeboden met instellingen als *seeds* en *temperatures*, die daarmee bewust variëteit in de output bewerkstelligen. Dit maakt de systemen commercieel aantrekkelijk omdat de gebruiker keuze heeft uit een in beginsel oneindig aantal opties in de output. Zie: Smit, 'Prompteursrecht', *AA 2025/1*.

78 Daarmee wordt niet het hele model als verveelvoudiging aangemerkt maar slechts de specifieke parameters daarbinnen.

79 Art. 14 ziet daarbij op de volledige beschermingsomvang: het kan ook zijn dat er bewerkingen in de zin van art. 13 Aw en niet alleen reproducties in het model zijn opgeslagen. Zie ook: USCO, *Part 3: Generative AI Training* (pre-publication), p. 28-29.

80 Tot op heden het enige uitvoerige artikel dat deze materie behandelt in de Nederlandse literatuur. Zelf heb ik me er wel kort over uitgelaten in: Smit, 'De verboden metafoer: AI en antropomorf taalgebruik', *IEF 22777*.

81 Dit wordt in de literatuur dus 'memorization' genoemd maar om voorname redenen noem ik het opslag of vastlegging.

82 Cooper & Grimmelman 2025, p. 54. Cooper en Grimmelman schrijven dat het risico op memorisering procentueel groter kan zijn naarmate de modellen op meer data zijn getraind.

83 Cooper & Grimmelman 2025, p. 53-54. Zie ook T. Dornis en N. Lucchi, 'Generative AI and the Scope of EU Copyright Law: A Doctrinal Analysis in Light of C-250/25', *SSRN 2025*, p. 16.

84 USCO, *Part 3: Generative AI Training*, p. 21.

naar 5,85 miljard afbeeldingen. Wanneer hiervan slechts 2 procent zou worden opgeslagen in het model gaat het alsnog om 117 miljoen afbeeldingen.⁸⁵ Belangrijker nog is dat de verveelvoudigingen in het model als *moederkopieën* kunnen fungeren; zonder adequaat filtersysteem kan deze moederkopie dan tot een onbeperkt aantal reproducties en bewerkingen in de output leiden (alook mededelingen aan het publiek). Elke prompteur kan immers met een enkele klik van de muis output opvragen door middel van een prompt. Een eenmaal in het model opgeslagen werk kan dus bijzonder schadelijk zijn en tot massale hoeveelheden aan dochterkopieën leiden.

5.2 Bewijslast bij maker of AI-aanbieder?

De aarzeling van Torenbosch en andere juristen⁸⁶ – om het verveelvoudigingsbegrip toe te passen op het AI-model – is ergens begrijpelijk: ook ik moet erkennen dat ik niet precies weet wat er in het model gebeurt. Sterker nog: *niemand* weet dit precies, zelfs de ontwikkelaars van de modellen hebben erkend het niet volledig te begrijpen. Juristen kunnen het dan wel schudden. Die onwetendheid is dus een *fact of life*. De grote vraag is echter voor wiens rekening deze ondoorzichtigheid moet komen: de vervaardigers en uitbaters van het model die het bewust zo hebben ontworpen of de schrijvers en kunstenaars wier werken – veelal zonder (expliciete) toestemming – voor de training zijn gebruikt? In dit licht acht ik de terughoudendheid van onder meer Torenbosch niet wenselijk, ik herhaal uit zijn artikel:

“Op basis van de publicaties van Gervais en Dornis en Stober is mij niet duidelijk hoe Gen AI-modellen permanente kopieën maken van hun trainingsdata, en waar die permanente kopieën zich dan bevinden in de modellen.”[...] “Hoewel het mij in het licht van de huidige stand van de computerwetenschap zeer onwaarschijnlijk lijkt dat memorization het gevolg is van het maken van reproducties van de trainingsdata, is het niet uitgesloten. Het is dan wel zaak om aan te tonen hoe Gen AI-modellen permanente kopieën maken van hun trainingsdata, en waar die permanente kopieën zich dan bevinden in de modellen.”⁸⁷

85 Onbekend is op hoeveel werken modellen exact zijn getraind en of het inderdaad om miljarden gaat. Duidelijk is wel dat het bij modellen zoals Dall-E en Midjourney minstens om honderden miljoenen afbeeldingen gaat. In het geval van Dall-E worden schattingen van 400 miljoen gemaakt. Ook dan gaat het dus nog om enorme aantallen. C. Schuhmann, ‘*Laion-400-million open dataset*’, (<https://laion.ai/blog/laion-400-open-dataset/>).

