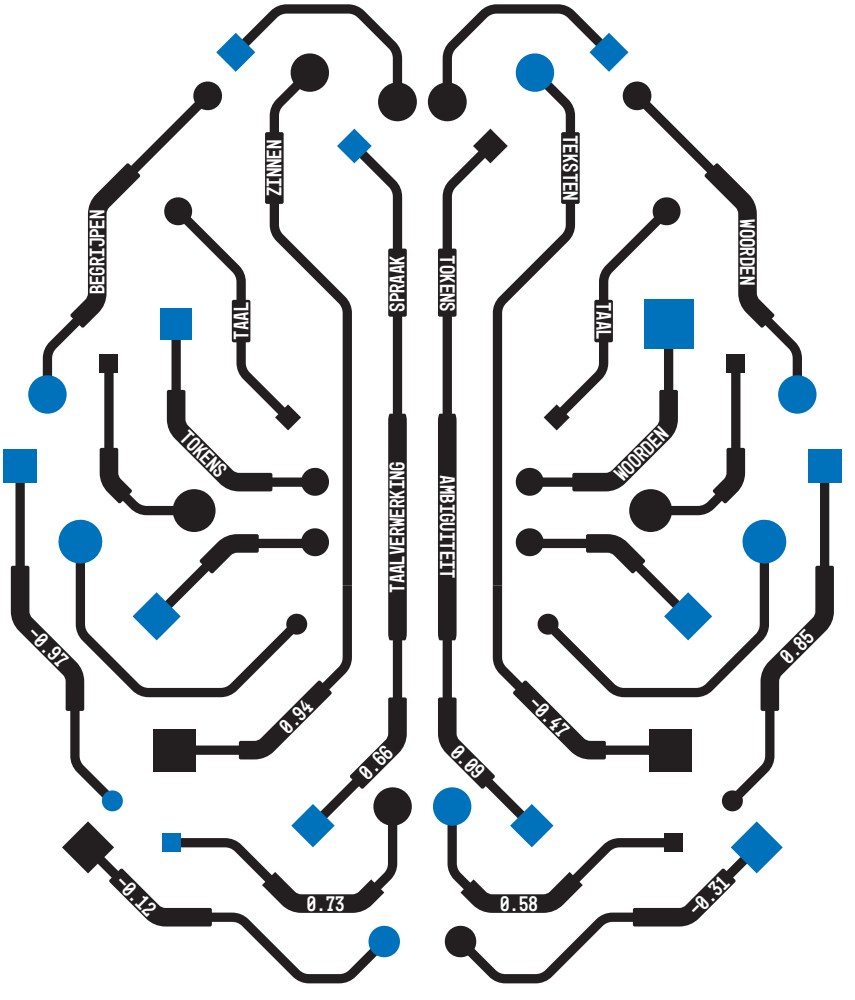
AI & TAAL

WALTER DAELEMANS



TAALTECHNOLOGIE NA CHATGPT

ACADEMIA PRESS

Academia Press
Coupure Rechts 88
9000 Gent
België

www.academiapress.be

ISBN 9789059964914
D/2026/45/514
NUR 616, 984

AI en taal
Walter Daelemans
Gent, Academia Press, 2026, 208 p.

Eerste druk, september 2026
Vormgeving cover: Studio Lannoo – Robin De Boelpaep
Vormgeving binnenwerk: Atelier Steve Reynders
© Walter Daelemans & Uitgeverij Lannoo nv, Tielt

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand en/of openbaar gemaakt in enige vorm of op enige wijze, hetzij elektronisch, mechanisch of op enige andere manier zonder voorafgaande schriftelijke toestemming van de rechtenhouder. Tekst- en datamining van (delen van) deze uitgave zijn uitdrukkelijk niet toegestaan.

All rights are reserved, including those for text and data mining, AI training and similar technologies.

Een inleiding tot taaltechnologie **9**
Voorwoord **11**

DEEL 1: DE BELOFTE VAN TAALTECHNOLOGIE 15

HOOFDSTUK 1. TAALTECHNOLOGIE NA CHATGPT **16**
Taal en artificiële intelligentie 17
Multimodale taaltechnologie 21

HOOFDSTUK 2. ARTIFICIËLE INTELLIGENTIE EN TAAL **40**
De ontwikkeling van AI 41
De rol van rekenkracht en schaal 44
Theoretische en filosofische fundamenten van AI 49
Ethische en maatschappelijke implicaties van AI 54

DEEL 2: HOE MACHINES TAAL VERWERKEN 61

HOOFDSTUK 3. HOE MACHINES TAAL BEGRIJPEN **62**
De architectuur van taalverwerking 63
Ambiguïteit oplossen met context 69
Data en machinelearning 74

