



Foto: Tom van Limpt

# Met ChatGPT wordt taalbewustzijn nóg belangrijker

Verbannen of omarmen? In de huidige discussies over het gebruik van ChatGPT op school lijken slechts twee posities mogelijk. Hoe houdbaar zijn die? En wat is de impact van de nieuwste *large language models* voor het talenonderwijs? ‘AI is geen magie...’

MARJOLIYN VOOGEL

‘A fox reading a newspaper surrounded by flowers, digital art.’ Dat is de *prompt* van de redactie van schoolkrant *Vulpes* voor het verkrijgen van de afbeelding op de cover van de zomereditie 2023. Het AI-programma Dall-E genereerde beelden van een vos met krant, van voren en op de rug. De docente die mij wees op deze *Vulpes*-editie, deed dat met een zekere trots: kijk hoe creatief mijn leerlingen zijn! *Vulpes* is de schoolkrant van het Vossius Gymnasium te Amsterdam. Ik lees daarin ook een groot, zelf-

geschreven, en goed geïnformeerd stuk over de invloed van kunstmatige intelligentie (artificial intelligence, of: AI), inclusief een verslag van een leerlingenpoll. Daaruit blijkt dat 56 procent van de Vossius-leerlingen ‘weleens ChatGPT gebruikt voor een schoolopdracht’. Ook het Grote Jongerenonderzoek 2023 van marktonderzoeksbureau Mediatest laat zien dat 15 procent van de jongeren tussen 15 en 21 regelmatig gebruikmaakt van AI-software; 41 procent doet dat af en toe. ChatGPT is verreweg de meest populaire tool, dat volgens hetzelfde onderzoek vooral voor school wordt gebruikt.

Veel docenten zullen daar minder blij mee zijn, sterker nog: het toenemende gebruik van ChatGPT baart menigeen zorgen. Niet alleen afbeeldingen, maar ook veel schoolopdrachten zijn met een simpele *prompt* in enkele minuten te genereren. Mijn zoons, alle drie uit de categorie 15–21 jaar, zijn eveneens grootgebruikers. In zijn eindexamenjaar kwam de oudste uitgelaten thuis: ‘Mam, dankzij Chat had ik een zes voor mijn schoolexamen Engels!’

Als wordt gesproken over kansen en bedreigingen van AI op school, gaat het sinds eind 2022 dan ook vooral over

ChatGPT. 30 november 2022 was het moment waarop het bedrijf OpenAI, nauw gelieerd aan Microsoft, dit zogeheten *large language model* (LLM) gratis ter beschikking stelde aan het publiek. En dat probeerde het massaal uit: tussen december 2022 en maart 2023 bezochten ruim 1,4 miljoen Nederlanders van 13 jaar of ouder de website van OpenAI. Velen waren *stupéfait* bij het zien van de prestaties van ChatGPT, dat verbluffend vloeiende taal genereert en uitstekend vertaalt. Zo slaagde ChatGPT voor het havo-eindexamen Nederlands, schreef taalkundige Marc van Oostendorp (2023b) op *Neerlandistiek*. Ook voor

## Met deze steeds geavanceerdere technieken is het dus niet vreemd dat voor veel docenten een zelfgeschreven schoolopdracht niet meer te onderscheiden is van een stuk dat door ChatGPT is gegenereerd

het vwo-examen, dat ChatGPT nét niet haalde, waren 'best veel antwoorden (...) verbazingwekkend goed' (Van Oostendorp, 2023a).

### Breinsimulatie

Voor wetenschappers waren de prestaties van GPT 3.5 (de versie die ten grondslag ligt aan de gratis variant van ChatGPT; GPT 4 wordt gebruikt voor de betaalde variant) minder verrassend. In zijn boekje *Slim, slimmer, slimst* (2023) vertelt wetenschapsjournalist Bennie Mols de geschiedenis van artificial intelligence, een term die in 1956 werd bedacht door John McCarthy. Een zomerproject in datzelfde jaar, zo'n tien jaar na de geboorte van de computer, wordt gezien als de start van het vakgebied. Het doel: menselijke intelligentie in een machine realiseren. Daarin zijn sindsdien behoorlijke stappen gezet. Mijnpalen waren 1997, toen IBM's supercomputer Deep Blue voor het oog van de wereld de schaker Kasparov versloeg. In maart 2016 won go-computer AlphaGo van Lee Sedol, de beste go-speler allertijden.