86 A. Guadamuz, ‘Disney and Universal sue Midjourney for copyright infringement’, *Technollama* <https://www.technollama.co.uk/disney-and-universal-sue-midjourney-for-copyright-infringement>: “But Midjourney does not store those images and serves them, image models know what a stormtrooper looks like and can reproduce it. Is it storage? I don’t think so”. En de rechtbank Hamburg is terughoudend omdat het volgens haar niet duidelijk is waar het verschil ligt tussen informatiewinning uit de trainingswerken en het opslaan van de geestelijke werken. Hier lijkt de rechtbank ook zekerheid te eisen over de manier waarop en waar precies een verveelvoudiging in het model – als gevolg van training – plaatsvindt. LG Hamburg 27 september 2024 (*Kneschke/LAION*), r.o. 72.

87 J. Torenbosch, ‘Memorization en generatieve AI’, *Auteursrecht* 2025/2, p. 69.

Hoewel Torenbosch’ artikel met name betrekking heeft op het technische fenomeen van ‘memorization’ en hij net als ik schrijft vanuit het perspectief van de academicus, is het toch van belang om stil te staan bij de gevolgen wanneer deze benadering *in de praktijk* wordt gevolgd. Dit zou er immers toe leiden dat de rechthebbende nu juist de meest schadelijke verveelvoudiging moet kunnen aanwijzen *in het model zelf* en ook nog eens het technologisch procedé hiervan dient te verklaren. Dit is rondit onmogelijk en leidt ertoe dat rechthebbenden slechts kunnen ageren tegen de inbreukmakende *output*; zij zouden kortom slechts aan symptoombestrijding kunnen doen terwijl de onderliggende bron hiervan – de moederkopie in het model – onaantastbaar blijft.⁸⁸ Daarmee samenhangend zal ook de aanbieder en ontwikkelaar de dans veelal ontspringen enkel en alleen omdat zij een technologie hebben ontwikkeld en exploiteren die dusdanig ontoegankelijk en complex is dat de werken hierin niet rechtstreeks kunnen worden waargenomen.⁸⁹

We dienen kortom ook voorzichtig te zijn in onze voorzichtigheid: indien we van rechthebbenden vragen dat zij met zekerheid een verveelvoudiging kunnen aanwijzen in het model en zelfs om te bewijzen hoe dit technisch is geschied, heeft dit ingrijpende consequenties qua aansprakelijkheid en bewijsrecht. Ten faveure van AI-bedrijven en ten detrimente van de menselijke creatief. Het zou effectieve auteursrechtelijke bescherming in het kunstmatige tijdperk serieus kunnen ondermijnen.⁹⁰ Los van de onwenselijkheid van een dergelijke benadering is deze zoals betoogd in dit artikel niet in overeenstemming met het geldende auteursrecht, met name vanuit het perspectief van art. 14 Aw en het reproductierecht: een werk hoeft niet rechtstreeks waarneembaar te zijn.

Verder is het evenmin te rijmen met de Nederlandse rechtspraak betreffende de bewijslast van (niet) ontlening. Uit *Barbie*⁹¹ en *Accordo/Tros*⁹² volgt dat wanneer er sprake is van opvallende overeenstemming – dusdanig dat een onafhankelijke dubbelschepping onwaarschijnlijk is⁹³ – er géén stelplicht van ontlening voor de eiser geldt en de bewijslast wordt omgekeerd. Dit laatste komt erop neer dat de beweerdelijke inbreukmaker *niet*-ontlening dient te bewijzen. In de context van GenAI betekent dit dat het niet aan de rechthebbende is om te bewijzen hoe het ontleningsproces precies heeft plaatsgevonden maar dat de ontwikkelaar en/of aanbieder van het model/sys-

88 Zeer vergelijkbaar was al de analyse van Pfeffer en Gerbrandy ten aanzien van verveelvoudigingen in de computer: H. Pfeffer, *Kort Commentaar op de Auteurswet 1912*, bewerkt door S. Gerbrandy, Haarlem: Bohn 1973, p. 144.