HOOFDSTUK 4. GROTE TAALMODELLEN **102**
De uitdagingen van neurale taalmodellen 103
De betekenis van woorden in vectoren 104
De kracht van generatieve taalmodellen 114
AI als alienintelligentie 131

DEEL 3: DE TOEPASSING EN IMPACT VAN TAALTECHNOLOGIE 139

HOOFDSTUK 5. TOEPASSINGEN VAN TAALTECHNOLOGIE **140**
Automatisch vertalen 141
Zoekmachines 145
Text mining: van tekst naar kennis 147
Dialoogsystemen en chatbots 154
Evaluatie van taaltechnologische toepassingen 156

HOOFDSTUK 6. CREATIVITEIT EN STIJL IN HET TIJDPERK VAN AI 164
De schrijvende zombie 166
Sporen van de auteur in de tekst..... 172
Het profiel van de auteur 176

WAT WE NIET MOGEN UITBESTEDEN 186
TREFWOORDENLIJST.....190
REFERENTIES.....204

Een inleiding tot taaltechnologie

Taaltechnologie is een tak van de informatica die computers en robots taalvaardig wil maken. Dat is een moeilijk probleem, omdat taal ambigu is, afhankelijk van context en verweven met kennis van de wereld. We schetsen in dit boek de methodes, mogelijkheden, problemen en toepassingen van dit vak.

Taaltechnologie was lange tijd een kleine niche tussen taalkunde en informatica. Niemand had zien aankomen dat statistische taalmodellen, een van de vele methodes uit dat vakgebied, zouden uitgroeien tot de kern van een technologische revolutie met wereldwijde impact. Taalmodellen zijn bijna een synoniem geworden voor artificiële intelligentie (AI). Vandaag wordt er miljarden in geïnvesteerd, is er een geopolitieke wedloop rond gestart en is het aantal AI-onderzoekers en -bedrijven sinds de jaren negentig explosief gegroeid, vooral in de laatste vijf jaar.

De technologie roept veel, vaak spectaculaire, vragen op, maar wordt niet altijd goed begrepen. Gaat deze variant van AI de mensheid uitroeien of is ze niet meer dan een hype en een economische zeepbel? Staan we aan de vooravond van ongeziene AI-gestuurde vooruitgang in geneeskunde, onderwijs en wetenschap, of komt er massale werkloosheid op ons af? Zullen we er gemakzuchtiger en afhankelijker door worden, of helpt AI ons net om onze talenten ten volle te ontwikkelen? En zijn deze systemen echt intelligent en creatief, of wekken ze alleen maar die illusie?

Dit boek is er voor wie zich over deze vragen een goed onderbouwde mening wil vormen. Door de keuze van de voorbeelden en de aandacht voor de ideeëngeschiedenis en filosofische context mikt het vooral op lezers met een interesse in de menswetenschappen.

Grote taalmodellen zoals ChatGPT zijn ontstaan uit een oud idee, maar verleggen grenzen. We begrijpen heel goed hoe ze werken, maar dat helpt niet om hun gedrag te verklaren of te voorspellen. Vaak zijn ze indrukwekkend goed, soms ontmoedigend slecht. Waar die paradox vandaan komt en hoe ermee om te gaan, is een van de onderwerpen in dit boek. De nadruk ligt daarbij op intuïtief begrijpen. Voor wie wat dieper wil gaan, zijn er de meer technische blauwe kaders en de suggesties aan het slot van elk hoofdstuk en op de bijbehorende website. Op die website volgen we ook de recentste ontwikkelingen op.

Hartelijk dank aan Antal van den Bosch, Ehsan Lotfi, Eleni Daelemans, Jens Lemmens, Jens Van Nooten, Lisa Hilte, Luna De Bruyne, Merel Naudts, Mike

Kestemont, Nicolas Daelemans en Pieter Fivez voor commentaar op eerdere versies van de tekst. Dank aan Peter Spyns voor een interessant gesprek over taaltechnologie in Vlaanderen. En dank aan de twintig generaties studenten in de bachelor Taal- en Letterkunde van de Universiteit Antwerpen die het vak taaltechnologie (computerlinguïstiek) hebben doorstaan, vaak met verbazing en gezonde scepsis.

Mijn grootste dank is voor Sonja, voor haar geduld en steun tijdens het schrijven van dit boek.

Bij dit boek hoort een website

Scan de QR-code of surf naar https://ai-en-taal.uantwerpen.be/ voor aanvullend materiaal, verdere leestips en updates over de nieuwste ontwikkelingen in AI en taal.