Die vooruitgang is ook te zien in wat Mols de twee AI-hoofdstromingen noemt, 'machinaal redeneren' en 'machinaal leren'. Machinaal redeneren is gebaseerd op het idee dat intelligente systemen functioneren door symbolen volgens logische regels te verwerken. Tot de jaren tachtig was dit binnen AI de belangrijkste stroming. De tweede stroming, machinaal leren, gaat verder. Met neurale netwerken ging men de werking van het menselijk brein simuleren. Ons brein bestaat uit een enorm netwerk van zo'n honderd miljard hersencellen of neuronen. De verbindingen daartussen worden tijdens het leerproces sterker of zwakker, en dat bootst men met computers na. Neurale netwerken van kunstmatige neuronen leren op basis van een heleboel voorbeelden. Zo'n netwerk wordt verdeeld in lagen, die elk een iets andere functie hebben. De sterkte van verbindingen tussen de neuronen verandert tijdens het leerproces (Mols, 2023). De afgelopen jaren werden de leerprocessen van die neurale netwerken steeds diepgaander, of liever gezegd: gelaagder. Hun vermogen om patronen te herkennen verbeterde sterk, door correctie van de

output door mensen maar vooral door training met enorme hoeveelheden data en met innovatieve neurale netwerkarchitectuur, zo vertelt Jelle Zuidema tijdens een door AlfaGamma- en Bètapartners georganiseerde bijeenkomst over de impact van ChatGPT in het voortgezet onderwijs. Zuidema is docent natuurlijke taalverwerking aan de Universiteit van Amsterdam. Voor de meeste grote LLM's wordt de transformerarchitectuur gebruikt, speciaal geschikt voor taalverwerking. GPT staat dan ook voor: Generative Pretrained Transformer. Al die ontwikkelingen hebben de afgelopen jaren geleid tot een gigantische schaalvergroting van de zelflerende taalmodellen.

### Babelvis

Een taalmodel als ChatGPT leert namelijk om zijn antwoorden zo nauwkeurig mogelijk te voorspellen. Zuidema geeft een voorbeeld. In de zin: *It is warm outside, open the ...* zou ChatGPT kunnen aangeven dat er 0,24 procent kans is dat het woord *ask* zal volgen, 1 procent kans dat het *what* moet zijn, 10 procent kans dat het *door* moet zijn en 40 procent kans dat het *window* is. Het taalmodel kiest dan voor het woord met de grootste kans, hier dus *window*. Beeldgeneratie werkt volgens hetzelfde voorstelprincipe. Mols: 'Een taalmodel als ChatGPT wordt ook wel "stochastische papegaai" genoemd, omdat het op basis van statistische patronen in grote hoeveelheden tekst nieuwe tekst genereert zonder zelf te begrijpen wat de tekst betekent en zonder een model te hebben van de wereld' (p. 87). Dat laatste wordt lastiger om je blijvend te realiseren, nu ChatGPT zich in taaluitingen steeds meer als mens lijkt te gedragen. Zo toonden onderzoekers van de universiteit van Berkeley aan dat GPT 4 recursie kent, dat door taalwetenschappers vaak als dé definiërende eigenschap van menselijke taal wordt gezien (Dąbkowski & Beguš, 2023). Recursie houdt in dat je zinnen en woord(groep)en kunt inbedden in andere zinnen, zoals in: *ik schrijf is een zin, die onderdeel is van jij denkt dat ik schrijf, die weer onderdeel is van Lina zegt dat jij denkt dat ik schrijf... et cetera.*

Met deze steeds geavanceerdere technieken is het dus niet vreemd dat voor veel docenten een zelfgeschre-

ven schoolopdracht niet meer te onderscheiden is van een stuk dat door ChatGPT is gegenereerd. Hoe erg is dat? Is het met deze nieuwe techniek überhaupt nog nodig om leerlingen te leren schrijven? Of zelfs om nog vreemde talen te leren, nu de Babelvis binnen handbereik is? Dat kleine, bloedzuigerachtige ding uit het sciencefictionboek *The Hitchhiker's Guide to the Galaxy* van Douglas Adams uit 1979, waardoor als je het in je oor steekt, je meteen begrijpt wat er tegen je wordt gezegd, in welke taal dan ook?