89 Een veelvoorkomende manier waarop AI-aanbieders en ontwikkelaars aansprakelijkheid proberen te ontlopen is door de schuld op de gebruiker van het systeem – de prompteur – af te schuiven. De inbreukmakende output zou het resultaat zijn van diens prompts (en niets te maken hebben met het model). Ook al is dit argument in veel situaties klaarblijkelijk onjuist, namelijk als de prompteur niet zelf de expressie aanlevert (zoals bij een simpele tekstprompt).

90 Zie over deze effectiviteit ook: HR 21 februari 1992, ECLI:NL:HR:1992:ZC0513 (*Barbie*).

91 HR 21 februari 1992, ECLI:NL:HR:1992:ZC0513 (*Barbie*), r.o. 4.4.

92 HR 29 november 2002, ECLI:NL:HR:2002:AE8456 (*Accordo/Tros*), r.o. 3.3.

93 Aldus samengevat door Spoor, Verkade & Visser 2025/4.14.

teem⁹⁴ moeten bewijzen dat de verveelvoudiging in de output het resultaat is van puur toeval en niet van ontle-ning (oftewel: training op het werk van eiser).⁹⁵ Indien in de output een trainingswerk wordt verveelvou-digd is het overigens *extreem* onwaarschijnlijk dat dit het resultaat van puur toeval⁹⁶ is en dat er geen causaal ver-band is met de trainingsfase en het onderliggende mo-del.⁹⁷ De genoemde Nederlandse jurisprudentie zal dan, in mijn ogen terecht, tot de conclusie moeten leiden dat een verveelvoudiging in de output ontleuning aan een werk uit de trainingsdata veronderstelt.⁹⁸ Hetgeen in veel gevallen alleen maar verklaard kan worden door een latent aanwe-zige, maar niettemin: *aanwezige* verveelvoudiging in het model; het werk is tijdens de trainingsfase op gecompri-meerde wijze opgeslagen door het model ('gememori-seerd') waarna het wordt opgehoest in de output ('regur-gitatie').⁹⁹ De verborgen 'statistische' verveelvoudiging in het model dus, zoals besproken in paragraaf 3.



Het AI-systeem Sora heb ik de enkele prompt 'Uderzo' gege-ven. Keer op keer komen daar vervolgens perfecte verveel-voudigingen van Asterix en Obelix uit. Moeten de erven van Uderzo nu zelf bewijzen hoe het onderliggende neurale net-werk van Sora dit exact doet, of is het juist aan Sora om de niet-ontleuning (niet-training) te bewijzen? Het laatste lijkt mij de juiste route. Overigens kan de prompteur misschien mede-aansprakelijk worden gehouden voor de verveelvou-diging in de output, maar in elk geval niet voor de onderlig-gende verveelvoudiging in het model.

5.3 De outputbenadering en de TDM-exceptie

Hoewel de verveelvoudiging in statistische vorm (in het model) voor alle processtrategieën¹⁰⁰ belangrijk kan zijn, alsook bij schikkingsonderhandelingen,¹⁰¹ beperk ik mij hier tot de benadering waarbij deze interne kopie het meest relevant lijkt: de outputbenadering. Deze benadering begint – in tegenstelling tot de inputstra-tegie – niet aan het begin maar aan het einde van de pro-ductieketen:¹⁰² dat wat uit de machine komt en volgens de

94 Afhankelijk van wie wordt aangesproken. Bij open-source aangeboden modellen zoals Stable Diffusion of Llama zal dit soms alleen de aanbieder en ontwikkelaar van het model kunnen zijn. Bij in systemen verwerkte modellen lijkt de aanbieder van het systeem het meest logische aanspreekpunt.