Voorwoord

Thomas Winters
Gastprofessor AI aan de KU Leuven

Als het boek dat je hier in handen hebt al had bestaan aan het begin van mijn onderzoekscarrière, zou ik het direct van voor naar achter en opnieuw hebben gelezen, en dan talloze keren herlezen in de jaren die volgden.

Meer dan tien jaar geleden, als beginnend onderzoeker in creatieve AI en taalmodellen, probeerde ik computers mopjes te laten schrijven met enkele statistische taalkundige trucjes en AI-technieken. Wanneer ik de resultaten trots aan collega's toonde, werd er soms gelachen. Al was het eerder om de machine dan om wat ze produceerde. Het was een curiositeit. Na de demo ging iedereen weer aan het echte werk: analytische AI-modellen op getallen ontwerpen. Niemand wist toen dat generatieve taalmodellen weldra de toekomst zouden worden.

Een hele generatie taaltechnologen zal die gematigde reacties, met een vleugje medelijden, wel herkennen. Systemen bouwen die zelf taal voortbrengen was lang een nichebezigheid: weinig mensen, weinig middelen en veel scepsis van buitenaf. Ik herinner me nog goed hoe ik in België amper mensen vond die ermee bezig waren en me wegwijs konden maken.

Tot ik een lezing van Walter Daelemans bijwoonde. Vanaf de eerste rij luisterde ik gefascineerd hoe hij computerlinguïstiek uitlegde: helder, geduldig en vooral met diepe, bevattelijke inzichten in het hele veld. Toen we nadien napraatten, kon hij meteen verwijzen naar papers en systemen met net dat soort taalkundige trucjes die ik nodig had voor de humor die ik wou genereren.

Wat ik altijd apprecieer in de gesprekken die ik sindsdien met hem mocht voeren: ook al is hij in Vlaanderen een van de pioniers van taaltechnologie, je hoort hem nooit klagen dat vroeger alles beter was. Integendeel. Hij gebruikt al die ervaring om zich ook van deze golf volledig meester te maken en ze dan toegankelijk uit te leggen, zonder de inhoud plat te slaan. Hij gaat het ongemakkelijke feit niet uit de weg dat zijn vakgebied plots voortgestuwd wordt door het opschalen van data en rekenkracht in plaats van door taalkundige inzichten. En hij steekt nog steeds zelf de handen uit de mouwen om met de nieuwste technieken taaltechnologiesystemen te bouwen. In de academische wereld is deze combinatie van eigenschappen zeldzamer dan je zou denken.

Daarin zit meteen ook de meerwaarde van dit boek. Veel Nederlandstalige AI-boeken bekijken het onderwerp in vogelvlucht, belichten vooral de maatschappelijke impact of behandelen toepassingen in een specifiek domein. Wil je echt begrijpen hoe het vanbinnen werkt, dan beland je al snel bij boeken voor computerwetenschappers, vol programmeercode en wiskunde. Tussen die twee uitersten gaapt een gat, en dit boek vult het: het opent de motorkap van taalmodellen en toont net genoeg wiskunde zonder de lezer in formules te verzuipen. Daelemans neemt je mee langs alles wat je nodig hebt om taalmodellen in hun volle context te snappen. Hij kadert ze in de geschiedenis van de AI en het brede taaltechnologieveld, en legt uit welke sterktes en zwaktes de belangrijkste taaltechnologieën hebben. Zo kan je zelf, met kennis van zaken, inschatten hoe je er als mens het beste mee omgaat.

Lees dit boek dus niet om bang te worden of om je te laten meeslepen in de AI-hype, maar om grip op taaltechnologie te krijgen. Zie het als een kompas waarmee je je eigen koers kan bepalen doorheen het hedendaagse taaltechnologielandschap, ook wanneer dat terrein straks weer verandert. Walter Daelemans reikt je in dit boek de inzichten aan om ook de komende generaties taalmodellen te begrijpen en ze verstandig in te zetten in je leven en werk, en misschien zelfs om er mopjes mee te schrijven.

DEEL 1:

DE BELOFTE VAN TAALTECHNOLOGIE

HOOFDSTUK 1
TAALTECHNOLOGIE NA CHATGPT

People are finding that, because of the way the machines are changing the world, more and more of their old values don't apply any more. People have no choice but to become second-rate machines themselves, or wards of the machines.

