

**Auteur:** Abel Hoogeveen is werkzaam als Legal Tech Consultant bij ICTRecht. Hij houdt zich binnen het Legal Tech Team met name bezig met het stroomlijnen en automatiseren van juridische bedrijfsprocessen. Daarnaast ontwikkelt hij regelmatig software om processen binnen ICTRecht en bij klanten te versnellen. Hij is bereikbaar via e-mail: a.hoogeveen@ictrecht.nl.

# Wie bezit GPT-modellen?

Large Language Models, zoals bijvoorbeeld ChatGPT, staan momenteel in het middelpunt van de belangstelling. Het aantal aanbieders van zulke modellen neemt sterk toe, waarbij de ene innovatie de andere innovatie lijkt op te volgen en de taalmodellen steeds beter worden. Een van de belangrijkste innovaties van de afgelopen paar maanden was de ontdekking dat deze modellen van elkaar kunnen leren (1). Door bijvoorbeeld ChatGPT op een slimme wijze uit te vragen is het mogelijk om de eigenschappen en kennis van ChatGPT over te dragen naar open source modellen. Op deze manier kan iedereen voortbouwen op de innovaties van anderen.

**M**et deze techniek is het goedkoper dan ooit om een fatsoenlijk taalmodel te trainen en te gebruiken voor allerlei al dan niet commerciële toepassingen. Het roept echter de vraag op wat hiervan de juridische gevolgen zijn en in het bijzonder de vraag wie zulke modellen nou eigenlijk bezit. ICTRecht gaat hier in dit artikel dieper op in.

Dat OpenAI, de maker van ChatGPT, niet blij was met bovenstaande ontwikkeling viel te verwachten. Het kwam dan ook niet als een verrassing dat OpenAI, Microsoft en Google hebben aangekondigd dat het niet langer toegestaan is om hun modellen te gebruiken voor het trainen van andere modellen (2). Dit nieuwe beleid hebben deze partijen geïncorporeerd in hun algemene voorwaarden. Naast de vraag wat dit met innovatie doet in de open source ruimte is er ook een juridische vraag voor gebruikers. Want als je als bedrijf een commerciële chatbot inkoopt, die gebouwd is op een open source model; en wanneer dat model geleerd heeft van ChatGPT, in hoeverre ben je dan juridisch kwetsbaar?

Het is belangrijk om op te merken dat het verdere trainingsverbod is opgenomen in de algemene voorwaarden van OpenAI en andere partijen. Deze voorwaarden hebben dan ook alleen betrekking op de partijen die eventueel een model verder trainen; en kunnen niet derde partijen bij het verbod betrekken. Als een commercieel bedrijf een taalmodel traint op basis van ChatGPT, dan is alleen dát commerciële bedrijf in overtreding en niet diens klanten.

## Auteursrecht

Maar de vraag is vooral in hoeverre OpenAI beschermd wordt op grond van het auteursrecht en andere intellectuele eigendomsrechten. Het moeilijke daarvan is dat het auteursrecht eigenlijk niet gebouwd is voor een creatie zoals een taalmodel. Het auteursrecht is gemaakt voor geschreven boekwerken, gemaakte muziek en andere creatieve voortbrengselen van de menselijke geest. Een 'large language model', dat bestaat uit allemaal 'neuronen' die patronen herkennen en labels toekennen aan woorden en concepten; dat past niet helemaal binnen de kaders die we kennen. Dus in hoeverre kan het auteursrecht taalmodellen beschermen?

Om bij het begin te beginnen: een auteursrechtelijk werk is ieder voortbrengsel van letterkunde, wetenschap of kunst, op welke wijze of in welke vorm het ook tot uitdrukking is gebracht. Het werk dient een eigen oorspronkelijk karakter te hebben en het resultaat te zijn van creatieve keuzes. Bij taalmodellen is vooral problematisch in hoeverre er sprake is van creatieve keuzes om het model tot stand te laten komen.