### Taallessen met ChatGPT

Aandacht voor taalonderwijs is er ook, op de bijeenkomst van de AlfaGamma- en Bètapartnersbijeenkomst over ChatGPT. Paul Halma is docent Duits aan het A. Roland Holst College te Hilversum en verbonden aan de lerarenopleiding van de Hogeschool van Amsterdam, waar hij onderzoek doet naar de inzet van chatten in taalonderwijs. Hij begeleidt zijn leerlingen bij het maken van een campagne in het Duits, en helpt hen met het formuleren van de juiste prompts in ChatGPT: 'Geef vijf zinnen die ik in het Duits kan gebruiken wanneer ik over mijn campagne-ideeën praat. B1-niveau.' Ik tik de prompt in met 'Frans' en krijg behulpzame input van ChatGPT. Dat geldt ook voor de volgende door Halma ingefluisterde prompts: 'Geef de tien belangrijkste woorden uit de vorige vijf voorbeeldzinnen', en: 'Geef drie verschillende typen oefeningen om de tien belangrijkste woorden te oefenen'. Het bedenken en formuleren van de juiste prompts, ook wel *prompt engineering* genoemd, vertelde Zuidema eerder ook al, wordt een steeds belangrijker vaardigheid. Naast digitale vaardigheid is ook taalvaardigheid hier cruciaal.

Maurice Schols is verbonden aan de lerarenopleiding van de Universiteit van Amsterdam en laat ons een nieuwe speech-plugin van ChatGPT zien, waarmee je leerlingen gesprekken kunt laten voeren met ChatGPT. Zijn startprompt luidt: 'Imagine we are two strangers who meet each other at Victoria Station in London. We start a random conversation and I will answer your questions in a casual way. Can you start by asking me a question?' Schols laat ons zien hoe vloeiend zijn gesprek met Chat verloopt, en hoe hij het model in verschillende accenten (Oxford of Amerikaans-Engels) antwoorden en vragen kan laten genereren. De Babelvis is inderdaad dichtbij, en helpt in dit geval om juist vaardiger te worden in een andere taal.

In de sessie 'Poëzie en ChatGPT' laat Martijn Koek, hoofddocent talendidactiek aan de Hogeschool van Amsterdam, goed zien hoezeer digitale vaardigheid en

### PROBEREN MET PROMPTS

Als ik ChatGPT op mijn smartphone vraag om aan een talendocent uit te leggen wat een *large language model* is, krijg ik een redelijk adequaat antwoord: 'Een large language model is een geavanceerd computeralgoritme dat getraind is op grote hoeveelheden tekstdata en in staat is om mensachtige tekst te genereren, vragen te beantwoorden en natuurlijke taal te begrijpen. Het fungeert als een virtuele gesprekspartner die taaluitingen kan produceren en begrijpen op basis van de patronen en de kennis die het heeft geleerd.'

Maar als je vraagt waar je de Engelse koning kan bezoeken, dan blijft ChatGPT het juiste antwoord schuldig. Taalmodellen als ChatGPT en Google Bard, een AI-chatbot voor conversaties die gebruikmaakt van PaLM 2, zijn geen zoekmachines als Google of Bing. ChatGPT is getraind met webteksten tot 2022. Om welke dataset het precies gaat, daar is OpenAI niet zo open over, hoogstwaarschijnlijk omdat auteursrechten zijn geschonden.