— Kurt Vonnegut, Player Piano (1952)

Taal en artificiële intelligentie

Taaltechnologie is een subdiscipline van artificiële intelligentie. In deze subdiscipline worden algoritmen ontwikkeld om taal in al haar complexiteit automatisch te verwerken met computers. Het doel is om kennis, uitgedrukt in taal, toegankelijk te maken. Daarnaast willen we oplossingen bouwen die toestaan om in mensentaal te communiceren met computers en informatiesystemen. Bovendien vallen onder taaltechnologie ook automatische vertaalsystemen die het mogelijk maken te communiceren met mensen die een andere taal spreken. Die systemen moeten in staat zijn om taal te begrijpen, te produceren en te vertalen. Taaltechnologie moet ook de brug kunnen slaan tussen de verschillende verschijningsvormen van taalgebruik, zoals tekst, spraak en gebarentaal, en die zelfs kunnen omzetten naar beeld, muziek en andere modaliteiten.

Taal is bijzonder. Taal, gesproken of geschreven, laat toe om met een beperkte hoeveelheid woorden een oneindige hoeveelheid ideeën, gevoelens en betekenissen uit te drukken. Het is dan ook het belangrijkste medium waarmee we als mensen kennis extern opslaan en uitwisselen, en de eigenschap die bij mensen opvallend veel sterker ontwikkeld is dan bij andere dieren. Taal is nauw verbonden met intelligentie, denken en bewustzijn. We gebruiken taal om verhalen te vertellen, inzichten te verwoorden, gebeurtenissen te beschrijven en anderen te beïnvloeden en te overtuigen. Soms is wat we zeggen en schrijven objectief en feitelijk, soms zijn het subjectieve opinies en gevoelens. We gebruiken ironie en humor. Maar bovenal wíllen we communiceren in taal; we hebben iets te zeggen. Taalgebruik is actie en heeft impact. Bij dat communiceren is context belangrijk; we rekenen op gedeelde kennis bij onze lezers en gesprekspartners. Het is door die gedeelde kennis dat we elkaar begrijpen. Want wat we zeggen is, als je het letterlijk bekijkt, vaak vaag, onvolledig en ambigu. Het is door de verwevenheid met denken en voelen, maar vooral door die gedeelde culturele en contextuele kennis dat het decennialang zo moeilijk is geweest om computers taal te laten leren.

Succesvolle taaltechnologie heeft veel maatschappelijke en economische waarde. Fortune Business Insights voorspelt een groei van de sector naar 200 miljard dollar tegen het begin van de jaren 2030, vanuit 10 miljard begin jaren 2020. Daarnaast dient taaltechnologie ook een wetenschappelijk belang: de bouw van een werkend taalvaardig computersysteem kan meer inzicht geven in hoe mensen taal leren en hoe we taalkennis opslaan en gebruiken in ons brein.

Taalvaardigheid automatiseren is geen gemakkelijke taak, en de vooruitgang sinds het prille begin in de jaren 1950 is moeizaam geweest, met verschillende vermeende doorbraken en teleurstellingen. En toen gebeurde er iets bijzonders: de introductie van ChatGPT door het Amerikaanse bedrijf OpenAI. Plots was er een voor iedereen toegankelijke chatbot die toeliet om met een computersysteem te communiceren alsof het een mens was. Was het al die tijd dan toch gemakkelijker dan we dachten?

Laten we maar meteen met de deur in huis vallen, met een stukje ChatGPT[1] als voorbeeld. Wat volgt is een conversatie met ChatGPT over een literair fragment.

> Analyze the following quote. What can you tell about the style of the author? Use no more than 200 words.
>
> ‘Je moet weten dat mijn moeder gestorven is. Een nare geschiedenis natuurlijk, niet alleen voor haar maar ook voor mijn zusters, die er zich bijna dood aan gewaakt hebben. Zij was oud, zeer oud. Op een paar jaar na weet ik niet hoe oud zij precies was. Ziek was zij eigenlijk niet, maar grondig versleten. Mijn oudste zuster, waar ze bij inwoonde, was goed voor haar. Zij weekte haar brood, zorgde voor stoelgang en gaf haar aardappelen te schillen om ze bezig te houden. Zij schilde, schilde, als voor een leger. Wij brachten allemaal onze aardappelen bij mijn zuster en dan kreeg zij die van madame van boven en van een paar buren ook nog, want toen ze eens geprobeerd hadden haar een emmer reeds geschilde aardappelen nog eens te doen overschillen, wegens gebrek aan voorraad, toen had zij ’t gemerkt en warempel gezegd ‘die zijn al geschild.’’
>
> –
>
> The author’s style in this quote is straightforward, colloquial, and unembellished. They use simple, direct language to describe a personal and somewhat grim situation—the death of their mother. The tone is conversational, with a focus on everyday details and the practical aspects

of caregiving. The use of short, clear sentences and the inclusion of specific, mundane tasks (like peeling potatoes) adds to the realism and immediacy of the narrative.