95 Een ander mogelijk verweer is dat de aanbieder van het systeem stelt dat de prompteur zelf het werk heeft samengesteld uit onbeschermde elementen – denk aan het lego-voorbeeld uit paragraaf 3 – of dat de prompteur het werk zelf heeft geupload of via RAG heeft opgevraagd. Daarmee zou een aanbieder soms kunnen bewijzen dat de expressie in de output niet het gevolg is van expressie in het model. De aanbieder moet kortom een 'non-parametrische' externe bron van de expressie aanwijzen en bewijzen.

96 Zie ook: Sesing-Wagenpfeil, p. 238-239; Dornis & Stober 2024, p. 53. Het bewijzen van niet-ontleuning is voor een mens met een biologisch brein al moeilijk. Voor de aanbieder van een model dat mogelijk op het gehele internet is getraind lijkt dit nog veel moeilijker.

97 De trainingsdatasets kunnen hierbij een belangrijke rol spelen. Indien betrouwbaar kan worden bewezen dat een model niet op een werk is getraind – hetgeen gezien de toenemende schaal van deze modellen steeds minder aannemelijk lijkt – dan lijkt daarmee het toevalsverweer te kunnen slagen. Het is echter zeer de vraag of op een betrouwbare manier kan worden bewezen waar het model wel en niet op is getraind, zeker gelet op het feit dat deze modellen op steeds meer data zijn getraind en gebruikmaken van datasets die praktisch het hele internet kunnen bevatten. Interessant is wel om op te merken dat de vernietiging van een dataset – zoals thans gebeurt bij Anthropic naar aanleiding van een schikking met Bartz – bewijsrechtelijke consequenties kan hebben voor de (niet)-ontleuning; een potentieel bewijsstuk wordt er immers mee vernietigd.

98 Oftewel: de inbreukmakende output bewijst dat het model én is getraind op dit werk én dat die training heeft geresulteerd in een duurzame verveelvoudiging in het model. Ontleuning dus.

99 Alleen wanneer de prompteur zelf de beschermde vorm construeert of deze door een derde wordt geupload – RAG bijv. – is een andere conclusie mogelijk.

100 Denk met name ook aan de inputbenadering: eisers stellen dat het ongeautoriseerde trainen van AI op hun werken inbreukmakend is omdat het getrainde model zal concurreren met hun werken, zelfs al worden er geen verveelvoudigingen in de output gedeeld. De *Anthropic/Bartz* zaak leert ons minder dan gehoopt omdat deze zaak uiteindelijk is geschikt naar aanleiding van de 'shadowlibrary' die Anthropic aanlegde. Het ging hierbij niet om de output of het model zelf maar vooral om de illegale datasets die werden gebruikt om het model te trainen en die ook daarna bleven bestaan. Deze bestonden uit o.a. 7 miljoen illegaal gedownloade boeken. In combinatie met het Amerikaanse fenomeen van 'statutory damages' resulteerde dit in een recordschikking. De input-strategie is desalniettemin – zeker in Europese context en zolang er geen illegaal verkregen data is gebruikt – hoogst onzeker voor rechthebbenden, met name doordat het zeer moeilijk zal zijn om aan te tonen dat er afbreuk wordt gedaan aan de normale exploitatie van iemands werken door een AI-model dat tevens op miljoenen of miljarden andere werken is getraind. Het werk of oeuvre van één auteur is slechts een druppel in een oceaan van data. Het schaalvoordeel dat GenAI-bedrijven hebben dankzij enorme trainingdatasets vertaalt zich dus in een causaliteitsprobleem voor de individuele rechthebbende. Het is dan ook zeer de vraag of de bredere – 'indirect substitution' – uitleg uit *Meta/Kadrey* in Europa zal worden gevolgd in de context van de driestaptoets. Zie over normale exploitatie en de driestaptoets ook: M. Senftleben, 'TDM, GenAI and the Copyright Three-Step Test, IIC 2025', *SSRN* 2025, p. 14.

101 Aangezien *memorization* bij alle grote AI-modellen voorkomt is het denkbaar dat dit ook in de *Anthropic/Bartz*-zaak een rol heeft gespeeld. Door voor dit recordbedrag te schikken heeft Anthropic dan voorkomen dat ook het model zelf (deels) vernietigd of aangepast zou moeten worden. In plaats daarvan moeten alleen de illegale datasets worden vernietigd.

