

UKB SCENARIO'S TEN AANZIEN VAN GPT-NL DEEL 2

Publieksversie

Van: UKB Werkgroep Auteursrecht (UKB WA)
Aan: UKB Plenair
Datum: 16 december 2025

SAMENVATTING

Verzoek UKB plenair: kom met een aanvulling op het advies "UKB SCENARIO'S TEN AANZIEN VAN GPT-NL" waarin de voorwaarden waaronder het gebruik van overige wetenschappelijke publicaties en content in onze repositoria voor het trainen van taalmodellen toegestaan is.

Overige wetenschappelijke publicaties en content betreft de onderdelen:

- scripties (bachelor/master);
- dissertaties; en
- artikelen, conference papers en hoofdstukken uit verzamelbundels die gepubliceerd zijn niet in open acces maar onder artikel 25fa Auteurswet (Taverne amendement).

Bij het opstellen van de voorwaarden, neem ook het databankenrecht in overweging.

Advies Werkgroep Auteursrecht:

Auteursrecht

Auteursrecht is een grondrecht dat toekomt aan de maker. Dit recht is verankerd in internationaal recht, Europees recht en nationaal recht. De ontwikkelaars van GPT-NL erkennen dit en nemen dat als hun uitgangspunt. Zij geven aan hun taalmodel uitsluitend te trainen met content waarvan de dataprovider auteursrechthebbende is of waarvoor deze over een licentie beschikt van de auteursrechthebbenden (zie punt 1 Advies en Bijlage 1).¹ Universiteiten zijn niet de auteursrechthebbenden van de wetenschappelijke publicaties en overige content in de repositoria van de universiteiten.

Scripties

Al in repositoria gepubliceerde scripties van studenten mogen niet gebruikt worden ter training van GPT-NL, tenzij de universiteit (of de ontwikkelaars van GPT-NL zelf) individuele toestemming van de betreffende oud-studenten verkrijgt. De eindgebruiker heeft op dit moment slechts een alleen-lezen recht.

Toekomstige scripties van studenten mogen pas gebruikt worden ter training van GPT-NL nadat (1) de licentie die studenten voorgelegd wordt is aangepast en (2) een workflow is ingericht om de opt-out van studenten te faciliteren.

Dissertaties

Dissertaties van promovendi die al in repositoria zijn gepubliceerd mogen niet gebruikt worden ter training van GPT-NL, tenzij de universiteit (of de ontwikkelaars van GPT-NL zelf) individueel toestemming verkrijgt van de betreffende oud-promovendi of van de uitgever indien het auteursrecht inmiddels aan een uitgever is overgedragen. De eindgebruiker heeft op dit moment slechts een alleen-lezen recht.

Toekomstige dissertaties van promovendi mogen pas gebruikt worden ter training van GPT-NL nadat (1) het promotiereglement van een universiteit is aangepast, (2) de licentie die aan promovendi

¹ <https://gpt-nl.nl/gpt-nl/>

voorgelegd wordt is aangepast en (3) een workflow is ingericht om de opt-out van promovendi te faciliteren.

Artikelen, conference papers en hoofdstukken uit verzamelbundels die zijn opgenomen in repositoria onder artikel 25fa Auteurswet (Taverne)

Artikelen die via Taverne in repositoria zijn gepubliceerd mogen niet gebruikt worden ter training van GPT-NL, ook niet in de toekomst, aangezien de uitgevers auteursrechthebbenden. Taverne omvat uitsluitend een alleen-lezen recht voor derden.

Databankenrecht

De universiteiten kunnen geen gebruik maken van hun databankenrecht. De data bestaan immers uit teksten waar auteursrecht op rust en waarvan de universiteiten niet de rechthebbenden zijn, en waarvan de rechthebbenden geen toestemming hebben gegeven voor gebruik anders dan het opnemen in het repositorium met een alleen-lezen recht voor derden.

Hoe nu verder?

De universiteiten, en dus ook de UB's, zijn niet (langer) auteursrechthebbenden van de werken in de repositoria. Daarom kan geen van beiden toestemming verlenen (of aangeven dat zij 'geen bezwaar' heeft) om de desbetreffende werken te gebruiken voor het trainen van GPT-NL.

M.b.t. huidige werken: die kunnen niet gebruikt worden omdat ze onder de wettelijke uitzondering van art. 4 DSM (lijken te) vallen. Bij een wettelijke uitzondering (zoals deze) zijn Creative Commons licenties niet van toepassing. Er is bij deze uitzondering wel een recht op opt-out, maar dit recht van opt-out is niet aangeboden noch gefaciliteerd door de universiteiten.

M.b.t. toekomstige werken: die zouden gebruikt kunnen worden als de voorwaarden daarvoor worden gecreëerd, door a) licentieovereenkomsten met studenten en promovendi aan te passen en daarbij een opt-out en/of opt-in op te nemen en b) het promotiereglement te (laten) aanpassen.

INLEIDING EN ACHTERGROND

GPT-NL is een Large Language Model (LLM) dat ontwikkeld wordt door TNO en is een reactie op bestaande taalmodellen.² TNO heeft daartoe een subsidie ontvangen van het ministerie van Economische Zaken en Klimaat (EZK). TNO is penvoerder, SURF en NFI zijn partners. De opdracht is een taalmodel aan te bieden aan commerciële partijen (de zgn. BV Nederland) en gratis beschikbaar te stellen aan niet-commerciële partijen zoals onderzoekers bij universiteiten en andere onderzoeksinstituten. Omdat dit taalmodel ook een commerciële component heeft, valt de ontwikkeling van dit taalmodel onder artikel 4 DSM Richtlijn (TDM ten behoeve van alles wat niet uitsluitend wetenschappelijk onderzoek betreft en waarbij een opt-out mogelijk is) en niet onder artikel 3 DSM Richtlijn (TDM ten behoeve van wetenschappelijk onderzoek waarbij een opt-out niet mogelijk is).

SURF heeft contact gezocht met (de bibliotheken van) Nederlandse universiteiten met het verzoek materiaal uit de universitaire repositoria beschikbaar te stellen om dit LLM te trainen.

UKB heeft haar Werkgroep Auteursrecht gevraagd om te inventariseren of, en zo ja welk, materiaal uit de universitaire repositoria aangeboden kan worden aan SURF ter training van dit LLM. Dat heeft geresulteerd in een advies dat de Werkgroep Auteursrecht op 27 mei 2025 heeft aangeboden aan UKB (dit advies is als Bijlage 2 bijgevoegd bij dit document).

² <https://gpt-nl.nl/gpt-nl/>

Over dit advies van 27 mei 2025 staat in het verslag van de UKB plenair vergadering het volgende:
De vergadering complimenteert de werkgroep met het advies. Op basis hiervan bespreken de directeurs het gebruik van content uit de repositories en wetenschappelijke publicaties van de universitaire wetenschappers met een creative commons licentie voor het trainen in GPT-NL. De directeurs kunnen dit gebruik formeel niet goedkeuren, omdat ze niet de rechthebbenden van de betreffende content zijn, maar UKB staat wel positief tegenover het initiatief.

De UB-directeurs komen overeen:

- dat UKB positief staat ten opzichte van de ontwikkeling van een Nederlands taalmodel in GPT-NL;
- dat UKB waardeert dat SURF zich bewust is van de ethische en auteursrechtelijke aspecten in deze;
- dat UKB geen bezwaar heeft tegen het inzetten van wetenschappelijke artikelen met CC0 en CC-BY voor het trainen van GPT-NL.

De voorzitter van UKB heeft inmiddels (via de voorzitter van) de UKB Werkgroep Auteursrecht een vervolgoopdracht gegeven. Deze luidt:

Maak een aanvulling op het advies "UKB SCENARIO'S TEN AANZIEN VAN GPT-NL" waarin de voorwaarden waaronder het gebruik van overige wetenschappelijke publicaties en content in onze repositories voor het trainen van taalmodellen toegestaan is.

Overige wetenschappelijke publicaties en content betreft onderdelen:

- scripties (bachelor/master);
- dissertaties; en
- artikelen, conference papers en hoofdstukken uit verzamelbundels die gepubliceerd zijn niet in open acces maar onder artikel 25fa Auteurswet (Taverne amendement), uit het hiervoor genoemd advies. Bij het opstellen van de voorwaarden, neem ook het databankenrecht in overweging.

In het bovenstaande lezen de leden van de Werkgroep een tegenstrijdigheid: aan de ene kant merken de directeurs op dat ze gebruik niet formeel kunnen goedkeuren (en dit is juist), aan de andere kant hebben de directeurs geen bezwaar tegen het gebruik van content met een CC0 en CC BY licentie.

Overigens zij opgemerkt dat de Werkgroep zich in haar advies beperkt tot de auteursrechtelijke aspecten van het onderwerp. Zij heeft zich niet gebogen over bijvoorbeeld privacy en (het vermijden van) het verwerken van persoonsgegevens. Denk daarbij o.a. aan de namen van de auteurs van de diverse teksten.

Vervolgens heeft de Werkgroep het uitgangspunt op de website van GPT-NL³ over de bescherming van intellectueel eigendom meegenomen in haar advies. Dit uitgangspunt betreft het gegeven dat een data provider eigenaar moet zijn van het auteursrecht van de content of geldige licentierechten moet hebben verkregen van de eigenaar van de content. Alleen dan zal de content gebruikt worden ter training van GPT-NL.

Het advies is gebaseerd op de tot op heden geldende visie dat het trainen van LLM's moet worden aangemerkt als een vorm van TDM. Dit wordt ter discussie gesteld in diverse rechtszaken. Als de visie dat het trainen van LLM's niet moet worden aangemerkt als TDM de overhand krijgt, is dat van grote invloed op dit advies. In de praktijk zou dit betekenen dat auteursrechthebbenden van de werken vooraf om toestemming moet worden gevraagd om deze te mogen gebruiken voor AI-training.