The author's style also includes subtle humor and irony. The anecdote about the mother recognizing already peeled potatoes, despite her age and condition, introduces a touch of dry wit. This balances the otherwise somber topic, giving the narrative a more human and relatable quality. Overall, the style is unpretentious and grounded in the reality of ordinary life.

—

Continue the story in the same style. No more than 5 sentences.

—

Mijn jongste zuster, die een eindje verderop woonde, kwam vaak langs met haar kinderen. Zij brachten altijd een beetje leven in huis, maar soms ook een hoop rommel. De kinderen noemden oma liefkozend 'oude schilster' en probeerden haar aan het lachen te maken met hun streken. Maar zelfs dan bleef ze onverstoorbaar schillen, alsof ze een belangrijke taak had die niemand anders kon volbrengen. Uiteindelijk, op een rustige zomermiddag, is ze in haar stoel ingeslapen en niet meer wakker geworden.

Eind 2022 was de wereld met stomheid geslagen bij de introductie van ChatGPT: een taalmodel met een gebruiksvriendelijke interface en een uitgesproken taalvaardigheid. Opeens kon iedereen zomaar communiceren met een 'intelligente' chatbot die eerdere pogingen ver achter zich liet. Ook de experts waren verrast, want bijna niemand had verwacht dat de strategie achter ChatGPT, het simpelweg opschalen van een al lang bekende methode, tot zo'n indrukwekkend resultaat zou leiden. Zonder overdrijving kun je stellen dat er een periode voor en na ChatGPT is in de taaltechnologie. In latere hoofdstukken komen we in detail terug op hoe neurale netwerken, en meer specifiek *transformers*, het algoritme waarop ChatGPT gebaseerd is, precies werken.

Maar waarom was iedereen zo onder de indruk? De korte interactie in het voorbeeld laat zowat alles zien wat we verlangen van een taalverwerkend systeem. Het moet eerst en vooral taal kunnen 'begrijpen'.[2] In dit geval niet alleen de vraag van de gebruiker (die we bij taalmodellen de prompt noemen), maar ook het tekstfragment waar de vraag over gaat, een tekst die bovendien geschreven

is in een wat oudere variant van het Nederlands. Het is een fragment uit de roman *Kaas* van Willem Elsschot (1933), maar dat weet het model niet. Er komt ook wat redeneren bij kijken. Het lijkt wel alsof het model echt kennis heeft over literaire stijlen. De genoemde, en correcte, stijlkenmerken (eenvoud, realisme en droge humor) worden geïllustreerd met voorbeelden uit de tekst. ChatGPT produceert de antwoorden met gemak in grammaticaal correct Engels, zelfs met gebruik van genderneutrale voornaamwoorden (het kent de auteur niet). Het gevraagde vervolgfragment, in het Nederlands, is narratief acceptabel, hoewel de stijl van Elsschot natuurlijk nooit geëvenaard wordt. Een verdienstelijke poging. Het model is dus meertalig en begrijpt en produceert tekst in het Nederlands of Engels zonder hapering. Ten slotte kan het systeem ook een conversatie volhouden en een consistente dialoog opbouwen. Is taaltechnologie dan een opgelost probleem? We zullen zien dat het antwoord op die vraag niet eenvoudig is.

Taaltechnologie is al langer alomtegenwoordig als ingebedde technologie. We gebruiken, vaak zonder het te beseffen, automatische taalverwerking in zoekmachines als Google, in vertaalsystemen als Google Translate en DeepL Translate, in conversatie met smartphones en informatiesystemen (zoals Siri en Alexa), bij automatische spelling- en grammaticacontrole, bij het automatisch modereren van wat er op sociale media wordt gezegd, bij het automatisch ondertitelen en vertalen van YouTubefilmpjes en in nog heel wat meer toepassingen.

Vanaf ongeveer 2015 is de kwaliteit van al deze systemen, waaraan al decennialang gewerkt werd, sterk toegenomen en met ChatGPT zelfs nog spectaculairder verbeterd. ChatGPT was het eerste systeem van zijn soort waarmee iedereen vrij kon experimenteren op een gemakkelijke manier, zonder technische voorkennis. Dat had een ongeziene impact. Met een miljoen actieve gebruikers binnen vijf dagen en 100 miljoen na drie maanden was het de snelst groeiende consumententoepassing ooit. ChatGPT is een chatbot die gedreven wordt door een ‘groot taalmodel’. Andere voorbeelden zijn onder meer Gemini, Claude, DeepSeek, Mistral en Llama. In dit boek zullen we regelmatig de Engelse afkorting LLM (*large language model*) gebruiken als parapluterm voor die systemen.

