



Universiteit
Leiden
The Netherlands

WetGPT? Kan een large language model wetten schrijven?

Meuwese, A.C.M.

Citation

Meuwese, A. C. M. (2023). WetGPT?: Kan een large language model wetten schrijven? *Regelmaat: Kwartaalblad Voor Wetgevingsvraagstukken*, 38 [abusievelijk 39](1), 90-97.
doi:10.5553/RM/0920055X2023039001008

Version: Publisher's Version
License: [Leiden University Non-exclusive license](#)
Downloaded from: <https://hdl.handle.net/1887/3728851>

Note: To cite this publication please use the final published version (if applicable).

WetGPT? Kan een Large Language Model wetten schrijven?

A.C.M. Meuwese

RegelMaat staat niet bekend als een blad dat zich veel aantrekt van *hypes*, maar soms kan ook een tijdschrift voor wetgevingsvraagstukken er niet omheen. Over het AI-systeem ChatGPT¹ zijn al allerlei voorspellingen gedaan: het programma gaat het onderwijs op zijn kop zetten, zoekmachines ouderwets doen lijken en tekstschrijvers buiten spel zetten. Tekstschrijvers? Vallen wetgevingsjuristen daar ook onder? Met andere woorden: kan ChatGPT wetsbepalingen ontwerpen?

Het zou voor de hand liggen deze vraag door middel van een empirisch onderzoek te toetsen door een grote hoeveelheid opdrachten tot het maken van wetgeven in te voeren in de chatbot-interface van ChatGPT. En inderdaad kunt u verderop in dit stuk voorbeelden vinden van door ChatGPT geschreven bepalingen. Ik heb mij echter beperkt tot een klein experiment, eigenlijk meer een voorbeeld. Hoewel er groot enthousiasme is over ChatGPT zijn er ook nadelen. Zo worden regelmatig feitelijke onjuistheden en verzinsels geconstateerd. Ook is het aannemelijk dat veelvuldig gebruik van ChatGPT niet milieuvriendelijk is. Het is niet eenvoudig betrouwbare bronnen te vinden over de exacte milieuoetafdrak van dit programma, maar in de universitaire wandelgangen doet het cijfer van duizend keer het energieverbruik van een Google-zoekopdracht de ronde. Een ander nadeel is dat eenzelfde verzoek niet per definitie dezelfde output oplevert, wat verklaart waarom traditionele plagiaatcontroleprogramma's niet werken bij door ChatGPT gegenereerde teksten. Een oordeel over de vaardigheden van het programma in relatie tot een bepaalde taak is dus altijd een momentopname. Ten slotte is het lastig om gebreken en successen van een AI-systeem te duiden als er geen transparantie is over de keuzes die het algoritme maakt. Wat dat betreft is ChatGPT een klassiek voorbeeld van een *black box*. Omdat de output voor ons vertrouwd overkomt (afgebakende teksten geschreven in behoorlijk Nederlands) zijn we geneigd deze in dezelfde categorie te plaatsen als door professionals geproduceerde teksten. De foutieve aanname dat aan gemaakte keuzes bij het opstellen van de tekst min of meer menselijke redeneringen ten grondslag liggen, is snel gemaakt. We wéten wel dat het allemaal gebaseerd is op statistiek ('welk woord heeft de meeste kans om te volgen op de tekst die er al ligt?'), maar zo voelt het niet.

Maar dit is geen stuk over de capaciteiten van juristen om met statistiek overweg te kunnen – een onderzoeksgebied op zichzelf. Mijn bedoeling is om een indruk te geven van wat we van ChatGPT mogen verwachten rond de activiteit 'wetsbepalingen ontwerpen'. Laten we vooropstellen dat dit geen activiteit is die de makers ex-

1 <https://chat.openai.com/chat>.

pliciet benoemen of uitsluiten – zoals wel expliciet wordt vermeld dat ChatGPT niet geschikt is om adviezen te geven. Tegelijkertijd is al sinds de lancering in november 2022 duidelijk dat het programma overweg kan met een breed scala aan opdrachten, van computercode schrijven tot het genereren van een lijstje ideeën voor een kinderfeestje.