³ <https://gpt-nl.nl/gpt-nl/> Zie ook punt 1 van dit Advies en Bijlage 1.

ADVIES

De Werkgroep komt tot het volgende advies.

1. Het auteursrecht

Auteursrecht is een grondrecht dat toekomt aan de maker. Dit recht is verankerd in internationaal recht, Europees recht en nationaal recht. Het bestaat uit een set van eigendomsrechten en morele rechten. De ontwikkelaars van GPT-NL erkennen dit en nemen dat als hun uitgangspunt. Zij geven aan hun taalmodel uitsluitend te trainen met content waarvan de data provider auteursrechthebbende is of waarvoor deze over een licentie beschikt van de auteursrechthebbenden.

Op de website van GPT-NL⁴ staat onder het kopje “Dataverzameling from scratch” bij het eerste criterium “Bescherming van intellectueel eigendom” het volgende:

“We gebruiken alleen data voor het trainen van GPT-NL als de data provider de juiste rechten heeft om ons een licentie te verlenen voor het gebruik van de data. Dit betekent dat de data provider de eigenaar moet zijn van eventuele auteursrechten of databaserechten in de dataset, of geldige licentierechten moet hebben gekregen van de eigenaar.”

Data provider zijn in dit geval de universiteiten.

De universiteiten zijn niet de eigenaar van de auteursrechten:

- op scripties (dat zijn de studenten);
- op dissertaties (dat zijn de promovendi of de uitgevers);
- op artikelen die via Taverne in de repositoria staan (dat zijn de uitgevers).

Universiteiten hebben ook geen geldige licentierechten gekregen van deze eigenaren. Universiteiten kunnen dus niet ‘geen bezwaar’ hebben dat materiaal wat in de repositoria staat, gebruikt wordt voor het trainen/ontwikkelen van het GPT-NL taalmodel.

Ter vergelijking: het beleid van de KB hierover is hetzelfde ten aanzien van content waarvan de KB niet de rechthebbende is.⁵ De ontwikkelaars van GPT-NL hebben daarom van de KB alleen content ontvangen die zich in het publiek domein bevindt.

Conclusie: Universiteiten kunnen de ontwikkelaars van GPT-NL geen toestemming verlenen (of aangeven dat ze ‘geen bezwaar hebben tegen’) om de bovengenoemde content te gebruiken voor het trainen/ontwikkelen van het GPT-NL taalmodel. Voor alle duidelijkheid: ook zonder de beginselen aangegeven op de website van GPT-NL komt de Werkgroep tot deze zelfde conclusie. Deze beginselen bevestigen alleen maar onze conclusie. Bovenstaande overwegingen staan geen enkel gebruik van content uit de universitaire repositoria toe, zoals hieronder gespecificeerd per contentgroep. Hieronder wordt ook per contentgroep advies gegeven om, indien gewenst, het gebruik voor trainingsdoeleinden in de toekomst mogelijk te maken.

2. Scripties (bachelor en master)

Het auteursrecht op een scriptie komt toe aan de auteur, de student. Het auteursrecht op scripties

⁴ <https://gpt-nl.nl/gpt-nl/>

⁵ <https://www.kb.nl/ai-statement>

komt in beginsel nooit toe aan de universiteit, ook niet aan universiteiten die werkgeversauteursrecht als beleid hebben. Studenten zijn immers geen werknemers. Universiteiten eisen dat studenten hun scriptie uploaden in het betreffende repositorium voor intern gebruik (publicatie op intranet) en/of extern gebruik (publicatie op internet). Er is geen wettelijke verplichting tot openbaarmaking van scripties. Sommige universiteiten hebben besloten dat scripties alleen op intranet inzichtelijk worden gemaakt, en niet daarbuiten. Er zijn universiteiten die geen enkele scriptie online hebben staan, er zijn universiteiten die alle scripties online hebben staan, er zijn universiteiten die wel masterscripties maar geen bachelorscripties online hebben staan, en er zijn universiteiten waar dit per faculteit verschilt.

Zoals al in het vorige advies uitgelegd, mag een rechthebbende geen bezwaar maken tegen TDM door een onderzoekinstelling met als doel wetenschappelijk onderzoek (art. 3 DSM). Wel is er voor de auteursrechthebbende een mogelijkheid tot opt-out voor TDM met een commercieel doel (art. 4 DSM), ook als dit uitgevoerd wordt door een onderzoekinstelling zoals TNO.

Mag nu gebruik gemaakt worden van de scripties die al op internet staan en waarvan studenten geen gebruik hebben gemaakt van hun opt-out?

Voor zover de Werkgroep weet, heeft nog geen student gebruik gemaakt van deze opt-out. De Werkgroep noemt twee redenen hiervoor. Ten eerste weten de meeste (zo niet alle) studenten helemaal niet dat deze mogelijkheid bestaat. Geen enkele universiteit heeft dit kenbaar gemaakt aan de studenten, noch in de licentie die studenten tekenen, noch op de universitaire websites. Ten tweede heeft geen enkele universiteit een workflow ingericht opdat, indien een student gebruik wil maken van het recht van opt-out, dit ook daadwerkelijk geregeld kan worden.

Aangezien geen enkele universiteit de mogelijkheid tot opt-out heeft gecommuniceerd, noch de mogelijkheid in het leven heeft geroepen om daadwerkelijk van deze opt-out gebruik te maken, acht de Werkgroep het ethisch en juridisch (wettelijke zorgplicht) uitgesloten dat met terugwerkende kracht UKB toestemming geeft om scripties in het repositorium te laten gebruiken voor het trainen van GPT-NL en voor het trainen van andere AI-systemen.

De Werkgroep acht het alleen mogelijk dat, onder voorwaarden, toekomstige scripties gebruikt worden voor het trainen van GPT-NL en voor het trainen van andere AI-systemen.

Toekomstscenario's voor het trainen van een AI-systeem op scripties

Scenario 1

Een universiteit kiest ervoor om hun scriptiedatabases offline te halen voor het publiek en deze slechts intern te blijven delen. Dan hoeft de universiteit verder niets te doen en kunnen ook andere modellen niet langer worden getraind op deze scripties. Nadeel is dat studenten van andere universiteiten niet langer deze scripties kunnen gebruiken ter inspiratie. Om inspiratie op te doen zijn de studenten dan beperkt tot de eigen universiteit.

Scenario 2

Een universiteit kiest ervoor om scripties online te publiceren en vermeldt dat de scripties zonder toestemming van de individuele student niet gebruikt mogen worden voor AI training. De universiteit neemt technische maatregelen om scrapen tegen te gaan en vermeldt op de website dat het partijen vrij staat om rechtstreeks contact op te nemen met rechthebbenden om deze toestemming te verkrijgen.

Scenario 3

Een universiteit kiest ervoor publicatie in het repository te verbinden met toestemming voor AI training. De student heeft dan de optie a) mijn scriptie is publiekelijk zichtbaar in het repository en mag gebruikt worden voor AI training of b) mijn scriptie is publiekelijk niet zichtbaar in het repository en kan dan niet gebruikt worden voor AI training.

Scenario 4

Een universiteit kiest ervoor alle scripties te publiceren in het repository, zichtbaar voor het grote publiek, en differentieert. Ze laat de studenten kiezen voor wel of geen beschikbaarstelling voor AI training en neemt technische maatregelen om dit te faciliteren. Dit zal wellicht in de praktijk moeilijk blijken te zijn.

Bij alle vier scenario's dient de licentie die de student wordt aangeboden te worden aangepast, en dient er een workflow te worden ingericht om de opt-out te faciliteren.

Samenvattend:

Huidige situatie

Scripties van studenten die nu al in repository staan mogen niet gebruikt worden ter training van GPT-NL, tenzij de universiteit (of de ontwikkelaars van GPT-NL zelf) individuele toestemming van de betreffende oud-studenten verkrijgt. Dit komt omdat de universiteiten deze oud-studenten tot nu toe nooit een opt-out hebben aangeboden en daarmee niet hebben voldaan aan hun zorgplicht.

Toekomstige situatie

Scripties van studenten mogen pas gebruikt worden ter training van GPT-NL nadat (1) de licentie die studenten voorgelegd wordt is aangepast en (2) een workflow is ingericht om de opt-out van studenten te faciliteren.

Als een universiteit besluit de scripties van studenten niet online te zetten, dan houdt dit in dat de universiteit ook *voor de toekomst* de mogelijkheid uitsluit om de scripties van de studenten te verzamelen en deze aan te bieden aan de ontwikkelaars van GPT-NL ter training van hun model. Want toestemming van de rechthebbenden (de studenten) ontbreekt dan.

3. Dissertaties

Dissertaties worden geschreven door promovendi. Er zijn promovendi in dienst van de universiteit (met arbeidsovereenkomst en salaris) en promovendi die niet in dienst zijn van de universiteit, zoals buitenpromovendi en beurspromovendi. Het auteursrecht op dissertaties van promovendi *niet* in dienst van de universiteit komt toe aan die promovendi. Het auteursrecht op dissertaties van promovendi *in* dienst van een universiteit komt ook (bijna) altijd toe aan de promovendi. Zelfs bij universiteiten die fervent voorstander zijn van werkgeversauteursrecht.⁶ In veel gevallen wordt de dissertatie daarnaast door een commerciële uitgever gepubliceerd en is het auteursrecht veelal volledig of exclusief overgedragen aan die uitgever.