Terminologie

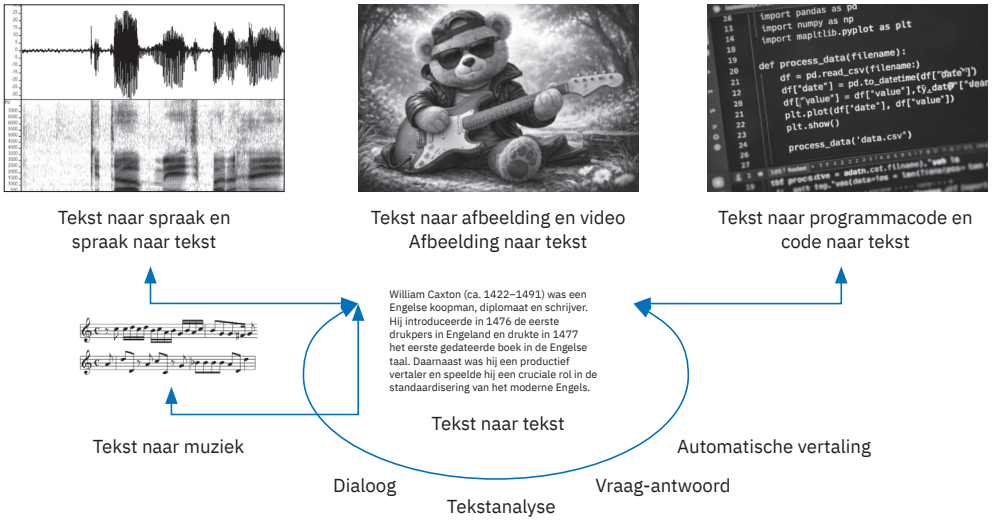
We gebruiken in dit boek de term **taaltechnologie** omdat die vandaag in het Nederlandse taalgebied de dominante benaming is voor onderzoek naar au-

tomatische taalverwerking met behulp van computermodellen. **Computertaalkunde** of **computerlinguïstiek** is een andere veelgebruikte benaming, die teruggaat naar de periode waarin taalkunde een grotere rol speelde in taaltechnologie. Nog een andere benaming is **natuurlijke taalverwerking** (letterlijke vertaling van *natural language processing*, NLP), een term die vaak gebruikt wordt wanneer taaltechnologie als onderdeel van artificiële intelligentie wordt beschouwd, zoals we in dit boek doen. Voor automatisering van gesproken taal wordt ook de term **spraaktechnologie** gebruikt. En voor het vakgebied van onderzoekers die vooral geïnteresseerd zijn in het simuleren, met behulp van computermodellen, van hoe mensen hun moedertaal verwerven, vreemde talen leren en taal begrijpen en produceren, gebruiken we de term **computationele psycholinguïstiek**.

In dit boek kijken we naar de oorzaken van deze plotse en snelle verbetering in de taaltechnologie en de impact ervan, maar we hebben ook aandacht voor de beperkingen, de ethische vraagstukken en wat we in de toekomst nog kunnen verwachten. We kunnen taaltechnologie ook niet los zien van de ontwikkelingen in de artificiële intelligentie, waar ze deel van uitmaakt. Tegenwoordig lijkt AI bijna een synoniem voor automatische taalverwerking. Nochtans zijn *large language models* een verschijningsvorm van maar één specifieke soort AI: generatieve artificiële intelligentie (GenAI). Deze modellen zijn gebaseerd op neurale netwerken die nieuwe beelden, geluiden, spraak, tekst en zelfs muziek kunnen voortbrengen (genereren) op basis van een instructie in tekstvorm (de prompt).

Multimodale taaltechnologie

GenAI heeft de taaltechnologie multimodaal gemaakt. Het voegt naast het begrijpen en produceren van taal, vertalen en dialoog tal van nieuwe mogelijkheden toe door de combinatie van tekst met spraak, beeld en muziek. Maar taal blijft daarbij centraal; de aansturing gebeurt in taal of het resultaat wordt weergegeven in taal.



Figuur 1. Multimodale taaltechnologie.

Laten we die groeiende ruimte van multimodale taaltechnologietoepassingen kort overlopen: van tekst naar spraak en van spraak naar tekst, van tekst naar beeld en van beeld naar tekst, van tekst naar programmeercode en van code naar de uitleg ervan, en van tekst naar muziek en van muziek naar de beschrijving ervan (*music captioning*). In de rest van het boek concentreren we ons vooral op de verwerking van tekst naar tekst (dialoog, vertaling, extractie van kennis uit tekst en andere taaltoepassingen).

