

Generatieve AI en machinevertaling

PRAKTISCHE HANDLEIDING VOOR HET ONDERWIJS

Auteurs: Dr. Gys-Walt van Egdom (Universiteit Utrecht)
en drs. Emma Hartkamp (Europese Commissie)

Deze module is onderdeel van het lesmateriaal voor European Parliament Ambassador Schools, maar kan ook door andere scholen in het voortgezet onderwijs (vo) en het middelbaar beroepsonderwijs (mbo) worden gebruikt.

De module richt zich tot docenten vo en mbo en is ook geschikt voor bovenbouwleerlingen en mbo-studenten. Met deze module willen we bewustwording van AI, met name onder jongeren, bevorderen.

Bij de totstandkoming van de module is er gebruikgemaakt van de expertise van Christophe Declercq, Anne-Marie Eekhout, Marloes Meijer en Jonah Zijlma, die kritische kanttekeningen hebben geplaatst en oefeningen hebben aangedragen.

Een beknopte versie van deze module is te vinden in de bijbehorende PowerPointpresentatie, die voor klassikaal onderwijs kan worden gebruikt.

De module is ook digitaal te lezen als online-versie.

Den Haag, september 2023, eerste druk

Inhoudsopgave

Inleiding	5
DEEL 1. Generatieve AI en machinevertaling	6
1A. Generatieve AI	6
Wat is Generatieve AI?	6
Een korte terugblik op het ontstaan van Generatieve AI	8
Mogelijkheden en beperkingen van Generatieve AI	14
1B. Machinevertaling	17
Wat is machinevertaling?	17
Een korte terugblik op het ontstaan van machinevertaling	17
Mogelijkheden en beperkingen van machinevertaling	22
DEEL 2. Wat doet de EU op het gebied van generatieve AI en machinevertaling?	24
2A. Europese AI-wetgeving	24
2B. eTranslation en een Europese ChatGPT?	26
2C. Europese projecten tegen desinformatie	26
DEEL 3. Oefeningen	29
3A. Quizvragen: test je kennis van AI-taaltools!	29
3B. Praktische oefeningen met AI	30
3C. Vertaalopdrachten	34
3D. Discussieonderwerpen	37
Nawoord	40
Diverse links	41
Bronnen	42

Inleiding

Kunstmatige intelligentie (AI) is volop in het nieuws en dat komt vooral door de doorbraak van zogenaamde *Generatieve AI*.

Zoals blijkt uit de eerste onderzoeken maken jongeren maar al te graag gebruik van deze nieuwe technologie, waarmee tools als ChatGPT, DALL-E en Bard zijn gemaakt. In de samenleving klinkt een positief geluid over veelbelovende AI-toepassingen, zoals superchatbots en tools om ouderen langer thuis te laten wonen, de communicatie tussen vluchtelingen en artsen te vergemakkelijken of sneller ziektes te detecteren. Steeds vaker worden echter ook de risico's voor democratie en rechtsstaat belicht. Er klinkt een steeds luider roep om voorwaarden en regelgeving, om desinformatie, schending van privacy of auteursrechten, discriminatie en fraude tegen te gaan. Verder is ook de ecologische voetafdruk van AI steeds vaker onderwerp van gesprek.

In het eerste hoofdstuk van deze module zoomen we in op Generatieve AI en machinevertaling. De belangrijkste begrippen en tools passeren de revue, en de mogelijkheden en beperkingen van de tools komen hier aan bod.

In hoofdstuk twee wordt beschreven wat de EU op dit gebied onderneemt om iedereen toegang te geven tot veilige tools en om de Europese democratie te beschermen.

In het laatste hoofdstuk staan quizvragen, praktische oefeningen en discussieonderwerpen die kunnen bijdragen tot een kritisch, (doel)bewust en efficiënt gebruik van de tools.

We willen polarisatie voorkomen, AI bespreekbaar maken in de klas en er zo voor zorgen dat leerlingen en studenten zelf de juiste balans kunnen vinden. We proberen hiermee de vraag over het uitbannen of omarmen van taaltools en AI in de klas of studie minder scherp te stellen en ruimte te bieden voor nuance, voor het vinden van de juiste balans tussen slim gebruik van de technologie en blijvende aandacht voor 'klassieke' taalvaardigheden en kritisch denken.

Deze module leidt tot een beter begrip van de nieuwe tools en maakt leerlingen en studenten wegwijs in het gebruik ervan. Degelijk taalonderwijs en onderwijs in tekstanalyse blijven hierbij van wezenlijk belang. Net als dat rekenkundig inzicht nodig blijft om een rekenmachine te kunnen gebruiken, blijven jongeren taalkennis en -inzicht nodig hebben om zich kritisch van de taaltools te kunnen bedienen. Door aandacht te geven aan beide vaardigheden, slim gebruik van digitale tools én taalvaardigheid, wordt niet alleen de kritische blik van jongeren op AI gescherpt, een vaardigheid die zij in de toekomst hard nodig zullen hebben, maar kunnen zij ook als digitaal geletterde burgers deelnemen aan de Europese democratie.