Voordat we toekomen aan een indruk van ChatGPT als wetgever, een korte inleiding op het fenomeen Large Language Models (LLM). Wat zijn en kunnen ‘grote taalmodellen’ nu eigenlijk? Large Language Models zijn algoritmes die gebruikmaken van *deep learning* op basis van zeer grote datasets. Deze datasets kunnen bestaan uit ‘natuurlijke taal’, maar ook uit andere ‘taal’, zoals bijvoorbeeld programmeertaal of eiwitsequenties. Daarmee tillen LLMs bestaande Natural Language Processing (NLP)-toepassingen, zoals chatbots, naar een hoger niveau en groter bereik, maar gaan zij ook verder dan ‘natuurlijke taal’. Het ‘leren’ van een LLM verloopt deels zonder menselijke supervisie. De hoofdtaak van ChatGPT bestaat uit het telkens voorspellen van het volgende woord. Dit is een moeilijke taak die behalve een woordenschat ook kennis van grammatica, betekenis, taalgebruik, en de wereld vereist. Een voldoende krachtig model met voldoende data om op te trainen zal (een vorm van) de benodigde kennis uit de data destilleren. Het algoritme leidt dan bijvoorbeeld uit contextuele informatie zelf af dat een bepaald woord twee verschillende betekenissen kan hebben. Daarnaast worden ook menselijke trainers ingezet om het model te verfijnen en om schadelijke inhoud eruit te vissen. Deze laatste activiteit heeft tot kritische geluiden geleid: laagbetaalde werknemers in Kenia hebben grote hoeveelheden schadelijke teksten uit de ‘donkere hoeken’ van het internet moeten verwerken teneinde de detectietool van OpenAI te trainen.²

Dan het experiment. Ik koos de casus van de gevaarlijkehondenproblematiek, een bekende in de reguleringstheorie, omdat over dit onderwerp als zodanig nog geen wetgeving bestaat in Nederland. Tegelijkertijd, om ChatGPT een eerlijke kans te geven, betreft het een onderwerp waarvan het niet heel vreemd zou zijn er wetgeving over te maken en waarover niet veel specialistische kennis nodig is om je daar iets bij voor te stellen.

Voorbeeld 1 Door ChatGPT gegenereerde output op basis van de opdracht ‘schrijf een wetbepaling die gevaarlijke honden verbiedt’, 13 februari 2023

Artikel 1 Verbod op gevaarlijke honden

1.1. In deze wet wordt verstaan onder een ‘gevaarlijke hond’: elke hond die, door haar aard of opvoeding, een direct gevaar vormt voor de veiligheid of gezondheid van personen, dieren of zaken.

1.2. Het is verboden om een gevaarlijke hond in Nederland te houden, te bezitten of onder zijn controle te hebben, tenzij de eigenaar of verzorger in het bezit is van een geldige vergunning, afgegeven door de gemeente waarin de hond verblijft.

2 B. Perrigo, ‘Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic’, Time.com, 18 januari 2023 (<https://time.com/6247678/openai-chatgpt-kenya-workers/>).

1.3. Om in aanmerking te komen voor een vergunning dient de eigenaar of verzorger aan te tonen dat hij over de nodige deskundigheid beschikt om de hond op een verantwoorde en veilige manier te houden en te beheren.

1.4. Een vergunning kan door de gemeente worden ingetrokken als blijkt dat de eigenaar of verzorger niet in staat is om de hond op een verantwoorde en veilige manier te houden of als de hond in strijd handelt met de bepalingen van deze wet.

1.5. Het is verboden om een gevaarlijke hond zonder toezicht of begeleiding op een openbare plaats te laten lopen.

1.6. Het is verboden om een gevaarlijke hond te gebruiken voor agressieve sporten of trainingen.

1.7. De eigenaar of verzorger van een gevaarlijke hond is verplicht om de hond te registreren bij de gemeente en de hond te laten chippen.

1.8. Overtreding van de bepalingen van deze wet is strafbaar en kan leiden tot een boete of het intrekken van de vergunning.