Tekst en spraak

Tekst en spraak zijn twee verschillende verschijningsvormen van dezelfde talige informatie. Spraaktechnologie bestaat uit het automatisch omzetten van tekst naar spraak (spraaksynthese) en het automatisch omzetten van spraak naar tekst (spraakherkenning of transcriptie van spraak). Het gaat hierbij meestal niet over de betekenis van wat gezegd wordt, maar puur over het omzetten van spraak naar tekst en omgekeerd. We kunnen spraaktechnologie dus beschouwen als een soort voor- of naverwerking van tekst. Er bestaan tegenwoordig ook systemen, bijvoorbeeld voor vertaling, die spraakinvoer rechtstreeks omzetten naar spraakuitvoer zonder tekst als tussenstap.

Spraaksynthese, het omzetten van tekst naar spraak, is een goed opgelost probleem voor de meeste talen, hoewel nog altijd ontwikkeling nodig is voor min-

der onderzochte talen. De kwaliteit is geëvolueerd van robotachtige uitspraak en intonatie naar natuurlijk klinkende stemmen in chatbots, voorleessoftware en toepassingen voor slechtzienden. Zelfs het uitdrukken van emoties in artificiële spraak is mogelijk en klinkt steeds natuurlijker. Prosodie (zinsmelodie) blijft nog een probleem in sommige toepassingen. De vooruitgang is de laatste jaren zo snel gegaan dat geavanceerde stemvervalsingen (deepfakes en voicecloning) een reëel gevaar vormen. Het is erg gemakkelijk geworden om woorden in iemands mond te leggen. Zeker in combinatie met beeldvervalsing is dit een gevaarlijke evolutie. Ze maakt bijvoorbeeld erg overtuigende vormen van phishing en fake news mogelijk.

Spraakherkenning, de transcriptie van spraak naar tekst, was in het verleden vooral succesvol bij dicteertoepassingen en stemgestuurde assistenten als Siri (Apple). In die toepassingen kan de software getraind worden om de spraak van een specifieke gebruiker te herkennen. Bij digitale assistenten heeft het systeem bovendien veel informatie over die gebruiker en relatief weinig verschillende commando's die herkend moeten worden. Recent is ook de accuraatheid van spontane spraakherkenning, zonder training op individuele stemmen, fenomenaal gestegen. Maar het blijft vaak een probleem als verschillende mensen tegelijk praten of als er veel achtergrondlawaai is. Toepassingen van spraakherkenning zijn onder meer handenvrije besturing van apparaten (zelfs auto's), automatische ondertiteling van spraak, spraakvertaling, conversaties met chatbots en het zoeken naar specifieke spraakfragmenten in audio of video.

Zowel voor spraaksynthese als voor spraakherkenning blijft de kwaliteit achter voor talen of dialecten waarvoor minder trainingsdata beschikbaar zijn: talen met beperkte digitale ondersteuning (low-resource-talen). Voor het Nederlands (en vooral de dialecten in Vlaanderen) is er zeker nog behoefte aan een inspanning om spraak goed te kunnen omzetten naar tekst. Dit is een illustratie van een groter probleem in taaltechnologie: de systemen worden ontwikkeld voor de gemiddelde gebruiker, in het nadeel van wie afwijkt door leeftijd, regionaal accent of moedertaal.

Tekst en beeld

GenAI-technieken hebben ook toepassingen mogelijk gemaakt waarin enerzijds beelden beschreven worden in tekstvorm en anderzijds met behulp van een beschrijving in tekstvorm een beeld (of zelfs video) geproduceerd kan worden. Dergelijke toepassingen waren tot voor kort ondenkbaar. Ze worden nu dagelijks gebruikt.

De beschrijving van beelden begon met toepassingen die in staat waren beelden te classificeren (‘dit is een hond’, ‘dit is een smartphone’). Het is onder meer het ongeziene succes van diepe neurale netwerken in 2012 bij deze taak dat een renaissance van deze aanpak inluidde. Een uitbreiding ervan, waarbij verschillende delen van een afbeelding gecategoriseerd worden in plaats van het beeld als geheel, wordt beeldannotatie genoemd en wordt onder meer gebruikt om databanken met afbeeldingen te indexeren. Dankzij de beschrijvingen kunnen de beelden later op basis van hun inhoud teruggevonden worden.

Beschrijf wat je ziet op de foto.