DEEL 1. GENERATIEVE AI EN MACHINEVERTALING

1A. Generatieve AI

WAT IS GENERATIEVE AI?

Generatieve AI (ook wel GenAI) is een vorm van kunstmatige intelligentie die nieuwe dingen maakt, zoals teksten, afbeeldingen, geluiden. Bekende voorbeelden van programma's die steunen op Generatieve AI zijn ChatGPT en DALL-E (van OpenAI) en Bard (van Google). De ontwikkelingen gaan zeer snel en er komen steeds nieuwe modellen en programma's bij. Bij Generatieve AI geef je als gebruiker het programma een opdracht, een zogenaamde 'prompt', meestal in de vorm van een tekst. Het AI-model ontleedt die opdracht en produceert dan nieuwe inhoud op basis van wat het uit ongekend grote hoeveelheden bestaande informatie of 'data' heeft geleerd.

Een voorbeeld:

- In je prompt vraag je om een stilleven in de stijl van Picasso te schilderen.
- Het programma analyseert je opdracht en zal vooral de aandacht vestigen op inhoudswoorden als 'stilleven', 'stijl', 'Picasso' en 'schilderen'.
- Het programma grijpt terug naar voorbeelden die het in de dataset heeft gebruikt om over stillevens, stijlen, Picasso en schilderen te leren.
- Het programma formuleert een antwoord door die verschillende soorten informatie uit de dataset te combineren.

Het antwoord vertoont dan gelijkenissen met al die informatiebronnen, maar het is nieuw en lijkt uniek. Door de prompt aan te passen, zal het programma teruggrijpen naar andere voorbeelden en zal het ook die weer combineren. Hiernaast staan enkele voorbeelden van uitkomsten na kleine aanpassingen van opdrachten in DALL-E. Onder elke afbeelding lees je de ingevoerde prompt.



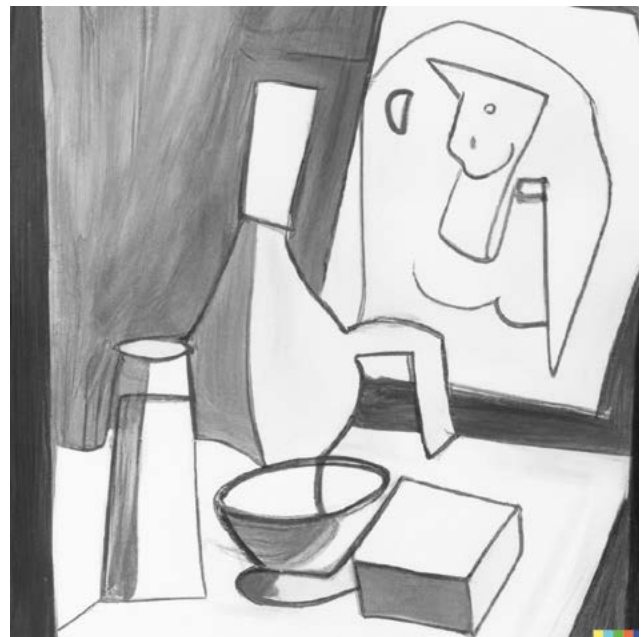
Schilder een stilleven in de stijl van Picasso.



Schilder een stilleven in de kubistische stijl van Picasso.



Schilder een stilleven in de stijl van Picasso in zijn 'blauwe periode'.



Schets een stilleven in stijl van Picasso.