102 Zie par. 3.1 voor deze productieketen, en in het bijzonder: Lee e.a., 2023, p. 5.

eiser een verveelvoudiging is van een van de werken waarop het model is getraind. De reeds toegelichte causale samenhang tussen de dochterkopie in de output en de moederkopie in het model verklaart waarom eisers in verschillende Amerikaanse zaken zich nadrukkelijk op kopieën in het model richten. Dit gebeurt in onder meer: Disney tegen Midjourney,¹⁰³ New York Times tegen OpenAI,¹⁰⁴ Getty tegen Stability AI¹⁰⁵ en Warner Bros tegen Midjourney.¹⁰⁶ In al deze zaken toonden eisers hoe de AI-systemen verveelvoudigingen produceerden in reactie op simpele prompts, die volgens hen louter te verklaren zijn door onderliggende kopieën in het model zelf. De New York Times vordert daarom zelfs de vernietiging van het AI-model.¹⁰⁷ Deze outputbenadering is in Europa vermoedelijk kansrijker dan de inputbenadering omdat er in Europa geen schade meer hoeft te worden bewezen om van inbreuk te spreken én omdat het naar mijn inschatting direct alle beroepen op beperkingen door AI-aanbieders¹⁰⁸ onschadelijk maakt. Wanneer er wordt uitgegaan van de – overigens omstrede¹⁰⁹ – gelijkstelling van tekst- en datamining met GenAI-training (*machine learning*), dan is de belangrijkste beperking uiteraard de TDM-exceptie. In dit artikel wordt inderdaad uitgegaan van de in *Kneschke/Laion*¹¹⁰ aangenomen toepasselijkheid van deze exceptie op het *trainen* van GenAI. De vraag die dit evenwel meteen oproept is hoe ver deze exceptie reikt, oftewel in welke fasen verveelvoudigingen hierdoor kunnen worden gerechtvaardigd. In het geval dat er kopieën achterblijven in een *getraind* model – dat doorgaans *open-source* wordt aangeboden op het internet of wordt verwerkt in een systeem – lijkt mij dat de TDM-exceptie *niet* van toepassing kan zijn.¹¹¹ Wanneer een kopie na de TDM in de parameters beklijft als gevolg van de training, is dat immers bewijs dat er geen 'schone' training

(en TDM) heeft plaatsgevonden. Hierbij is het tweede lid van art. 15o relevant: "De reproductie gemaakt overeenkomstig het in het eerste lid bepaalde mag worden bewaard zolang dit nodig is voor tekst- en datamining". Als we tekst- en datamining gelijkstellen met *machine learning* dan volgt daaruit dat de reproducties verdwenen moeten zijn na het afronden van de training. Deze mogen dus ook niet in een andere (wiskundige) draagtaal permanent achterblijven in het model.¹¹² Andere beperkingen dan de TDM-exceptie lijken in deze context ten aanzien van de aanbieder of ontwikkelaar van AI-modellen en systemen niet relevant.¹¹³

In Europa is het eerste antwoord op dit vraagstuk te verwachten in de zaak van GEMA tegen OpenAI. In reactie op simpele prompts gaf het taalmodel van OpenAI beschermde liedteksten weer, hetgeen volgens GEMA bewijs is van interne kopieën in het model. OpenAI erkent dat het model op deze teksten is getraind maar verweert zich op de manier zoals uitgelegd in paragraaf 3 en stelt dat modellen geen trainingsdata op kunnen slaan vanwege de statistische aard van het productieproces: de output is slechts het resultaat van een "sequenziel-analytisch, iteratief-probabilistisch Synthese".¹¹⁴ Het 'verklaart' de inbreukmakende output door naar de prompteurs en hun malafide prompts te wijzen, waarmee het duidelijk de aandacht en blaam probeert weg te leiden van het eigen AI-model en dus van het eigen handelen.¹¹⁵

Een uitgebreider en definitief antwoord van het HvJ EU volgt mogelijk naar aanleiding van de prejudiciële vragen rond de zaak 'Like Company'. Een deel van deze vragen

103 "Midjourney's process of training and making multiple copies of Plaintiffs' Copyrighted Works was done without Plaintiffs' approval or authorization. In other words, Midjourney used software, servers, and other technology to store and fix data associated with Plaintiffs' Copyrighted Works in such a manner that those works are thereby embodied in the model, from which Midjourney is then able to generate, reproduce, publicly display, and distribute unlimited 'copies' of Plaintiffs' works as defined by the Copyright Act.". Disney e.a. tegen Midjourney, complaint 11 juni 2025 bij United District Court Central District of California, p. 56.