Het regelen van het gebruik van dissertaties ter training van GPT-NL moet vooraf gebeuren, omdat later de promovendus/a de dissertatie als monografie of per hoofdstuk als artikel kan publiceren, en daarbij een contract met de uitgever zal worden getekend waarbij overdracht of exclusieve licentie

⁶ <https://uba.uva.nl/ondersteuning/auteursrecht/onderzoekers/onderzoekers.html#anker-eigen-werk>

aan de uitgever wordt gegeven. Ook aan promovendi moet wel de mogelijkheid van een opt-out geboden worden, zoals de wet voorschrijft. In de meeste gevallen komt het auteursrecht op de dissertatie namelijk toe aan de promovendus/a of aan de beursverstrekker.

Het advies is om het promotiereglement te wijzigen en daarin de mogelijkheid tot deze licentie alsmede de opt-out op te nemen. Tevens dient dan de niet-exclusieve licentie die de promovendus/a de universiteit verleent voor opname van de dissertatie in het repository aangepast te worden. In deze licentie wordt de promovendus/a de keuze gegeven om wel of niet in te stemmen met het gebruik van zijn dissertatie tot het trainen van LLM's en andere AI-systemen. Een workflow dient te worden ingericht om dissertaties met toestemming te scheiden van dissertaties zonder toestemming en om in de dissertaties zonder toestemming aan te geven dat deze niet gebruikt mogen worden voor TDM. Dit is noodzakelijk omdat de dissertaties immers voor eenieder toegankelijk zijn in de repositoria.

Aangezien geen enkele universiteit tot op heden de mogelijkheid tot opt-out heeft gecommuniceerd, noch een workflow in het leven heeft geroepen om daadwerkelijk van deze opt-out gebruik te maken, acht de Werkgroep Auteursrecht het ethisch en juridisch (wettelijke zorgplicht) uitgesloten dat met terugwerkende kracht UKB toestemming geeft om dissertaties in het repository te laten gebruiken voor het trainen van GPT-NL of voor het trainen van andere AI-systemen.

Bovendien zegt de wet dat alleen de rechthebbende gebruik mag maken van het recht van opt-out. Van de meeste dissertaties zijn de promovendi allang geen rechthebbenden meer, die rechten zijn inmiddels overgedragen aan uitgevers. Wij nemen aan dat uitgevers niet willen dat hun content gebruikt wordt voor TDM met een commercieel doel (inclusief en met name het trainen van LLM's en andere AI-systemen), althans niet zonder hun toestemming. Voor dat doel willen ze mogelijk hun content in licentie geven, maar daar zullen hoogstwaarschijnlijk kosten aan verbonden worden. Indien universiteiten deze rechten wel weggeven, zoals bijvoorbeeld aan de partijen die GPT-NL ontwikkelen, zou dat niet alleen de verhoudingen tussen universiteiten en uitgevers op scherp kunnen zetten, maar universiteiten lopen ook het risico aansprakelijk gesteld te worden voor auteursrechtinbreuk.

Samenvattend:

Huidige situatie

Dissertaties van promovendi die al in repositoria staan mogen niet gebruikt worden ter training van GPT-NL, tenzij de universiteit individueel toestemming van de betreffende oud-promovendus/a verkrijgt. Dit geldt in elk geval voor buitenpromovendi en voor promovendi in dienst van een universiteit die het auteursrecht op dissertaties doet toekomen aan de promovendi zelf. Zelfs indien de universiteit auteursrechthebbende is van de dissertatie van de promovendi in dienst via het werkgeversauteursrecht, wordt het vragen van toestemming geadviseerd, aangezien promovendi nooit eerder zijn geconfronteerd met de mogelijkheid dat hun dissertaties voor het trainen van AI-systemen zouden kunnen worden gebruikt.

Toekomstige situatie

Toekomstige dissertaties van promovendi mogen pas gebruikt worden ter training van GPT-NL nadat (1) het promotiereglement van een universiteit is aangepast, (2) de licentie die aan promovendi voorgelegd wordt is aangepast en (3) een workflow is ingericht om de opt-out van promovendi te faciliteren.

4. Artikelen, conference papers en hoofdstukken uit verzamelbundels die gepubliceerd zijn niet in open access maar zijn opgenomen in de universitaire repositoria onder artikel 25fa Auteurswet (Taverne)

De in de titel genoemde korte werken van wetenschap worden hierna gemakshalve samengevat met het woord 'artikel'.

Wanneer artikelen van medewerkers en andere met de universiteit geaffilieerden via Taverne geplaatst worden in het repositorium, is er al een embargo van zes maanden na publicatie verlopen. Het contract van de auteur met de uitgever is dan al geruime tijd getekend, het auteursrecht op het artikel is overgedragen of in exclusieve licentie gegeven aan de uitgever. De rechten van derden, inclusief de universiteiten zelf, op het artikel beperken zich tot een 'alleen-lezen' recht. Wij nemen aan dat uitgevers niet willen dat hun content is gebruikt wordt voor TDM met een commercieel doel (inclusief en met name het trainen van LLM's en andere AI-systemen), althans niet zonder hun toestemming. Voor dat doel willen ze mogelijk hun content in licentie geven, maar daar zullen hoogstwaarschijnlijk kosten aan verbonden worden.

Samenvattend: artikelen die via Taverne in repositoria staan mogen niet gebruikt worden ter training van GPT-NL, ook niet in de toekomst, aangezien de rechten liggen bij de uitgevers.

5. Het databankenrecht

Het databankenrecht kan worden ingeroepen als er een substantiële investering is gedaan in het verkrijgen/verzamen/orderen van de data. Of dit voor universitaire repositoria geldt is onduidelijk, nader onderzoek door experts op het gebied van het databankenrecht wordt daarom aangeraden.

De universiteiten kunnen geen gebruik maken van hun (eventuele) databankenrecht. De data bestaan immers uit teksten waar auteursrecht op rust en waarvan de universiteiten niet de rechthebbenden zijn, en waarvan de rechthebbenden geen toestemming hebben gegeven voor gebruik anders dan het opnemen in het repositorium met een alleen-lezen recht voor derden. De universiteiten kunnen misschien wel gebruik maken van hun eventuele databankenrecht voor wat betreft de "kale" metadata van de publicaties/teksten die zich in de repositoria bevinden. Met kale metadata bedoelen we dan bibliografische gegevens zoals titels, jaar van uitgave, ISBN, etc. van publicaties, maar zonder persoonsgegevens. Het is wel de vraag of een corpus van uitsluitend kale metadata interessant is voor de ontwikkelaars van GPT-NL.

6. AI/TDM-voorbehoud content in repositoria

Vanuit juridisch en ethisch perspectief hebben de universiteit(sbibliothek)en geen zeggenschap over de content die in hun repositoria is opgenomen en waarbij de rechthebbenden niet de keuze hebben gehad om af te zien van een TDM-voorbehoud of dit voorbehoud juist in te roepen. Ook nadat de universiteiten deze keuze wel in hun workflows hebben opgenomen, zal bij een aantal werken in de repositoria aangegeven moeten worden dat deze niet gebruikt mogen worden voor TDM met een commercieel doel, inclusief het trainen van LLM's en andere AI-systemen. Dit AI/TDM-voorbehoud moet duidelijk aangegeven worden in de gebruiksvoorwaarden en zal ook machine

leesbaar aangegeven moeten worden op elke webpagina en in elk document waar dit voor geldt. Elke opt-out zal technisch verwerkt moeten worden in het repositorium. Dit zal hoogstwaarschijnlijk een hele puzzel/klus blijken op te leveren. Afstemming met de verantwoordelijke collega's en uitvoerende collega's zal dienen plaats te vinden om dit in goede banen te leiden.

Aan externen die wel graag TDM met een commercieel doel willen uitoefenen op content in de repositoria, zal op de website van de repositoria uitgelegd dienen te worden hoe zij toestemming kunnen verkrijgen en bij wie zij zich voor deze toestemming dienen te melden: uitgevers (in geval van gepubliceerde artikelen en gepubliceerde proefschriften), individuele studenten (in geval van scripties), etc.

7. Recente jurisprudentie

Er worden zowel in de VS als in Europa rechtszaken gevoerd tegen de bedrijven die hun AI-systemen hebben getraind op auteursrechtelijk beschermd materiaal. Dit kan zowel tekst als beeld, audio (muziek) en video (film) betreffen. In de VS zijn een aantal procedures inmiddels gestaakt omdat er overeenkomsten en licenties gesloten zijn. Dat wil zeggen dat de bedrijven een licentie gesloten hebben met een of meerdere rechthebbenden en daar een (fors?) bedrag voor hebben neergelegd. De bedrijven voerden de procedures in eerste instantie met een beroep op het zgn. 'fair use' dat geldt in de VS.

In de EU, waar het concept van 'fair use' niet als wettelijk recht bestaat, doen de bedrijven een beroep op art. 4 DSM⁷. Dit artikel bepaalt dat TDM met een commercieel doel mag, tenzij de rechthebbende gebruik maakt van zijn opt-out. Ook in Europa zijn er rechtszaken aanhangig gemaakt omdat rechthebbenden menen dat hun rechten geschonden zijn omdat bedrijven hun auteursrechtelijk beschermde werken gebruikt hebben om hun AI-systemen te trainen. De uitkomst van deze procedures variëren. En tot nu toe zijn het uitspraken van lagere rechters, er zijn nog geen uitspraken van de Nederlandse Hoge Raad of van het Europees Hof. De rechtsvraag die beantwoord moet worden is: is het trainen van een LLM te beschouwen als een vorm van TDM? Zo ja, dan is art. 4 DSM van toepassing. Zo nee, dan is art. 4 DSM niet van toepassing en is het ontbreken van een opt-out niet voldoende: er moet dan vooraf toestemming gevraagd worden door het bedrijf/de entiteit die het AI-systeem wil trainen met materiaal van rechthebbenden.