Wel geeft ChatGPT vloeiend antwoord op de vraag naar de Engelse koning. Door die altijd vloeiende antwoorden zijn wij sterk geneigd om de chatbot mensachtige eigenschappen toe te kennen. Antropomorfisme ligt steeds op de loer. Maar, aldus Bennie Mols (2023) in zijn eerdergenoemde boekje: 'AI is geen magie. Elke AI-toepassing wordt nog steeds ontworpen, gebouwd, geprogrammeerd en onderhouden door mensen' (p. 11). Om die reden is het voor leerlingen van belang om, in ieder geval op basaal niveau, te snappen hoe AI-(taal)modellen werken. En te weten hoe je de juiste vragen kunt stellen, of liever gezegd: de juiste prompts ingeeft.

In het boek *Chatten met Napoleon* (2023) geven blended-learningexpert Barend Last en docent Engels Thijmen Sprakel veel werkvormen en tips voor het werken met AI-modellen. Voor optimale resultaten bij *prompting* zijn volgens hen de volgende punten cruciaal:

- Geef een duidelijke taakomschrijving.
- Vertel voor wie de reactie op jouw *prompt* bedoeld is.
- Schets de stappen om de taak te voltooien.
- Geef context en beperkingen.
- Specificeer het doel.
- Geef instructies over de presentatie van de output, zoals lengte of stijl.

Naast deze tips is het belangrijkste advies van Last en Sprakel: experimenteer met *prompt engineering*, het liefst samen met je leerlingen.

## ‘Ik heb geleerd over het inzetten van computers bij Nederlands dat je erg specifiek moet opschrijven wat je wilt’

geletterdheid met elkaar verbonden zijn. In zijn lessen voor havo 3 laat hij leerlingen aan de slag gaan met het maken van gedichten met behulp van ChatGPT, nadat hij aandacht heeft besteed aan poëzie en in het bijzonder figuurlijk taalgebruik en metaforen. ChatGPT een gedicht laten schrijven over vriendschap met een metafoor, zonder het woord *vriendschap* te gebruiken, blijkt nog geen sinecure. Leerlingen leren te oefenen met *prompting* én ze lopen tegen de grenzen van ChatGPT aan. Want zo fraai zijn die Chatgedichten niet. De eerste evaluaties van leerlingen spreken boekdelen. De een ziet in ‘dat computers dommer zijn dan mensen’, de ander vindt ze juist slimmer. Weer een ander zegt: ‘Ik heb geleerd over het inzetten van computers bij Nederlands dat je erg specifiek moet opschrijven wat je wilt’, en: ‘dat je teksten geschreven door computers goed moet controleren op fouten’.

### Plausible and pleasing

Talen en de wens om die te leren blijven natuurlijk bestaan, ook als die taal in de toekomst steeds vaker kunstmatig zal worden gegenereerd. Een kritische taalblik is dan belangrijk. ‘Om te denken in prompts is taalbewustzijn een vereiste’, aldus Maurice Schols. Maar ook met goed geformuleerde prompts (zie kader), is taalwaakzaamheid geboden. Taalmodellen zijn namelijk niet geprogrammeerd om de waarheid te vertellen, maar om een plausibel antwoord te geven op vragen. Zuidema: ‘ChatGPT is *plausible and pleasing*. De waarheid is secundair.’ Bovendien is het is niet helder welke data er precies gebruikt zijn bij de training van ChatGPT; soms zitten er gekke gaten in de kennis. En aangezien ChatGPT een *pleaser* is, kan het model gaan ‘hallucineren’: complete nonsens genereren. Met de verbetering van de taalmodellen wordt het steeds moeilijker om afwijkingen te detecteren. Ook moeten leerlingen zich bewust zijn van de vooroordelen van ChatGPT, bijvoorbeeld bij gender, en de ongelijkheden die de ontwikkelingen mogelijk kunnen versterken. ChatGPT legt zelf uit: ‘Als taalmodel dat getraind is op grote hoeveelheden tekst, genereer ik geen bewuste of opzettelijke genderstereotypen. Mijn gegenereerde antwoorden weerspiegelen echter soms de vooroordelen en genderstereotypen die mogelijk aanwezig

zijn in de trainingsdata. Het is essentieel om bewust te zijn (sic) van deze mogelijke vooroordelen en ze kritisch te bekijken om inclusie en gelijkheid te bevorderen.’ Ongelijkheid geldt ook de talen zelf. Zo zijn de Verenigde Staten en China de landen waar op dit moment het meest wordt geïnvesteerd in kunstmatige intelligentie – in de VS voornamelijk door de bigtechbedrijven. De prestaties van taalmodellen zijn in het Engels dan ook veel beter dan in het Nederlands. Meer perifere talen komen er nog bekaaid van af. Verder moeten leerlingen snappen dat alle teksten die zij invoeren, opnieuw gebruikt zullen worden door ChatGPT.