1.9. Deze wet treedt in werking op een bij koninklijk besluit te bepalen tijdstip.

Het mag duidelijk zijn dat op deze artificiële wettekst voldoende aan te merken is. Interessant aan dit experiment is wat er goed gaat en wat er minder goed gaat. Je zou namelijk zeggen dat voor een programma dat leert uit grote hoeveelheden tekst waaronder bestaande wetten, jargon relatief eenvoudig beheersbaar is. Het is dus opvallend dat ‘de gemeente’ de vergunning verleent en niet een bestuursorgaan van een gemeente. Ook is bijvoorbeeld de delegatiesystematiek niet bepaald levensecht, blijft de ‘strafbaarheid’ wel erg onbepaald en bestaat de hele ‘wet’ uit één artikel. Op wetgevingstechniek, systematiek en begrijpelijkheid (bijvoorbeeld, ‘gelden de verbodsbepalingen 1.5 en 1.6 ook als de eigenaar een vergunning heeft?’) scoort ChatGPT matig. ‘De hond’ is volgens het programma blijkbaar een vrouwelijk woord en kan zelf ook rechtssubject zijn (zie 1.4, ‘als de hond in strijd handelt met de bepalingen van deze wet’, die verder netjes op de eigenaar of verzorger gericht is). Tegelijkertijd komt ChatGPT inhoudelijk met best interessante suggesties, overigens zonder dat het programma veel blijk geeft van besef van de handhavingsproblematiek. Wat we niet kunnen achterhalen: stelt ChatGPT een vergunningensysteem voor voor gevaarlijke honden, omdat dat nu eenmaal veel voorkomt in bestaande (wet)teksten? Of ligt hier een geavanceerdere berekening aan ten grondslag? Het programma bevat geen knop met ‘geef toelichting’. Die zal er gezien de ontwikkelingen in het deelgebied ‘*Explainable AI*’ wellicht wel komen, maar niet van vandaag op morgen.

Deze uitkomst (relatief slecht op formele aspecten, relatief goed op de inhoud) is mogelijk deels te verklaren doordat ChatGPT niet specifiek op wetteksten is getraind. Het lijkt erop dat ChatGPT beter werkt bij Engelstalige opdrachten, simpelweg omdat er meer Engelstalige tekstdata, en dus misschien ook wel meer wetgeving, in het model zijn gestopt. Vandaar dat ik ter vergelijking ook een Engelstalige wetgevingsopdracht heb ingevoerd. In het Verenigd Koninkrijk bestaat al wel wetgeving op het gebied van gevaarlijke honden.³ Het ChatGPT-resultaat naar aanlei-

3 Dangerous Dogs Act 1991.

ding van een Engelstalige versie van mijn opdracht is, wellicht niet geheel toevallig, langer dan bovenstaande Nederlandstalige 'wet', maar een stuk korter en minder genuanceerd dan de echte Britse wet.

Voorbeeld 2 Door ChatGPT gegenereerde output op basis van de opdracht 'draft a statutory provision banning dangerous dogs', 13 februari 2023

Draft Statutory Provision Banning Dangerous Dogs

Section 1. Purpose The purpose of this statutory provision is to ensure public safety by regulating the ownership and control of dangerous dogs within the jurisdiction.

Section 2. Definitions For the purposes of this statute, the following terms shall have the meanings ascribed to them:

- 1 'Dangerous Dog' means any dog that has a known history of biting, attacking, or causing injury to humans or other animals, or any dog that is trained for fighting or aggressive behavior.
- 2 'Owner' means any person who keeps, harbors, or has custody or control of a dog.
- 3 'Control' means the restraint of a dog by a leash, chain, or other similar device.

Section 3. Requirements for Owners of Dangerous Dogs The owner of a dangerous dog shall comply with the following requirements:

- 1 The owner shall register the dangerous dog with the local animal control agency.
- 2 The owner shall secure the dangerous dog at all times, either indoors or in a secure, enclosed outdoor area.
- 3 The owner shall secure the dangerous dog with a leash, chain, or other similar device whenever the dog is off the owner's property.
- 4 The owner shall display a warning sign on their property warning of the presence of a dangerous dog.
- 5 The owner shall obtain liability insurance in an amount not less than \$100,000 to cover any damage or injury caused by the dangerous dog.