Een voorbeeld is de uitspraak van de Rechtbank in München, Duitsland inzake GEMA vs OpenAI.⁸ GEMA vertegenwoordigt rechthebbenden, OpenAI traint AI-systemen ontwikkelt taalmodellen en chatbots die op die taalmodellen zijn gebaseerd. Het taalmodel reproduceerde de teksten van bestaande liedjes. De rechtbank oordeelde dat het taalmodel deze teksten heeft memoriseerd, en dat deze memorisatie niet valt onder TDM:

"Both the memorization of the lyrics in the language models and the reproduction of the lyrics in the chatbot's outputs constitute infringements of copyright exploitation rights. These infringements are

⁷ <https://eur-lex.europa.eu/legal-content/NL/TXT/HTML/?uri=CELEX:32019L0790>

⁸ Zie https://openjur.de/by/ig_munchen_i.html. Voor blogs hierover, zie bijvoorbeeld <https://mediareport.nl/persrecht/11112025/ai-alert-ruling-of-munich-district-court-in-european-landmark-case-gema-openai/> en <https://www.osborneclarke.com/insights/gema-vs-openai-ai-memorisation-reproduction-relevant-copyright-law-and-tdm-exception-does> en https://www.ai-forum.nl/art/90-6575_GEMA-v-OpenAI-Duitse-rechtbank-zet-punt-achter-ongelicenseerde-generatieve-AI-training

not covered by limitations, in particular the limitation for text and data mining.”⁹

Dit betekent dat gebruik van content zonder opt-out niet mag, er moet actief en vooraf toestemming worden gevraagd.

Vooraf collectieve beheersorganisaties (CBO's) en uitgeverscollectieven zijn van mening dat het trainen van AI-systemen zoals LLM's niet vallen onder de uitzondering van artikel 4 DSM. Zij dringen er bij de EU op aan om dit artikel 4 DSM te herzien (maak van een opt-out een opt-in).¹⁰ Zowel de Spaanse¹¹ als de Portugese¹² overheid zijn van oordeel dat artikel 4 DSM niet van toepassing is op het trainen van AI-systemen.

Als deze trend zich doorzet (en dat is niet onwaarschijnlijk) dan heeft dat consequenties.

Ten eerste kan een AI-systeem dan niet getraind worden op content zonder de expliciete toestemming van de rechthebbenden van deze content (een opt-in).

Ten tweede kunnen universiteiten aansprakelijk gesteld worden door rechthebbenden (studenten, promovendi, uitgevers) voor het toestemming geven tot gebruik aan de ontwikkelaars van GPT-NL zonder toestemming van de daadwerkelijke rechthebbenden, deze rechthebbenden kunnen dan een schadevergoeding eisen.

Ten derde kunnen de rechthebbenden middels een gerechtelijke procedure eisen dat het taalmodel van de markt wordt gehaald indien GPT-NL getraind is op content zonder toestemming van deze rechthebbenden.

Creative Commons licenties

UKB heeft besloten aan de ontwikkelaars van GPT-NL aan te geven dat UKB 'geen bezwaar heeft tegen' het inzetten van het inzetten van wetenschappelijke artikelen met CC0 én CC BY voor het trainen van GPT-NL. Er is binnen de Werkgroep geen bezwaar tegen het gebruik van content met een CC0. Dit is ook overeenkomstig het beleid van de KB, die aan de ontwikkelaars van GPT-NL publiek domein materiaal heeft aangeboden.

Toch zijn er binnen de Werkgroep toch wel bezwaren tegen het gebruik van content met een CC BY (en dus ook content met andere CC licenties).

Voorwaarden voor hergebruik van materiaal met een CC BY (en ook alle overige CC licenties, behalve CC0) licentie is tweeledig:

- er moet verwezen worden naar de oorspronkelijke maker/auteur van het materiaal;
- er moet worden aangegeven welke wijzigingen de hergebruiker heeft aangebracht.

Dit wordt op de website van Creative Commons als volgt beschreven:

“Attribution – you must give appropriate credit, provide a link to the licence, and indicate if changes were made.”¹³ En verder: “You must indicate if you modified the material and retain an indication of previous modifications.”¹⁴ Indien een AI-systeem is getraind op materiaal met een CC BY licentie kan bij de output (het antwoord op een zgn. ‘prompt’) aangegeven worden wat de bron is, maar dat gebeurt lang niet altijd. Het gebeurt feitelijk alleen indien gebruik gemaakt wordt van de zgn. memorization, want dan is duidelijk waar de output vandaan is gekomen. Heel vaak zal de output op

⁹ Uit het persbericht van de rechtbank, zie: <https://www.justiz.bayern.de/gerichte-und-behoerden/landgericht/muenchen-1/presse/2025/11.php>. De uitspraak is nog niet definitief, er kan op moment van schrijven nog hoger beroep worden ingesteld.

¹⁰ [https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/776366/IUST_BRI\(2025\)776366_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/776366/IUST_BRI(2025)776366_EN.pdf)

¹¹ https://www.boe.es/diario_boe/txt.php?id=BOE-A-2021-17910

¹² <https://diariodarepublica.pt/dr/detalhe/decreto-lei/47-2023-214524782>

¹³ <https://creativecommons.org/licenses/by/4.0/>

¹⁴ <https://creativecommons.org/licenses/by/4.0/#ref-indicate-changes>

basis van (veel) verschillende input content samengesteld worden, en dan is niet duidelijk wat de bronnen en de makers zijn. De output moet dan ook aangeven welke wijzigingen zijn aangebracht. Voor zover de Werkgroep weet, gebeurt dit niet.

De Werkgroep adviseert UKB plenair daarom dringend om terug te komen op haar beslissing tot het afgeven van een 'verklaring van geen bezwaar' om content met CC BY licentie in de repositoria te gebruiken voor het trainen van GPT-NL.

De Werkgroep adviseert daarom dat universiteit(sbibliotheek)en geen toestemming geven (of een verklaring van geen bezwaar geven) aan de ontwikkelaars van GPT-NL om dit model te trainen met content in de universitaire repositoria, met uitzondering van content in het publiek domein. De KB heeft al heel wijs besloten om alleen toestemming te geven voor gebruik van content in het publiek domein. De Werkgroep raadt UKB plenair aan een en ander voor te leggen aan de afdeling Juridische Zaken van hun respectieve universiteiten om een risico assessment te maken. De ontwikkelaars van GPT-NL doen er goed aan om hun taalmodel niet te trainen met content zonder expliciete toestemming van de rechthebbenden.

8. Hoe nu verder?

UKB noch de universiteiten kunnen de ontwikkelaars van GPT-NL toestemming verlenen, of aangeven dat zij 'geen bezwaar' hebben, om de huidige content in de repositoria te gebruiken voor het trainen van hun taalmodel.

UKB kan de universiteit(sbestuur)en adviseren om beleid op te stellen ten aanzien van commercieel gebruik van content in de repositoria, zoals het trainen van AI-systemen waaronder LLM's. Mocht een universiteit deze content beschikbaar willen stellen voor commercieel gebruik, dan kunnen auteursrechtsspecialisten in samenwerking met de juridische afdelingen binnen de instellingen adviseren over de inrichting van de juridische kaders en workflows die dit in de toekomst mogelijk kunnen maken. De workflows dienen zo ingericht te worden dat de opt-out of (indien wetgeving wijzigt) opt-in van studenten, promovendi en andere academische auteurs gefaciliteerd wordt. Dit beleid zou concreet moeten gaan over:

- a) licentieovereenkomsten met studenten en promovendi aanpassen en daarbij een opt-out en/of een opt-in opnemen, en
- b) het promotiereglement (te laten) aanpassen.

9. Tot slot

De leden van de Werkgroep Auteursrecht hebben uitgebreid over dit onderwerp vergaderd. Een van de leden vroeg zich af of de discussie niet achterhaald was omdat buitenlandse grote bedrijven allang het internet hebben gescraped ter training van hun taalmodellen en andere AI-systemen.

De leden hebben geconcludeerd dat dit niet het geval is: het is waar dat een aantal grote bedrijven dit al hebben gedaan, maar tegen dit gebruik/misbruik zijn inmiddels een groot aantal rechtszaken gestart door rechthebbenden die dit gebruik zonder hun toestemming aanvechten. De redenering 'Anderen hebben het gedaan dan moeten/mogen wij het ook maar doen' is een drogreden. Het is niet zo dat de leden er per sé tegen zijn dat universiteiten meewerken aan de ontwikkeling van GPT-NL. Echter, het auteursrecht dient te worden gerespecteerd en universiteiten zijn geen rechthebbenden van de content in hun repositoria. Universiteiten kunnen daarom geen

toestemming geven voor gebruik van deze content door de ontwikkelaars van GPT-NL. Laatstgenoemden hebben zich gebonden aan hun eigen 'code of conduct', namelijk om uitsluitend content te gebruiken met toestemming van de rechthebbenden.

Ook werd tijdens de vergadering de vraag gesteld of de ontwikkelaars van GPT-NL dan hun beginselen en uitgangspunt op de website inzake de bescherming van intellectueel eigendom¹⁵ niet moeten afzwakken of zelfs van de website zouden moeten verwijderen. En vervolgens net als de bedrijven achter de andere modellen de universitaire repositoria gaan scrapen en de content gebruiken voor het trainen van hun taalmodel.

De werkgroep concludeerde echter dat, als de ontwikkelaars van GPT-NL dit zouden doen, dit onethisch, onoorbaar, oneerlijk en zelfs ronduit schijnheilig zou zijn. De ontwikkelaars van GPT-NL dienen zich aan hun eigen beginselen van betrouwbaarheid en transparantie te houden.