## ‘ChatGPT is plausible and pleasing.’

### De waarheid is secundair’

In zijn presentatie toont Zuidema een krantenkop uit 1965: ‘Math teachers protest against calculator use’. Een dergelijke ‘ban’ heeft niet lang standgehouden. Veel jongeren maken dankbaar gebruik van ChatGPT. Ook meer docenten zullen gaan experimenteren met ChatGPT en het eerste boek met praktische werkvormen is inmiddels verschenen (zie kader). ‘Mensen zijn sociale dieren, machines niet’, schrijft Bennie Mols (2023, p. 38). Maar ook: ‘ons gevoel wat het betekent om mens te zijn gaat door AI veranderen’ (p. 104). Taal is een sociale praktijk; kunstmatige taal comfortabele calculatie. Ook onze menselijke taal zal meer en worden beïnvloed door kunstmatige taal, net zoals het Engels en taal op sociale media in ons dagelijks taalgebruik doorsijpelt. Taalbewustzijn wordt daarmee alleen maar belangrijker. Kennis van de werking van taalmodellen hoort daarbij. ■

#### LITERATUUR

- Dąbkowski, M., & Beguš, G. (2023). *Large language models and (non-) linguistic recursion*. arXiv. <https://doi.org/10.48550/arXiv.2306.07195>
- Last, B., & Sprakel, T. (2023). *Chatten met Napoleon: Werken met generatieve AI in het onderwijs*. Boom.
- Mols, B. (2023). *Slim, slimmer, slimst: Hoe kunstmatige intelligentie de mens een turboboost geeft*. New Scientist.
- Oostendorp, M. van. (2023a, 18 mei). ChatGPT maakt het vwo-examen Nederlands. *Neerlandistiek: Online Tijdschrift voor Taal- en Letterkunde*. <https://neerlandistiek.nl/2023/05/chatgpt-maakt-het-vwo-examen-nederlands/>
- Oostendorp, M. van. (2023b, 19 mei). Slaagt ChatGPT voor het havo-examen Nederlands? *Neerlandistiek: Online Tijdschrift voor Taal- en Letterkunde*. <https://neerlandistiek.nl/2023/05/slaagt-chatgpt-voor-het-havo-examen-nederlands/>

## nieuws

### Het Engels van universitaire docenten

Nederlandse docenten die Engelstalig universitair onderwijs geven, maken allerlei taalkundige fouten in hun lessen. Dat blijkt uit het proefschrift van Klazien Tilstra. Zij onderzocht welke competenties docenten nodig hebben om goed les te geven in het Engels en waar het vaak fout gaat. Het veelvuldig maken van fouten schaadt niet alleen de autoriteit van de docent, het leidt ook makkelijk tot misverstanden en ondermijnt de begrijpelijkheid van de les.

Gedurende haar promotieonderzoek (zie <[vlt.nl/tilstra](http://vlt.nl/tilstra)>) onderzocht Tilstra met welke taalkundige en communicatieve uitdagingen Nederlandstalige docenten aan een universiteit te maken krijgen als ze moeten lesgeven in het Engels. Daartoe ontwikkelde ze een model waarmee ze dertig Engelstalige lessen van dertig Nederlandstalige docenten analyseerde. Hieruit blijkt welke fouten het vaakst worden gemaakt door Nederlandstalige docenten die moeten lesgeven in het Engels.

Nederlandstalige docenten maken allerlei soorten fouten in hun Engelstalig onderwijs, wijst de analyse van de dertig bestudeerde lessen uit. ‘De algemene kwantitatieve analyse toont dat zich in alle lessen problemen voordeden met vloeiend taalgebruik, lexico-grammaticale nauwkeurigheid, uitspraak en helderheid’, schrijft Tilstra.