–

Op de foto zie je vier giraffen die samen in een ruim, omheind buitenverblijf staan. Ze bevinden zich op een zandpad, omringd door een groene, bosrijke omgeving met hoge dennenbomen en struiken. De giraffen staan dicht bij elkaar, drie meer naar links en één iets verder naar rechts. Op de voorgrond is een metalen afrastering zichtbaar, met wat gras, mos en enkele stenen. Het is een zonnige dag met een heldere blauwe lucht, wat de warme, natuurlijke sfeer van de omgeving versterkt.

–

Figuur 2. Een interactie met ChatGPT die beeldannotatie illustreert.

Afbeeldingen categoriseren (labelen) was trouwens maar een eerste stap. Vanaf 2015 waren er de eerste resultaten van het automatisch beschrijven van beelden in tekst (*image captions*), die uiteindelijk culmineerden in uitgebreide beschrijvingen, zoals hier geïllustreerd in figuur 2.

De tekst wordt automatisch gegenereerd op basis van de analyse van het beeld. Dergelijke systemen zijn een combinatie van de eerdergenoemde beeldannotatie en tekstgeneratie.

Nog indrukwekkender is de omzetting van tekstbeschrijvingen naar beelden, zoals we dat gewend zijn geraakt in systemen als DALL-E en Midjourney en tegenwoordig in alle commerciële systemen. In figuur 3 vinden we de output van DALL-E 3 op de prompt 'Een pentekening van een eenzame robot in de stijl van Spilliaert op het strand bij ondergaande zon.' Interessant is dat het systeem de prompt van de gebruiker zelf omzet naar een betere prompt. Vertaling naar het Engels ligt voor de hand omdat de trainingsdata Engelstalig zijn, maar de prompt wordt ook geparafraseerd naar een meer specifieke, beeldende taal en lost mogelijk een copyrightprobleem op door de naam Spilliaert te vervangen door *early 20th century symbolism*. Daarmee toont het systeem trouwens dat het Spilliaert en zijn stijl kent. De uiteindelijke, door het systeem gebruikte prompt wordt dan: *Black and white pen drawing of a tall, slender lonely robot standing on a sandy beach with the waves gently crashing nearby. The sky, while devoid of color, shows contrasts of light and dark as if the sun is setting. The scene has moody, atmospheric, and shadowy elements reminiscent of early 20th century Symbolism.* Intussen is die expliciete tussenstap niet meer zichtbaar bij ChatGPT van OpenAI. De opvolger van DALL-E (GPT-Images) zit nu geïncorporeerd in de chatinterface.

Figuur 3. Uitvoer van DALL-E 3 (GPT-Images) op de prompt 'Een pentekening van een eenzame robot in de stijl van Spilliaert op het strand bij ondergaande zon'.

Deze modellen zijn gebaseerd op grote hoeveelheden referentiedata, in dit geval afbeeldingen met bijbehorende tekstbeschrijvingen. Voorlopig lijkt er nog geen einde te komen aan de kwaliteitsverbetering van deze multimodale toepassingen. Zelfs het genereren van video op basis van tekst ligt binnen bereik, bijvoorbeeld met Gen-4 van Runway of Veo van Google.

Benchmarkdatasets

In generatieve AI en taaltechnologie komt vooruitgang vaak voort uit goed gekozen verzamelingen data die als referentie of ijkingsdata voor de computermodellen dienen. Die zogenaamde benchmarkdatasets zorgen ervoor dat iedereen met hetzelfde materiaal kan werken en modellen objectief kan vergelijken dankzij gestandaardiseerde evaluaties. Bij beeldclassificatie zorgde ImageNet voor de grote doorbraak van diepe neurale netwerken: een databank die in 2009 werd gelanceerd en miljoenen beelden bevatte, onderverdeeld in duizenden categorieën (hond, kat, stoel).

Ook in de taaltechnologie zien we hetzelfde patroon. Al sinds de jaren negentig hebben datasets als de Penn Treebank, de CoNLL-datasets of recenter de GLUE- en SuperGLUE-benchmarks de vooruitgang in die discipline enorm versneld. Dankzij dit soort datasets is het mogelijk om precies te meten welke aanpak beter presteert. Onderzoekers kunnen sneller ontdekken wat werkt en wat niet, en delen hun inzichten tijdens conferenties rond deze zogenaamde *shared tasks*: taken waarvoor verschillende teams een oplossing ontwikkelen die geëvalueerd wordt op de benchmarkdataset.

Een nadeel is dat benchmarks de ontwikkeling van de modellen zo sterk beginnen te beïnvloeden dat ze uiteindelijk onbruikbaar worden door Goodhart's law: wanneer een meetinstrument het doel wordt, houdt het op een goed meetinstrument te zijn. Het is dus noodzakelijk om regelmatig nieuwe benchmarks te ontwikkelen.