¹⁵ <https://gpt-nl.nl/gpt-nl/>

The screenshot shows a web browser window with the address bar displaying 'gpt-nl.nl/gpt-nl/'. The website header includes the 'GPT-NL' logo, navigation links for 'Home', 'Contact', and a search icon, and a secondary menu with 'Over GPT-NL', 'Commitments', 'Samenwerken', 'Planning', 'Veelgestelde vragen', and 'Nieuws'. The main content area is titled 'Dataverzameling from scratch' and contains a paragraph about using state-of-the-art model architecture while training from scratch to avoid biases. It lists five criteria for data collection, with the first two being 'Bescherming van intellectueel eigendom' and 'Bescherming van persoonsgegevens', each followed by a brief explanation. The third criterion, 'Uitsluiten van informatie onder vertrouwelijkheidseisen', is partially visible.

gpt-nl.nl/gpt-nl/

Site Edit All Bookmarks

GPT-NL Home Contact

Over GPT-NL Commitments Samenwerken Planning Veelgestelde vragen Nieuws

Dataverzameling *from scratch*

We zullen state-of-the-art modelarchitectuur hergebruiken, maar de training zelf doen we helemaal *from scratch* om te voorkomen dat er onbekende factoren van eerdere modellen worden geërfd, zoals copyrightschendingen of het onrechtmatig gebruik maken van persoonsgegevens. Hierbij houden we ons aan de volgende 5 criteria voor een eerlijke dataset:

- **Bescherming van intellectueel eigendom**
We gebruiken alleen data voor het trainen van GPT-NL als de data provider de juiste rechten heeft om ons een licentie te verlenen voor het gebruik van de data. Dit betekent dat de data provider de eigenaar moet zijn van eventuele auteursrechten of databaserechten in de dataset, of geldige licentierechten moet hebben gekregen van de eigenaar.
- **Bescherming van persoonsgegevens**
We trainen GPT-NL niet op persoonsgegevens, behalve voor publieke personen waar we wel (een deel van de) persoonsgegevens van meenemen tijdens de training. Dit is nodig om het model basiskennis mee te geven en in lijn met de AVG.
- **Uitsluiten van informatie onder vertrouwelijkheidseisen**

<https://gpt-nl.nl/gpt-nl/> Geraadpleegd op 4 november 2025

UKB SCENARIO'S TEN AANZIEN VAN GPT-NL

Publieksversie

Van: UKB Werkgroep Auteursrecht (UKBWGAR)

Aan: UKB Plenair

Datum: 27 mei 2025

Inleiding en achtergrond

GPT-NL is een Large Language Model (LLM) dat ontwikkeld wordt door TNO. TNO heeft daartoe een subsidie ontvangen van het ministerie van Economische Zaken en Klimaat (EZK). TNO is penvoerder, SURF en NFI zijn partners. De opdracht is een taalmodel aan te bieden aan commerciële partijen (de zgn. BV Nederland) en gratis beschikbaar te stellen aan niet-commerciële partijen zoals onderzoekers bij universiteiten en andere onderzoeksinstellingen. De bedoeling is een LLM aan te bieden als tegenwicht voor ChatGPT (een bedrijf uit de VS) en zo Nederland (en Europa) onafhankelijker te laten worden van partijen buiten Europa.

SURF heeft contact gezocht met (de bibliotheken van) Nederlandse universiteiten met het verzoek materiaal uit de universitaire repositoria beschikbaar te stellen om dit LLM te trainen. Het verzoek is dan om alle documenten die in de repositoria beschikbaar zijn gesteld met een CC0 licentie en een CC BY licentie daarvoor te mogen gebruiken.

Het proces

Op verzoek van (de voorzitter van) de UKB heeft de UKB werkgroep Auteursrecht zich gebogen over dit verzoek en presenteert het onderhavige document. Er wordt uitgelegd wat een LLM is, wat de wet bepaalt over TDM (het trainen van een LLM is een vorm van TDM) en hoe het in de praktijk werkt met de diverse Creative Commons (CC) licenties. Vervolgens wordt een inventarisatie gemaakt van de diverse documenten die zich bevinden in de universitaire repositoria en aan wie het auteursrecht op deze documenten toekomt. Daarna worden scenario's geschetst die aan UKB ter overweging worden aangeboden. Tot slot geeft de UKB werkgroep Auteursrecht aan op welke overwegingen en beginselen UKB en anderen een beslissing kunnen nemen. Na bespreking in de vergadering van UKBWGAR van 26 mei 2025 is dit document vervolgens gezonden aan (de voorzitter van) de UKB. De UKB liaison kan dit document in UKB plenair nader toelichten.

Wat is een Large Language Model (LLM)?

Een LLM is een tool waarbij een persoon een vraag stelt of een opdracht geeft aan de tool. Dat wordt de 'prompt' genoemd. De tool geeft antwoord en/of voert de opdracht uit. Het resultaat is een tekst. Hoe genereert de tool de tekst? Dit is mogelijk omdat de tool is gevoed/getraind op een grote hoeveelheid data door middel van Tekst en Datamining (TDM). Deze data bestaan uit een grote hoeveelheid (soms miljoenen) teksten. Deze teksten zijn vaak auteursrechtelijk beschermd. Wanneer is TDM nu wettelijk toegestaan?

Hoe werkt het trainen van een LLM in de praktijk?

Commerciële partijen hebben voor het trainen van hun LLM's inmiddels al veelvuldig gebruik gemaakt van materiaal dat via het internet beschikbaar is gemaakt. Amerikaanse bedrijven hebben dit gebruik naar eigen zeggen geëgitimeerd door het Amerikaanse zgn. 'fair use'. Andere partijen zijn het daar niet mee eens en hebben rechtszaken geïnitieerd. De uitkomsten van deze rechtszaken kunnen vergaande gevolgen hebben voor het trainen van LLM's en het huidige business model, maar met die uitkomsten kunnen we op dit moment nog geen rekening houden. In de VS zijn ook elkaar tegensprekende uitspraken gedaan.¹⁶

We kijken nu hoe dit werkt in Europa onder EU wetgeving.

Wat bepaalt de wet?

Nederland is als lid van de EU onderworpen aan EU wetgeving. We kijken naar materiaal dat algemeen toegankelijk is via het internet en naar materiaal dat bij universiteiten toegankelijk is middels een inlog via een licentie van een derde partij (meestal een uitgever).

Het gebruik van TDM is geregeld in de artikelen 3 en 4 van de DSM Richtlijn, voluit EU Richtlijn 2019/790 inzake auteursrechten en naburige rechten in de digitale eengemaakte markt.¹⁷ Deze artikelen 3 en 4 zijn geïmplementeerd in de Nederlandse Auteurswet in de artikelen 15n en 15o Aw.¹⁸ De artikelen 3 en 4 DSM alsmede de artikelen 15n en 15o Aw zijn te vinden in Bijlage 1. Overigens, de richtlijn en dus ook deze artikelen 3 en 4 zijn tot stand gekomen voordat de 'hype' van LLM's was begonnen, dus in de tijd van het meer 'traditionele' TDM. Toch, het uitvoeren van TDM om een LLM te trainen valt onder de omvang van zowel de artikelen 3 en 4 van de DSM Richtlijn als de artikelen 15n en 15o Aw.

Artikel 3 DSM (15n Aw) bepaalt dat TDM ten behoeve van wetenschappelijk onderzoek altijd mag op teksten waar de onderzoeker/de instelling rechtmatig toegang toe heeft. Rechtmatige toegang is dan alles wat vrijelijk op internet staat en alles waar de onderzoeker zelf of via de instelling toegang toe heeft via licenties en abonnementen.¹⁹ Deze TDM t.b.v. onderzoek mag niet contractueel worden uitgesloten. Ook een niet-EU partij mag dit recht niet contractueel uitsluiten, want deze partij biedt zijn waren aan binnen de EU en is daarmee onderworpen aan EU wetgeving. Het LLM van GPT-NL valt niet onder dit artikel 3 DSM aangezien TNO dit model niet traint ten behoeve van wetenschappelijk onderzoek en TNO dit model (ook) aan commerciële partijen zal aanbieden.

Artikel 4 DSM (15o Aw) gaat over TDM voor andere doeleinden dan niet-commercieel wetenschappelijk onderzoek en is daarmee van toepassing op het onderhavige GPT-NL. Artikel 4 DSM bepaalt dat TDM met een commercieel oogmerk is toegestaan tenzij de rechthebbende op passende wijze een 'uitdrukkelijk voorbehoud' maakt. Daarna is voor deze partij TDM alleen nog mogelijk indien die partij toestemming heeft verkregen van de (individuele) auteursrechthebbende. Artikel 15o Aw spreekt alleen van een voorbehoud en niet van een uitdrukkelijk voorbehoud. Wat is nu een voorbehoud maken 'op passende wijze'? Artikel 15o lid 1 spreekt van 'zoals door middel van machinaal leesbare middelen bij een online ter beschikking gesteld werk'. Wat wordt nu verstaan onder een 'machinaal leesbaar middel'?

¹⁶ Zie bijvoorbeeld <https://kvdl.com/artikelen/ai-alert-training-van-ai-systemen-valt-onder-amerikaanse-fair-use-exceptie> en <https://www.mr-online.nl/trainen-van-ai-is-geen-fair-use/>

¹⁷ <https://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=CELEX:32019L0790>

¹⁸ <https://wetten.overheid.nl/BWBR0001886/2025-02-04>

¹⁹ Zie overwegingen 10 en 14 DSM inzake rechtmatige toegang.