104 NYT tegen OpenAI e.a., complaint 27-12-2023 bij United District Court Southern District of New York, p. 29-32.

105 Getty Images US tegen Stability AI, complaint 14-8-2025 bij United District Court Northern District of California.

106 Warner Bros e.a. tegen Midjourney, complaint 9-4-2025 bij United District Court Central District of California.

107 NYT tegen OpenAI, p. 68.

108 Het beroep van een *gebruiker* van een AI-systeem kan mogelijk anders uitpakken; bijvoorbeeld als iemand een parodie maakt met behulp van generatieve AI.

109 De kritiek op deze gelijkstelling groeit (onder meer) omdat TDM iets wezenlijk anders zou zijn dan het trainen van een GenAI-model. Het eerste zou een puur analytisch oogmerk hebben terwijl ML voor GenAI resulteert in een model dat content kan genereren die concurreert met de makers van de trainingswerken. Zie bijvoorbeeld: T. Dornis & N. Lucchi, 'Generative AI and the Scope of EU Copyright Law: A Doctrinal Analysis in Light of C-250/25', SSRN 2025.

110 LG Hamburg 27 september 2024 (*Kneschke/LAION*), ro. 79-80. De Duitse rechter baseert dit primair op de AI-verordening.

111 Dit is mijn interpretatie: zekerheid zullen we pas hebben wanneer het HvJ een uitspraak doet. Zie ook Torenbosch 2025, p. 67.

112 Al helemaal niet omdat deze moederkopie er zoals gezegd toe kan leiden dat er in de output dochterkopieën worden openbaargemaakt.

113 Zij kunnen zich niet op beperkingen beroepen zoals de parodie-exceptie omdat deze er zijn omwille van de vrijheid van meningsuiting. Aanbieders doen zelf geen uiting maar exploiteren slechts een middel waarmee dergelijke uitingen gemaakt kunnen worden door derden. De individuele gebruikers zouden mogelijk wel een beroep kunnen doen op de parodie-exceptie maar dat lijkt mij slechts relevant in de relatie tussen prompteur en rechthebbende. Dit rechtvaardigt niet zonder meer de kopie die in het model is vastgelegd door de aanbieder van het model, die immers in beginsel slechts met commercieel oogmerk zal zijn gemaakt: het aanbieden en uitbaten van een AI-systeem. Dit alles geldt mutatis mutandis voor de citaatexceptie. De tijdelijke reproductie is verder klaarblijkelijk niet toepasselijk bij de duurzame opslag van kopieën in een model. Hoe dan ook is het de vraag of deze beperking vanuit wetssystematisch oogpunt nog een rol mag spelen als inderdaad de TDM-exceptie de aangewezen beperking is voor machine learning bij GenAI.

114 Daarmee focust OpenAI bewust op het 'generaliserende' vermogen en het genereringsproces door het systeem. Zoals uitgelegd neemt dat echter allerm minst weg dat in het onderliggende model dat daarvoor wordt gebruikt auteursrechtelijke beschermde trekken zijn opgeslagen ('gememoriiseerd') in een gecompriëerde vorm, die in de output kunnen worden weergegeven. Dit is volgens mij dus geen valide argument maar slechts een afleidingsmanoeuvre, net als het wijzen naar de gebruikers. Zie hierover: Cooper & Grimmelmann 2025, p. 4-5.

115 Pressemitteilung 9 vom 29.09.2025 'Verhandlung GEMA geg. OpenAI', Landgericht München I. <https://www.justiz.bayern.de/gerichte-und-behoerden/landgericht/muenchen-1/presse/2025/9>.