De keuze hoe interne neuronen en labels zijn georganiseerd heeft weinig met creativiteit te maken en meer met feitelijke en technische keuzes. Dat een taalmodel opmerkt dat in een zin na een onderwerp vaak een werkwoord volgt, dat is een feitelijke observatie. Dat heeft weinig te maken met creatieve keuzes. Hetzelfde is het geval als een model opmerkt dat Koningin Wilhelmina de Koningin van Nederland was; dat is ook een feitelijke constatering. Natuurlijk maken de makers van taalmodellen hierbij keuzes en zullen sommige modellen het daardoor beter of slechter doen, maar dit betreffen slechts rationale wetenschappelijke keuzes en weinig schepping vanuit de creativiteit. Het is dan ook niet waarschijnlijk dat de werking van een large language model beschermd kan worden door het auteursrecht.

### Databankenrecht

Het auteursrecht biedt dus waarschijnlijk geen bescherming aan taalmodellen, maar kan het databankenrecht misschien soelaas bieden? Een databank is een verzameling van gegevens die systematisch geordend zijn en waarbij het maken daarvan een substantiële investering nodig had (3). Het recht heeft als doel bedrijven te beschermen die met veel geld en moeite een databank of dataset maken; zodat niet iedereen daar zonder compensatie zomaar mee vandoor kan gaan.

Large Language Models passen best goed binnen die definitie. Ze zijn heel anders opgebouwd dan traditionele databases waarvoor de wet is gemaakt, maar de definitie is breed genoeg dat zo een taalmodel er waarschijnlijk wel onder valt. Het concept 'substantiële investering' is niet exact vastgelegd in de wet, maar gezien de enorme financiële investering en moeite die bedrijven als OpenAI en Google hebben gestoken in de ontwikkeling van taalmodellen vallen deze vrijwel zeker onder de beoogde beschermingsruimte.

Maar zelfs als taalmodellen onder het databankenrecht vallen, is nog maar de vraag of dit recht ook beschermt tegen de manier waarop andere modellen daarvan leren.

Het databankenrecht beschermt op twee wijzen tegen het maken van kopieën van de databank (4). Ten eerste beschermt het tegen het maken van een gehele kopie; en ten tweede beschermt het tegen het maken van een kopie van een 'substantieel deel'. Het overnemen van kennis en vaardigheden van taalmodellen ziet op dat tweede, maar er is niet echt sprake van het maken van letterlijke kopieën.

Wanneer een open source taalmodel leert van ChatGPT dan stelt deze vragen aan ChatGPT net zoals wij dat zelf ook doen. Het vraagt niet: 'geef mij een lijst van al je labels en neuronen', maar het stelt simpelere vragen zoals: 'Geef het proces weer van de evolutietheorie' of 'Vertel mij hoe een koelkast werkt'. ChatGPT geeft hierop dezelfde antwoorden zoals het doet aan iedere andere gebruiker. Het open source taalmodel analyseert deze antwoorden en leert hier uit de kennis en patronen hoe het antwoord in elkaar steekt. Hierdoor leert het langzaam hoe het ChatGPT kan nabootsen. Is dit te kwalificeren als het maken van een kopie? Ik durf dat niet met vertrouwen te zeggen.

Wat het antwoord op de vraag ook is, het zal OpenAI niets helpen. Het databankenrecht beschermt alleen databanken opgezet door bedrijven in de Europese Unie. OpenAI is tot de dag van vandaag een Amerikaans bedrijf en geniet dan ook geen bescherming op basis van dit recht. Toch biedt het databankenrecht kansen voor bescherming van commercieel ontwikkelde GPT-modellen.

### Conclusie

De ontwikkelingen van de afgelopen maanden laten zien dat het mantra van Silicon Valley - 'Go fast and break things' - vol in leven is. Voor even leek het of de voorsprong van OpenAI zo groot was dat niemand deze meer in zou kunnen halen; en nu korte tijd later is dat vertrouwen verdampt. De schrik zit erin en OpenAI en Google halen de ophaalbrug omhoog om hun koninkrijk te beschermen. Of het genoeg is, is nog maar de vraag. Want met zulke rappe ontwikkelingen is het onmogelijk om te zeggen wie morgen de kroon van beste large language model draagt.

### Referenties

- (1) <https://arxiv.org/abs/2212.10560>
- (2) <https://www.businessinsider.com/openai-google-anthropic-ai-training-models-content-data-use-2023-6?international=true&r=US&IR=T>
- (3) Artikel 1 lid 1 onder a Databankenwet
- (4) Artikel 2 lid 1 onder a Databankenwet