Voorbehoud op passende wijze/machinaal leesbaar middel: Volgens overweging 18 van de DSM Richtlijn wordt het voorbehouden als passend beschouwd ‘... indien machinaal leesbare middelen worden gebruikt, waaronder metagegevens en de voorwaarden van een website of een dienst’. Het gebruik van contractuele overeenkomsten of een eenzijdige verklaring worden evenveel als passend beschouwd. Verder gaan we ervan uit dat de teksten (op websites) machinaal leesbaar zijn en dat een mededeling in de trant van: “De op deze website aangeboden teksten/artikelen/scripties mogen niet gebruikt worden voor het trainen van AI-modellen of andere vormen van TDM, tenzij dit gebeurt voor niet-commercieel wetenschappelijk onderzoek.” voldoende is. We gaan ervan uit dat het gebruik van een dergelijke tekst op een website een passend voorbehoud is in de zin van art. 4 DSM en art. 15o Aw.

Dit heeft gevolgen voor het gebruik onder art. 4 DSM van werken toegankelijk via abonnementen en licenties: indien op de website van de uitgever of in de licentieovereenkomst met de uitgever staat dat de data niet gebruikt mogen worden voor TDM (en het trainen van een AI model is een vorm van TDM) dan is het voorbehoud op passende wijze gemaakt en mag de commerciële partij die toegang heeft tot deze data, deze data niet voor het trainen van een AI model gebruiken.

Samenvattend: een LLM mag getraind worden met alle werken waartoe de LLM-ontwikkelaar rechtmatig toegang heeft tenzij er sprake is van een voorbehoud op passende wijze (het zgn. opt-out). Als in een licentie of contract geen sprake is van een voorbehoud, dan is TDM toegestaan. Als op een website (zoals een repository) geen voorbehoud wordt gemaakt voor alle daar te vinden werken, of in de werken zelf geen voorbehoud wordt aangegeven, dan is TDM toegestaan. En nogmaals, het trainen van een LLM is een vorm van TDM.

Inventarisatie documenten in universitaire repositoria

Het verzoek van SURF/TNO aan de universiteiten is om werken (teksten) uit de universitaire repositoria beschikbaar te stellen om het taalmodel van GPT-NL te trainen.

In de repositoria van universiteiten bevinden zich (globaal) de volgende “full tekst” documenten:

1. scripties van studenten (zowel bachelor als master scripties);
2. dissertaties van promovendi;
3. artikelen die gepubliceerd zijn in open access met een CC licentie;
4. artikelen, conference papers en hoofdstukken uit verzamelbundels die gepubliceerd zijn niet in open access, maar die onder art. 25fa Aw (het zgn. Taverne Amendement) beschikbaar gesteld zijn en aan de eindgebruiker een alleen-lezen recht toekennen.

We behandelen deze categorieën documenten en geven per categorie aan welke opties er zijn.

1. Scripties van studenten

Het auteursrecht op scripties komt toe aan de student.

De meeste universiteiten eisen of vragen van studenten dat zij hun scripties uploaden en beschikbaar stellen in het repository. Dit geldt niet voor alle universiteiten en ook binnen universiteiten worden niet alle scripties beschikbaar gesteld. De universiteit verkrijgt daartoe een niet-exclusieve licentie van de student. Deze licentie houdt niet meer in dan het recht van de universiteit om de scriptie te uploaden in het repository en het recht van de universiteit om aan de eindgebruiker een alleen-lezen recht te geven. Alle andere (auteurs)rechten blijven in handen van

de student.²⁰

Dat dit een alleen-lezen recht is, is vanaf het begin dat scripties online verschenen de bedoeling en het doel geweest van universiteiten. Het was nooit de bedoeling dat commerciële partijen de scripties zouden mogen gaan gebruiken voor het trainen van commerciële LLM's. Ook ander commercieel gebruik door eindgebruikers werd niet toegestaan in de licentie van de student aan de universiteit. Ten tijde van het beschikbaar stellen van het overgrote deel van de scripties bestonden art. 4 DSM en art. 15o Aw nog niet, evenmin als TDM en GenAI. Nu dus wel.

Op basis van de wet is commercieel TDM toegestaan, tenzij er een uitdrukkelijk voorbehoud wordt gemaakt. Dit voorbehoud is inmiddels gemaakt door een aantal universiteiten voor hun documenten die op PURE staan.²¹ Dit kunnen scripties zijn, maar ook dissertaties en korte werken die beschikbaar zijn gesteld via Taverne. De tekst is dan:

All content on this site: Copyright © 2025 the University of ..., its licensors, and contributors. All rights are reserved, including those for text and data mining, AI training, and similar technologies. For all open access content, the relevant licensing terms apply.

GPT-NL zegt geen materiaal/teksten te willen gebruiken voor het trainen van haar taalmodel zonder toestemming van de rechthebbenden. Echter, universiteiten zijn veelal niet de rechthebbenden van de teksten/artikelen die zich in haar repositoria bevinden. Zo ook in het geval van scripties. Toestemming van een universiteit aan SURF om deze scripties beschikbaar te stellen voor training van het LLM kan daarom niet gegeven worden, aangezien het auteursrecht niet toekomt aan de universiteit maar aan de student. En omdat (veel) universiteiten dit voorbehoud hebben gemaakt in de gebruiksvoorwaarden.

2. Dissertaties van promovendi

Het auteursrecht op dissertaties komt toe aan de promovendus/a. Universiteiten eisen van promovendi dat zij hun dissertatie uploaden en beschikbaar stellen in het repository, eventueel met een embargo. De universiteit verkrijgt daartoe een niet-exclusieve licentie van de promovendus/a. Deze licentie houdt niet meer in dan het recht van de universiteit om de dissertatie te uploaden in het repository en het recht van de universiteit om aan de eindgebruiker een alleen-lezen recht te geven. Alle andere (auteurs)rechten blijven in handen van de promovendus/a. In veel gevallen wordt de dissertatie later gepubliceerd. In dat geval wordt het auteursrecht op de dissertatie (bijna altijd) overgedragen aan de uitgever. Toestemming van een universiteit aan SURF om deze dissertaties beschikbaar te stellen voor training van het LLM kan daarom niet gegeven worden, aangezien het auteursrecht niet toekomt aan de universiteit maar aan de promovendus/a of aan de uitgever.

3. Artikelen met een Creative Commons licentie

Eerst even een opsomming van de CC licenties en waar de licenties (geen) toestemming voor geven, let op, er zijn combinaties mogelijk.

CC0 publiek domein, alles toegestaan, geen verplichting tot naamsvermelding

²⁰ Let op: op dit moment zijn er universiteiten die 'meer' eisen van de student. Er is een universiteit die op scripties een CC BY-NC licentie plaatst. Dit is dus nog een discussiepuntje. Een andere universiteit is ertoe overgegaan om scripties niet meer zichtbaar te maken voor het algemene publiek.

²¹ Zie bijvoorbeeld hier: <https://pure.eur.nl/> en hier: <https://research.rug.nl/>. Bedoelde tekst staat onderaan de pagina. De tekst is kennelijk toegevoegd door Elsevier.

CC BY-NC	hergebruik, afgeleide werken, geen commercieel gebruik , wel naamsvermelding
CC BY-ND	hergebruik, wel commercieel gebruik, geen afgeleide werken , wel naamsvermelding
CC BY-SA	hergebruik, wel commercieel gebruik, wel afgeleide werken, wel naamsvermelding, verplichting tot het delen onder dezelfde licentie van het eindproduct
CC BY-SA-NC	hergebruik, geen commercieel gebruik , wel afgeleide werken, wel naamsvermelding, verplichting tot het delen onder dezelfde licentie van het eindproduct
CC BY	hergebruik, wel commercieel gebruik, wel afgeleide werken, wel naamsvermelding

Wij van de taakgroep dachten eerst:

- CC BY-NC + CC BY-SA-NC geeft geen toestemming voor het trainen van een commercieel LLM, want commercieel gebruik is niet toegestaan;
- CC BY-ND geeft geen toestemming voor het trainen van een commercieel LLM, want de output van het LLM levert een afgeleid werk op;
- CC BY-SA geeft geen toestemming voor het trainen van een commercieel LLM, want GPT-NL wordt later niet beschikbaar gesteld onder dezelfde licentie;
- CC BY geeft geen toestemming voor het trainen van een commercieel LLM, want GPT-NL geeft bij de output gegeneerd door een prompt niet aan waar de output van de prompt op gebaseerd is. Dit bleek niet waar te zijn. Het is al technisch mogelijk om bij de output van een prompt aan te geven waar de output vandaan komt. Zie bijvoorbeeld GENIA-L, app.rechtsorde.nl. Deze LLM is getraind op literatuur die is opgenomen in rechtsorde.nl, op rechtspraak.nl, op BOOM portaal en op overheid.nl. Het zou mogelijk zijn dat GPT-NL dit in de toekomst ook kan. Voor de input kan GPT-NL op haar website een lijst opnemen van alles waar GPT-NL op getraind is om te voldoen aan de 'BY' in alle licenties. Voor de output zou aan naamsvermelding kunnen worden voldaan indien het mogelijk is om de gegenereerde tekst van voetnoten te voorzien, zodat duidelijk wordt waar de gegenereerde tekst vandaan komt. Toch is ook een voorwaarde van CC BY dat aangegeven moet worden of, en zo ja welke, veranderingen zijn aangebracht. Dat lijkt op dit moment nog niet mogelijk te zijn.
- alle documenten met een CC0 zijn in het publieke domein en mogen gebruikt worden voor elk commercieel gebruik, dus ook het trainen van commerciële LLM's.

Het lag toch anders.