Section 4. Prohibited Acts The following acts are prohibited:

- 1 Abandoning or surrendering a dangerous dog to an animal shelter, rescue organization, or other entity without providing written notification of the dog's dangerous nature.
- 2 Selling, giving away, or transferring ownership of a dangerous dog without providing written notification of the dog's dangerous nature to the new owner.
- 3 Allowing a dangerous dog to run at large or roam free, unsupervised and uncontrolled.

Section 5. Enforcement

- 1 The local animal control agency is responsible for enforcing this statutory provision.
- 2 Any person who violates any provision of this statute may be subject to a fine or imprisonment, or both, as determined by the court.
- 3 Any dangerous dog that is seized or impounded due to a violation of this statute may be subject to destruction or other disposition as deemed appropriate by the animal control agency.

Section 6. Severability If any provision of this statute is held invalid or unconstitutional by a court of competent jurisdiction, the remaining provisions shall remain in full force and effect.

Section 7. Effective Date This statute shall become effective on [insert date].

Dan is er nog een vermeldenswaardig praktijkexperiment dat we bij de Nederlandse wetgever toch niet zo snel zullen tegenkomen. In de Verenigde Staten heeft een *state senator* voor de staat Massachusetts, Barry Finegold, een wetsvoorstel ingediend over de regulering van AI chatbots, dat geschreven is met behulp van ChatGPT.⁴ Blijkbaar weigerde ChatGPT eerst Finegold's verzoek, maar reageerde na meerdere pogingen toch met een tekstvoorstel dat volgens de senator '70 percent there' was.⁵ Finegold's ambtelijke staf heeft het ontwerp nog wel verfijnd en geredigeerd en heeft van dit proces screenshots bewaard om met journalisten te delen. Een hoge ambtenaar die in het in het notenapparaat aangehaalde *Washington Post*-artikel aan het woord komt, rapporteert dat ChatGPT wel degelijk met een aantal originele, inhoudelijke ideeën kwam, zoals over de-identificatie en dataveiligheid. Die bevinding komt wel overeen met mijn gevaarlijkheidsexperiment, waarbij de hierboven weergegeven producten (voorbeelden 1 en 2) overigens het resultaat waren van de eerste de beste poging. Hieronder volgt de tekst van het wetsvoorstel, zoals dat dus ook door Finegold officieel is ingediend.

*Voorbeeld 3 Door ChatGPT gegenereerde en vervolgens door mensen bewerkte output*⁶

SECTION 1. The General Laws are hereby amended by inserting after chapter 93A the following chapter:

CHAPTER 93A½. Generative Artificial Intelligence Models

Section 1. Purpose

The purpose of this chapter is to regulate generative artificial intelligence models, such as ChatGPT, in order to protect the public's safety, privacy and intellectual property rights.

Section 2. Definitions

4 Bill SD.1827, An Act drafted with the help of ChatGPT to regulate generative artificial intelligence models like ChatGPT (<https://malegislature.gov/Bills/193/SD1827>).

5 C. Lima, 'ChatGPT is now writing legislation. Is this the future?', *The Washington Post* 23 januari 2023.

6 Zie noot 4.

For the purposes of this chapter, the following words shall have the following meanings, unless the context clearly requires otherwise:

- a A 'large-scale generative artificial intelligence model' shall mean a machine learning model with a capacity of at least one billion parameters that generates text or other forms of output, such as ChatGPT.
- b 'Parameter' shall mean any variable or value used to control the operation or output of a generative artificial intelligence model.

Section 3. Operating Standards

Any company operating a large-scale generative artificial intelligence model shall adhere to the following operating standards:

- 1 the model shall not be used to engage in discrimination or bias against any individual or group based on protected characteristics, as defined by state or federal law;
- 2 in order to prevent plagiarism, the model shall be programmed to generate all text with a distinctive watermark or offer an authentication process that allows a user to determine whether a particular output was generated by the model;
- 3 the company shall implement reasonable security measures to protect the data of individuals used to train the model;
- 4 the company shall obtain informed consent from individuals before collecting, using or disclosing their data;
- 5 the company shall delete or de-identify any data collected from individuals if it is no longer needed for the intended purpose of the model; and
- 6 the company shall conduct regular risk assessments to identify, assess and mitigate reasonably foreseeable risks and cognizable harms related to their products and services, including in the design, development and implementation of such products and services.