– in zoverre geen sprake is van RAG¹¹⁶ – heb ik vanuit mijn optiek reeds beantwoord in dit artikel.¹¹⁷

5.4 Samenvattend

Centraal in dit artikel stond niet het AI-systeem maar het onderliggende AI-model. Daarom wordt hier niet stilgestaan bij de verplichtingen van de aanbieder van het systeem,¹¹⁸ maar louter bij de rol die het model kan spelen in de verhouding tussen aanbieders en rechthebbenden.¹¹⁹

Of het maatschappelijk wenselijk en in een concreet geval proportioneel is om een AI-model (deels) te verwijderen of offline te halen, is natuurlijk de vraag. Om een gewenst evenwicht tussen GenAI-aanbieders en rechthebbenden te bewerkstelligen lijkt het niettemin belangrijk dat de bestaande auteursrechtelijke beginselen consequent en techniekneutraal worden toegepast, ook ten aanzien van generatieve AI-modellen. Zelfs wanneer de 'nucleaire' maatregel – een sanctie met betrekking tot het model – nooit wordt ingezet, heeft de enkele mogelijkheid hiervan mogelijk al een corrigerende werking. Dat wil zeggen: het dwingt AI-bedrijven ertoe om zich beter aan het auteursrecht te houden in de omzomende

fasen: het rechtmatig trainen van AI en het filteren van inbreukmakende output of afsluiten van licenties.

6. Conclusie

In dit artikel is allereerst beschreven dat art. 14 Aw een reactie was op dragerfabrikanten die in het verleden succesvol auteursrecht ontrokken door het geestelijke werk te maskeren als louter patronen, perforaties, groeven en andere technische deeltjes van de drager. Omdat het werk in pianolarollen en andere mechanische dragers niet direct waarneembaar was, werd de aanwezigheid ervan ontkend en werden componisten niet betaald.

Vervolgens is uiteengezet met welke vergelijkbare argumenten de fabrikanten en aanbieders van GenAI-modellen eveneens het auteursrecht – het verveelvoudigingsrecht in het bijzonder – trachten te ontgaan. Zoals hier is betoogd in met name par. 3 zijn deze argumenten naar het geldende recht te weerleggen en kunnen we in het AI-model wel degelijk van verveelvoudigingen spreken. Ook al kunnen we die niet direct waarnemen en de technologie zelfs niet volledig begrijpen. Artikel 14 Aw speelt daarbij rechtshistorisch een centrale rol omdat dit al 113 jaar onderstreept dat het reproductierecht techniekneutraal is en het dus niet uitmaakt op welk voorwerp of in welke draagtaal de vastlegging geschiedt. De vastlegging van een werk in statistische en wiskundige taal of vorm is evengoed een verveelvoudiging, zolang deze waarneembaar kan worden gemaakt in de output. De toepassing van het verveelvoudigingsbegrip op delen van een GAI-model zal door sommigen als een (ongewenste) uitbreiding van het auteursrecht worden gezien. Naar mijn inzien is het echter een correctie op een (dreigende) kunstmatige inperking daarvan, zoals dit ook in het verleden is gebeurd naar aanleiding van mechanische reproducties. Art. 14 Aw gaat in tegenstelling tot art. 13 Aw niet om de beschermingsomvang – of het uitbreiden daarvan – maar om de levensvatbaarheid van het auteursrecht in een wereld waarin (draag)technologieën continu veranderen. Zonder dit beginsel van techniekneutraliteit zou het auteursrecht binnen korte tijd grotendeels obsoleet kunnen worden omdat auteurs hun rechten niet meer kunnen handhaven. Hoewel het klopt dat art. 14 Aw strikt genomen overbodig is (en altijd is geweest!), namelijk wanneer het reproductierecht perfect wordt toegepast, leert het verleden en heden ons dat dit bij nieuwe technologieën veelal niet direct gebeurt. Dit is precies waarom art. 14 ooit is ontstaan en waarom het nu weer uiterst actueel is. Zoals De Beaufort