De kwantitatieve analyse stelde haar in staat om een top tien van de meest voorkomende fouten te maken, aangevuld met een inschatting van de kans om tot misverstanden te leiden. De meest voorkomende fout betreft de uitspraak, meer precies de uitspraak van de *ash* (de *æ*, de wat langer aangehouden klinker in bijvoorbeeld het Engelse *trap*), het al dan niet stemhebbend uitspreken van medeklinkers aan het eind van een woord (*word-final fortis-lenis contrast*) en aspiratie (het uitspreken van een medeklinker met wat meer adem). Hoewel fouten in uitspraak niet meteen tot misverstanden leiden, komen de genoemde varianten wel heel vaak voor.

De fouten die het snelst tot misverstanden leiden, zijn het gebruik van incomplete zinnen, onduidelijk woordgebruik, onduidelijke verwijzingen en onduidelijke logische proposities. Het veelvuldig maken van fouten schaadt niet alleen de autoriteit van de docent, het komt ook de begrijpelijkheid van de les niet ten goede. Vooral fouten in taalgebruik en een opeenstapeling van fouten kunnen een negatief effect hebben op het begrip bij studenten, aldus Tilstra. *ScienceGuide*

### Smartphones en welbevinden

Gebruik van smartphones in de klas en op school lijkt niet bij te dragen aan het welbevinden van leerlingen in het voortgezet onderwijs, zo blijkt uit een recente literatuurstudie van Kennisrotonde. Er zijn aanwijzingen dat leerlingen minder direct contact met elkaar hebben, waardoor ze bepaalde communicatievaardigheden minder goed ontwikkelen.

Het bezit en gebruik van smartphones onder jongeren is zeer groot. In 2018 gebruikte 99 procent van hen internet op een mobiele telefoon. Een op de drie leerlingen in het voortgezet onderwijs heeft bijna de hele dag onlinecontact met anderen. Vijf procent vertoont kenmerken van verslaving.

Een eerder literatuuroverzicht van Kennisrotonde (zie <[vlt.nl/phone-leren](http://vlt.nl/phone-leren)>) liet zien dat vrij gebruik van smartphones, tablets en laptops tijdens de lessen een negatief effect heeft op de leerprestaties van leerlingen en studenten. Leerlingen worden te veel afgeleid en kunnen zich niet concentreren op de les of de leertaak.

Er is weinig onderzoek gedaan naar (een verbod op) smartphonegebruik in de les en het welbevinden van leerlingen in het voortgezet onderwijs. Daarom is ook gekeken naar studies in het basisonderwijs en het hoger onderwijs. Uit een Tsjechische studie in het basisonderwijs blijkt dat leerlingen tussen de acht en twaalf jaar op scholen zonder smartphoneverbod meer onderlinge conflicten hebben. Ze zijn minder tevreden over de sfeer in de klas en ervaren meer onderlinge competitie dan leerlingen op scholen met een smartphoneverbod. Dat gaat dan ten koste van hun welbevinden. Volgens de onderzoekers leidt smartphonegebruik tot minder directe communicatie tussen leerlingen, waardoor ze vaardigheden als empathie, assertiviteit en het met elkaar eens willen worden minder goed ontwikkelen. Een Amerikaans onderzoek onder studenten sociale wetenschappen laat zien dat studenten die de beschikking hebben over hun smartphone tijdens het college meer stress ervaren en hun aandacht minder goed bij de les kunnen houden.

Deze bevindingen uit het primair en hoger onderwijs zijn terug te vinden in een studie in twee Spaanse regio's waar het gebruik van smartphones op middelbare scholen is verboden. Na invoering van het smartphoneverbod is er een daling te zien in het aantal leerlingen die een melding van pesten doen ten opzichte van regio's zonder een dergelijk verbod. Het gaat om een daling tussen de tien en twintig procent.

Het volledige rapport is te raadplegen op <[vlt.nl/phone-welbevinden](http://vlt.nl/phone-welbevinden)>. Kennisrotonde