Op de website van Creative Commons staat het volgende:²²

"The terms of the Creative Commons (CC) licenses cannot be construed or interpreted as a reservation of a right in the context of Article 4 of the EU Directive on Copyright in the Digital Single Market (EU 2019/790) or any of its national transposition instruments. CC licenses do not operate as an opt-out of an exception or limitation to copyright. Both the language (legal code) and purpose (spirit) of the licenses prohibit such an interpretation."

TDM van zowel art. 3 DSM (TDM voor wetenschappelijk onderzoek) als art. 4 DSM (commercieel TDM) zijn een uitzondering op het auteursrecht.²³ De website van Creative Commons geeft aan dat een CC licentie niet geldig is wanneer een uitzondering op het auteursrecht van toepassing is. En dat een CC licentie niet gebruikt mag worden als een opt-out voor commercieel TDM, met andere woorden, een CC licentie houdt niet in het 'uitdrukkelijke voorbehoud' van art. 4 DSM om commercieel TDM tegen te gaan.

²² <https://creativecommons.org/2021/12/17/creative-commons-statement-on-cc-licenses-and-the-ex-ant-data-mining-exception-under-article-4-eu-cds-m-directive/> Lees vooral ook het artikel waar op deze website pagina naar wordt verwezen, de rechtstreekse link is deze: <https://creativecommons.org/wp-content/uploads/2021/12/CC-Statement-on-the-TDM-Exception-Art-4-DSM-Final-updated.pdf>

²³ Resp. artt. 15n en 15o Auteurswet.

Genoemd artikel leidt tot best wel ingewikkelde situaties.

- Een artikel is rechtmatig toegankelijk op internet via het repositorium van een universiteit en heeft een CC BY-NC licentie. Het trainen van een commercieel LLM valt onder de uitzondering van art. 4 DSM, de CC BY-NC licentie is daarom niet van toepassing, het commerciële bedrijf mag dus wel het artikel gebruiken om het commerciële LLM te trainen. Tenzij de auteur op het voorblad van het artikel of de universiteit op de website van het repositorium (of allebei!) het uitdrukkelijke voorbehoud heeft gemaakt.

Dit is dan zo volgens de letter van de wet en volgens de letter van de CC licentie. Maar hoe zit het met de geest van de wet en de geest van de licentie?

De auteur denkt: ik heb alles afgedekt, want ik heb gekozen voor een NC variant bij mijn licentie, maar nu wordt de auteur toch 'gepakt' en wordt er toch commercieel gebruik gemaakt van zijn artikel!

- Andersom is ook mogelijk. Stel, een artikel is rechtmatig toegankelijk op internet via het repositorium van een universiteit en heeft een CC0 licentie. Onder een CC0 licentie is het werk in het publieke domein. Toch kan het uitdrukkelijke voorbehoud gemaakt op het platform voorkomen of tegenhouden dat het werk gebruikt mag/zal worden voor het trainen van een LLM.

Dan de vraag aan de UKB: wat willen we? Welke rechten willen we geven of onthouden aan commerciële partijen? En in dit geval aan GPT-NL? Want als het zo is dat een CC licentie niet geldt bij een uitzondering, dan kunnen we dus in onze repositoria zetten bij de gebruiksvoorwaarden: wij maken een uitdrukkelijk voorbehoud voor commercieel TDM zoals vereist in art. 4 DSM en dat geldt voor alle artikelen, ongeacht welke CC licentie de artikelen hebben.

4. Taverne artikelen/hoofdstukken

Het auteursrecht op artikelen die via Taverne beschikbaar zijn gesteld in de repositoria komt in (bijna) alle gevallen toe aan de uitgever. SURF zoekt toestemming van de universiteiten om deze teksten te gebruiken voor het trainen van haar model. Maar de universiteiten zijn geen rechthebbenden. De universiteiten kunnen daarom geen toestemming geven aan SURF om deze artikelen te gebruiken voor training van het LLM. Omdat veel uitgevers ertoe zijn overgegaan om het trainen van LLM's aan derde partijen te verbieden (in contracten en op hun websites), verdient het aanbeveling om in onze repositoria duidelijk te vermelden (onder een kopje 'gebruiksrechten') dat artikelen gepubliceerd onder Taverne niet gebruikt mogen worden voor het trainen van LLM's.

De scenario's

De leden van de UKB werkgroep Auteursrecht begrijpen de motivatie van de minister van EZK, TNO, NFI en SURF om te komen tot een zelfstandig LLM dat in Nederland/Europa ontwikkeld en gemaakt wordt en dat Europa een alternatief biedt voor niet-Europese LLM's. De roep om Europese onafhankelijkheid, autonomie en digitale soevereiniteit wordt steeds luider. Het wordt bijzonder op prijs gesteld dat SURF niet zonder toestemming van universiteiten (en andere auteursrechthebbenden) het model gaat trainen. Dit terwijl we weten dat ChatGPT zonder toestemming van welke auteursrechthebbende maar ook het hele wereldwijde web heeft gescraped om haar model te trainen, onder het mom van het Amerikaanse 'fair use'. Dat dit ook in de VS betwist wordt, blijkt uit het aantal rechtszaken dat op dit moment tegen de

makers van ChatGPT in de VS gevoerd wordt. Auteursrecht is een wettelijk recht dat gerespecteerd dient te worden. Maatschappelijke wensen en geopolitieke situaties kunnen nopen tot andere afwegingen en besluiten. Het probleem is echter dat de universiteiten in (bijna) alle gevallen niet de auteursrechthebbenden zijn van de documenten die zich in hun repositoria bevinden.

1. Scenario 1 (meest restrictief)

Bij dit scenario wordt gekozen om het zwaartepunt te laten vallen op het auteursrecht en zoveel mogelijk alle rechten te laten bij de auteur. Universiteiten geven bij alle documenten in hun repositoria aan dat deze niet gebruikt mogen worden voor commercieel TDM en maken een uitdrukkelijk voorbehoud zoals vereist in art. 4 DSM. Dit geldt dan voor scripties van studenten, dissertaties van promovendi, alle artikelen met een CC licentie en alle Taverne artikelen.

2. Scenario 2 (minder restrictief)

In dit scenario wordt toestemming verleend om alle documenten met een CC0 licentie te leveren voor het trainen van het LLM. Documenten waar een andere CC licentie aan verbonden is worden niet geleverd.

3. Scenario 3 (meer open)

In dit scenario wordt toestemming verleend om alle documenten met een CC0 en met een CC BY licentie te leveren voor het trainen van het LLM. We redeneren dan dat SURF/TNO op een bepaalde manier op de website van het LLM informatie verschaft over de documenten die gebruikt zijn ter training van het model, en dat dit voldoende is om aan het attributie vereiste van de CC BY licentie te voldoen.

Documenten waar een andere CC licentie aan verbonden is worden niet geleverd.

4. Scenario 4 (meest open)

In dit scenario worden documenten aangeboden zoals onder scenario 3, alsmede alle scripties van studenten.

In dit scenario mag SURF dan gebruik maken van:

- alle documenten met een CC0 verklaring/licentie;
- alle documenten met een CC BY licentie;
- alle scripties van studenten.

Scenario 4 kan ook een uitsluitend toekomstig scenario zijn. Dat houdt in dat de licentie die de studenten van de universiteit moeten tekenen (om de scriptie te kunnen uploaden in het repositorium) aangepast wordt, en dat daarin een alinea komt te staan met een opt-out voor commercieel TDM gebruik. Dan is de keuze aan de student. Dit heeft twee nadelen:

- niet alleen SURF kan dan gebruik maken van de scripties, maar ook andere commerciële partijen;
- het levert veel administratief werk op voor de universiteiten: zij moeten dan een corpus van scripties maken waar alleen de scripties in staan van studenten die niet van hun opt-out hebben gebruik gemaakt.

Mocht wel van dit scenario gebruik gemaakt gaan worden in de toekomst, dan verdient het aanbeveling hierover landelijke afspraken te maken die gelden voor elk van de 14 universiteiten.

Advies van de UKB Werkgroep Auteursrecht

Op 26 mei heeft de Werkgroep vergaderd.

De structuur van de wet (commercieel TDM is toegestaan op alles wat op internet staat (en dus ook op alles wat zich in universitaire repositoria bevindt), tenzij de auteursrechthebbende gebruik maakt van een opt-out) werd als moeilijk ervaren. Met name omdat er iets van universiteit(sbibliotheek)en wordt verwacht waar zij niet de mogelijkheid toe hebben: UB's/universiteiten zetten materiaal online (via beleid, Taverne of anderszins) waarvan zij niet de auteursrechthebbende zijn. En volgens de wet kan alleen de auteursrechthebbende gebruik maken van een opt-out. Dat zal bij scripties eenvoudig de student zijn en blijven, maar bij artikelen ligt dat anders: eerst is de auteur rechthebbende, vervolgens veelal de uitgever. Dat er dan via een Creative Commons licentie bepaalde rechten aan derden worden gegeven, maakt de zaak er ook niet makkelijker op, omdat eindgebruikers vaak niet begrijpen of die gegeven rechten TDM nu wel of niet mogelijk maken, of dat die gegeven rechten commercieel TDM wel of niet uitsluiten.

De Werkgroep geeft hierna advies vanuit de volgende vier perspectieven:

- ethiek;
- auteursrecht;
- praktische uitvoerbaarheid;
- waarde van het gezamenlijke materiaal dat in repositoria wordt aangeboden.

1. Ethiek

GPT-NL zegt groot belang te hechten aan ethische waarden. Ze wil geen teksten gebruiken zonder toestemming van de auteursrechthebbenden. Voor het trainen van hun LLM zoeken ze naar teksten waarvoor een geldige licentie is gegeven of teksten waarvoor geen toestemming nodig is, omdat de teksten zich in het publieke domein bevinden.²⁴

Universiteiten hebben open access hoog in het vaandel staan. Met publiek geld gefinancierd onderzoek en het daaruit voortvloeiende materiaal moet open toegankelijk zijn voor iedereen. Inclusief grote commerciële partijen.