116 Taalmodellen worden tegenwoordig veelal aangeboden in (multimodale)-systemen die ook Retrieval Augmented Generation gebruiken. Dankzij deze techniek wordt actuele informatie eerst opgezocht op doorgaans het internet en die tekst wordt dan – buiten het zicht van de prompteur om – toegevoegd aan de originele prompt. Deze gelardeerde prompt gaat vervolgens het AI-model in waarna een antwoord wordt gegenereerd. In deze situaties kunnen auteursrechtelijk beschermde trekken zowel uit het parametrische geheugen van het model afkomstig zijn als van de externe RAG-geraadpleegde bron op het internet. Dit maakt de situatie dus complexer en vooral relevant voor LLM-modellen, minder voor diffusie of muziekmodellen. Wanneer via RAG beschermde expressie wordt weergegeven in de output is de aanbieder van het systeem overigens alsnog aansprakelijk voor deze reproductie en mededeling aan het publiek, naar mijn inzien. Expressie die via RAG van het internet is getrokken kan echter niet als bewijs dienen voor een verveelvoudiging in het model. De TDM-exceptie lijkt RAG overigens niet te kunnen rechtvaardigen omdat er überhaupt niet wordt getraind: er wordt informatie van het internet toegevoegd aan een prompt om vervolgens op basis daarvan output te genereren. RAG is dus geen training van AI. Het gaat hier primair om de vraag of een nieuw publiek wordt bereikt (mededeling aan publiek). Zie hierover: Dornis & Lucchi 2025.

117 Immers heb ik betoogd dat we van reproducties in het model kunnen spreken, maar ook dat de gehele AI productieketen in acht moet worden genomen. Het is te hopen dat het HvJ de verschillende fasen en niveaus waarop reproducties plaats kunnen vinden helder uiteenzet en daarbij ook rekening houdt met RAG, dat deze analyse nog ingewikkelder maakt. Verder heb ik betoogd dat de TDM-exceptie niet gebruikt kan worden om verveelvoudigingen in het model zelf te rechtvaardigen en al helemaal niet om verveelvoudigingen in de output te rechtvaardigen.

118 De aanbieders van AI-systemen, waarin het model is verwerkt, zullen in sommige opzichten vergelijkbaar zijn met internet tussenpersonen zoals Youtube, waarmee 29c Aw ook relevant lijkt. In andere opzichten zijn GenAI-aanbieders vergelijkbaar met aanbieders van zoekmachines; dankzij RAG groeien GenAI en zoekmachines steeds meer naar elkaar toe. ChatGPT en Google AI Overview kunnen in feite reeds als *generatieve zoekmachines* worden beschouwd. Tegelijkertijd gaan GenAI-aanbieders verder dan enkele zoekmachines en platformaanbieders in die zin dat zij in het geval van 'memorization' zelf inbreukmakende werken hebben vastgelegd in een model, vervolgens aanbieden en openbaar maken – dat laatste op verzoek van een prompt. In dit opzicht heeft het weer meer weg van The Pirate Bay.

119 Voor de duidelijkheid: wanneer inderdaad wordt aangenomen dat er in een model verveelvoudigingen aanwezig zijn, heeft dit grote gevolgen voor de rest van de productieketen. Het impliceert zoals gezegd dat de training niet voldoet aan de TDM-exceptie maar ook dat de aanbieder van het systeem aansprakelijk kan worden gehouden voor deze verveelvoudigingen, in elk geval wanneer deze niet gefilterd worden.

al in 1909 waarschuwde, sneuvelen auteursrechtelijke beginselen juist als de 'geldelijke belangen' groot zijn. En op dit moment zouden de financiële belangen rond generatieve AI nauwelijks groter kunnen zijn.¹²⁰

120 De verwachte inkomsten van OpenAI alleen worden geschat op 12.7 miljard dollar in 2025. De *marktwaarde* van het bedrijf is in augustus 2025 geschat op 500 miljard dollar, een bedrag dat vergelijkbaar is met het BBP van landen als Noorwegen en Bangladesh. D. Kerr, 'OpenAI eyes world's largest valuation for private company in stock sale talks', *The Guardian* 20-8-2025 https://www.theguardian.com/technology/2025/aug/19/openai-chatgpt-stock-sale-reports?utm_source=chatgpt.com; 'OpenAI's annualized revenue hits \$10 billion, up from \$5.5 billion in December 2024', *Reuters* 10-6-2025. Zie met name ook het Stargate Project: https://en.wikipedia.org/wiki/Stargate_LLC.