Section 4. Registration with the Attorney General

- a Any company operating a large-scale generative artificial intelligence model shall register with the attorney general within 90 days of the effective date of this act.
- b The registration shall include the following information:
 - 1 the name and contact information of the company;
 - 2 a description of the large-scale generative artificial intelligence model, including its capacity, training data, intended use, design process and methodologies; and
 - 3 information on the company's data collection, storage and security practices.
- c The attorney general shall maintain a public registry of all companies registered under this act.

Section 5. Enforcement

- a The attorney general shall adopt regulations for the purposes of carrying out this chapter.
- b To remedy violations of this chapter and for other relief that may be appropriate, the attorney general may bring an action pursuant to section 4 of chapter 93A against a person.

SECTION 2. Effective Date

Chapter 93A½. of the General Laws shall take effect on the ninetieth day following the passage of this act.

SECTION 3. Disclaimer

This act has been drafted with the help of ChatGPT and any errors or inaccuracies in the bill should not be attributed to the language model but rather to its human authors.

Het is mij onduidelijk of het invoegen van een ‘Chapter 93A½’ overeenkomstig de stijlrichtlijnen van de wetgevende macht van Massachusetts is of dat dit een eigenaardigheid is van ChatGPT die men expres heeft laten staan. De journalist van de *Washington Post* die over deze *gimmick* rapporteert, vergelijkt het gebruik van ChatGPT door de senator met het invoeren van hulp bij het ontwerpen van wetgeving van belangengroeperingen of commerciële partijen.

Het voorspelbare antwoord op de vraag of ChatGPT wetteksten kan produceren (echt ‘schrijven’ doet het programma dus niet) is ‘ja en nee’. Moeten wetgevingsjuristen toch massaal aan de ChatGPT, bij wijze van ondersteuning van hun creatieve proces? Op dit moment zeker niet. Niet alleen is nog veel onduidelijk over belangrijke zaken als milieu-impact, *bias* en de ethiek bij OpenAI, het bedrijf achter ChatGPT, ook lijkt de meerwaarde voorlopig beperkt. In specifieke gevallen zou een wetgevingsjurist door een door ChatGPT gegenereerde oplossing wellicht wel eens op nieuwe ideeën en invalshoeken kunnen komen.

De verwachting is dat Large Language Models nog sterk zullen verbeteren in de komende jaren. Daarnaast zijn er ook veel kleinschaligere juridische NLP-toepassingen, die juist door hun focus en beperkte reikwijdte beter zullen zijn in bepaalde taken. Het WetSuite-project, waar ik bij betrokken ben, richt zich bijvoorbeeld specifiek op het faciliteren van kwantitatief onderzoek op basis van Nederlandse overheidsdocumenten, waaronder ook wetgeving. Daarbij kan men denken aan patroonherkenning, clustering en het geautomatiseerd linken van verschillende typen documenten. LLMs kunnen ook gebruikt worden als basis voor dit soort specifieke toepassingen, al zal dit voor kleinschalige *legal tech*-toepassingen niet altijd een haalbare kaart zijn. De projecten RegelSpraak en Calculemus/FLINT⁷ kiezen een iets andere aanvliegroute voor het ontwikkelen van technologie die regels kan ontwerpen. Door de inzet van formele taal kunnen regels makkelijker naar digitale systemen worden omgezet. Deze technologie zal niet snel zo laagdrempelig

7 Zie A.C.M. Meuwese & I. Timmer, “Law smells”. Over digitale vertalingen van regelgeving en de uitvoeringspraktijk’, *RegelMaat* 2022, afl. 3, p. 272-283.

in gebruik worden als ChatGPT en is ook niet bedoeld om inhoud te genereren, maar een toekomstige integratie met *generative* AI-systemen zoals ChatGPT is potentieel zeer krachtig.