Echter, veel individuen binnen de universiteiten hebben geen idee wat er allemaal gebeurt in verband met grote commerciële partijen die gebruik maken van hun werk, of, als ze dat wel al doorhebben of doorkrijgen, zijn ze daar niet blij mee. Studenten zullen zich echt niet realiseren dat hun scripties gebruikt worden door deze partijen om een LLM te trainen. Of ze dit erg vinden is de vraag. In de vergadering kwam naar voren dat de AIP's regelmatig benaderd worden door auteurs die er niet blij mee zijn dat hun publicaties gebruikt worden om LLM's te trainen, zij weten ook niet hoe ze zich hiertegen moeten verzetten.

De Werkgroep vraagt UKB zich te realiseren dat de universiteiten / de UB's een verantwoordelijkheid hebben naar individuele studenten en auteurs, aangezien het de universiteiten zijn die van studenten en auteurs 'eisen' dat hun werken online komen te staan in de repositoria. Hier moet zorgvuldig mee worden omgegaan. Nagedacht zou kunnen worden over een workflow om awareness hierover te creëren, en om rekening te houden met opt-outs van de daadwerkelijke (eerste) auteursrechthebbenden.

²⁴ Zie slide 13 in het Progress Report van GPT-NL,
https://gpt-nl.nl/publish/pages/6374/progress_report_1_q42024.pdf

2. Auteursrecht

Zoals in het begin van het advies al gesteld, was auteursrecht ten aanzien van het trainen van LLM's moeilijk. Volgens de letter van de wet is commerciële TDM toegestaan tenzij gebruik wordt gemaakt van een machine-leesbare opt-out. Door de auteursrechthebbende. Niet door de partij die het materiaal online zet, te weten de universiteiten. In hoeverre wordt in dit geval het auteursrecht van de makers geschonden door de acties van de universiteiten? Misschien niet volgens de letter van de wet, maar misschien wel volgens de geest van de wet? Zie ook de overwegingen onder het kopje 'Ethiek'. Hoever moet de rol van de universiteiten gaan? Het volgende onderwerp is daarbij ook van belang, namelijk de praktische uitvoerbaarheid.

3. Praktische uitvoerbaarheid

Indien gekozen wordt voor een workflow waarbij de auteursrechthebbende (student, auteur) een opt-out kan doen voor commercieel TDM, hoe gaat dat er dan uitzien in de repositoria? Moet er een tekst op de website komen te staan bij 'Terms of Use' zoals bijvoorbeeld: *"Sommige van onze auteurs/studenten hebben gekozen voor een opt-out voor commercieel TDM, gebaseerd op artikel 4 van de DSM Richtlijn van de EU, dit kunt u lezen op elk voorblad voorafgaand aan elke scriptie en elk artikel."*? Is dit nog praktisch uitvoerbaar voor universiteiten?

Op dit moment staat in PURE een tekst onderaan (ook genoemd eerder in dit document) dat alle rechten voorbehouden zijn voor alle content op deze website, ook namens de licensors en de contributors. Deze tekst is erop gezet door Elsevier. Dat mag Elsevier helemaal niet doen. Zij hebben niet de rechten op artikelen die daarop staan en bij andere uitgevers zijn gepubliceerd, of waarvan de rechten nog in handen zijn van studenten of medewerkers (auteurs). Als UKB vindt dat commercieel TDM wel mogelijk moet zijn voor alle partijen, dan moet deze tekst er af.

4. Waarde van het gezamenlijk materiaal in repositoria

De dag voor de vergadering van de Werkgroep Auteursrecht werd op TV en elders bekend gemaakt dat Meta alle teksten van de kleine berichtjes op Instagram en Facebook zou gaan gebruiken om een NL taalmodel te trainen. Gebaseerd dus op deze teksten met allerlei soorten taalfouten en slang. Dat zal een mooi taalmodel worden! De teksten die staan in de repositoria van de universiteiten zijn van een veel hogere kwaliteit, en daarom bijzonder waardevol om gebruikt te worden voor het trainen van een LLM of ander AI model. Laten we dit niet vergeten in tijden van bezuinigingen bij de universiteiten. We mogen hier best wel wat voor vragen aan een commercieel bedrijf (in dit geval dan TNO) die een business model gaat maken van GPT-NL.

Conclusie van de Werkgroep

De Werkgroep concludeert dat deze materie voor haar leden nieuw en redelijk onontgonnen terrein is. Wel is het zo dat vragen die in de diverse instellingen hierover gesteld worden, toch wel terecht komen bij de leden van deze Werkgroep en dat zij daarom wel stakeholder zijn. De Werkgroep realiseert zich dat de uiteindelijke beslissing over wel/niet meedoen met GPT-NL en zo ja, in welke mate dan, niet een beslissing is die op het niveau van deze Werkgroep genomen dient te worden. Ook toekomstige beslissingen over het al dan niet meedoen (en onder welke voorwaarden) met andere AI gerelateerde modellen dienen op een ander niveau in de organisaties genomen te worden. Dit kan UKB niveau zijn, of zelfs CvB niveau. Als besloten wordt tot CvB niveau is de Werkgroep van mening dat de CvB's wel heel duidelijk de consequenties van elke optie voorgelegd moeten worden.

Bijlage 1: Artikelen 3 en 4 DSM en artikelen 15n en 15o Aw

Artikel 3

Tekst- en datamining met het oog op wetenschappelijk onderzoek

1. De lidstaten voorzien in een uitzondering op de rechten bedoeld in artikel 5, onder a), en artikel 7, lid 1, van Richtlijn 96/9/EG, artikel 2 van Richtlijn 2001/29/EG, en artikel 15, lid 1, van deze richtlijn voor reproducties en opvragingen door onderzoeksorganisaties en instellingen voor cultureel erfgoed om met het oog op wetenschappelijk onderzoek tekst- en datamining te verrichten op werken of andere materialen waartoe zij rechtmatige toegang hebben.
2. Kopieën van werken of andere materialen die worden gemaakt overeenkomstig lid 1, worden opgeslagen met een passend beveiligingsniveau en mogen worden bewaard voor doeleinden in verband met wetenschappelijk onderzoek, inclusief de verificatie van onderzoeksresultaten.
3. Rechthebbenden kunnen maatregelen nemen om de veiligheid en de integriteit te verzekeren van de netwerken en databanken waar de werken of andere materialen worden gehost. Deze maatregelen gaan niet verder dan wat nodig is om die doelstelling te verwezenlijken.
4. De lidstaten moedigen rechthebbenden, onderzoeksorganisaties en instellingen voor cultureel erfgoed aan om algemeen aanvaarde beste praktijken vast te stellen met betrekking tot de toepassing van de in lid 2 bedoelde verplichting en de in lid 3 bedoelde maatregelen.

Artikel 4

Uitzondering of beperking voor tekst- en datamining

1. De lidstaten voorzien in een uitzondering op of beperking van de rechten als bedoeld in artikel 5, onder a), en artikel 7, lid 1, van Richtlijn 96/9/EG, artikel 2 van Richtlijn 2001/29/EG, artikel 4, lid 1, onder a) en b), van Richtlijn 2009/24/EG en artikel 15, lid 1, van deze richtlijn voor reproducties en opvragingen van rechtmatig toegankelijke werken en andere materialen met het oog op tekst- en datamining.
2. Reproducties en opvragingen op grond van lid 1 mogen worden bewaard zolang dit nodig is voor tekst- en datamining.
3. De uitzondering of beperking als bedoeld in lid 1 is van toepassing op voorwaarde dat het gebruik van de in dat lid bedoelde werken en andere materialen door de rechthebbenden ervan niet op passende wijze uitdrukkelijk is voorbehouden, zoals machinaal leesbare middelen in het geval van online ter beschikking gestelde content.
4. Dit artikel doet geen afbreuk aan de toepassing van artikel 3 van deze richtlijn.

Artikel 15n

- 1** Als inbreuk op het auteursrecht op een werk van letterkunde, wetenschap of kunst wordt niet beschouwd de reproductie door onderzoeksorganisaties en cultureel erfgoedinstellingen om met het oog op wetenschappelijk onderzoek tekst- en datamining te verrichten op het werk waartoe zij rechtmatige toegang hebben.
- 2** De reproductie bedoeld in het eerste lid wordt opgeslagen met een passend niveau van bescherming en mag worden bewaard voor doeleinden in verband met wetenschappelijk onderzoek.
- 3** De makers van werken van letterkunde, wetenschap of kunst en hun rechtverkrijgenden kunnen maatregelen nemen met het oog op de veiligheid en de integriteit van de netwerken en databanken waar die werken zijn opgeslagen. Deze maatregelen gaan niet verder dan wat nodig is om die doelstelling te verwezenlijken.
- 4** Van de bepalingen in dit artikel kan niet bij overeenkomst worden afgeweken.

Artikel 15o

- 1** Onverminderd het bepaalde in artikel 15n wordt een reproductie in het kader van tekst- en datamining niet als inbreuk op het auteursrecht op een werk van letterkunde, wetenschap of kunst beschouwd mits degene die de tekst- en datamining verricht rechtmatig toegang heeft tot het werk en het auteursrecht door de maker of zijn rechtverkrijgenden niet uitdrukkelijk op passende wijze is voorbehouden, zoals door middel van machinaal leesbare middelen bij een online ter beschikking gesteld werk.
- 2** De reproductie gemaakt overeenkomstig het in het eerste lid bepaalde mag worden bewaard zolang dit nodig is voor tekst- en datamining.