EEN KORTE TERUBLIK OP HET ONTSTAAN VAN GENERATIEVE AI

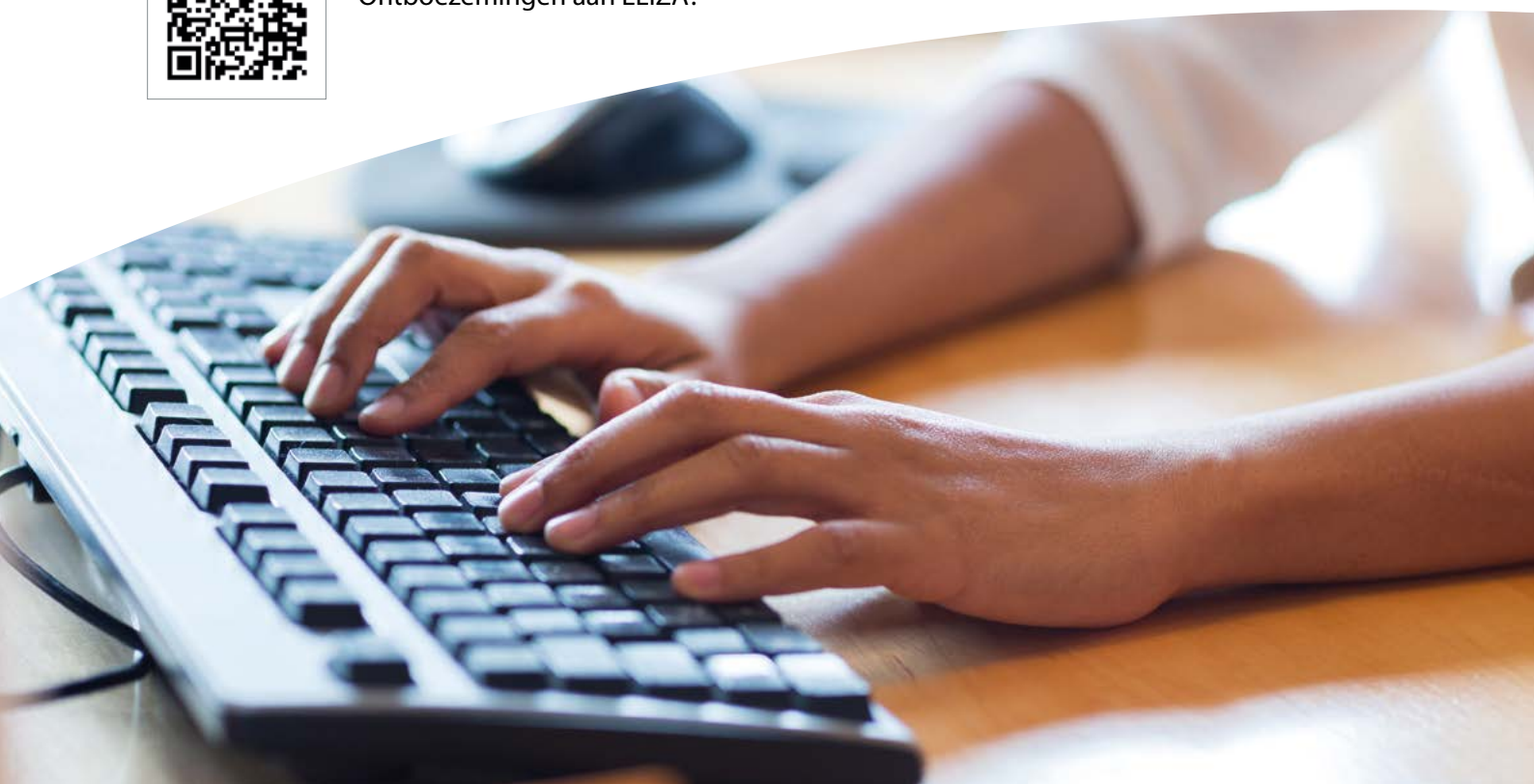
Hoe is Generatieve AI eigenlijk ontstaan? We kunnen stellen dat Generatieve AI een zeer lang traject heeft afgelegd. Het is moeilijk om een beginpunt aan te wijzen. De mens is al eeuwen gefascineerd door de werking van taal. Vooral aan het begin van de Koude Oorlog begon men hier op een meer systematische manier over na te denken. Het idee dat ten grondslag ligt aan alle taaltechnologie is dat taal bepaalde patronen volgt om betekenis te krijgen. Kortom, wie de patronen blootlegt, heeft de sleutel tot taaltechnologie.

Velen zien **ELIZA** als een voorloper van de Generatieve AI die wij kennen. Tussen 1964 en 1966 werkte Joseph Weizenbaum aan deze 'chatterbot'. Het programma was in staat om een stevige conversatie te voeren. Althans, voor veel gebruikers leek het alsof het programma echt begreep wat je zei. Bekend is het voorbeeld dat Weizenbaum zelf aanhaalde om zijn programma aan te prijzen: na een paar zinnestjes zou zijn secretaresse de computerwetenschapper hebben gevraagd om de ruimte te verlaten, omdat het gesprek zo persoonlijk werd (zie link aan het eind van de module). Van begrip was echter geen sprake: het programma analyseerde de input van de gebruiker, herkende daarin bepaalde woorden, en toverde daarop een aantal standaardreacties tevoorschijn.

Na ELIZA was er nog veel activiteit in de computerwetenschap, maar omdat veel bollebozen te zeer uitgingen van het idee dat taal met simpele wiskundige modellen kon worden doorgrond, waren de resultaten niet echt indrukwekkend.



[Ontboezemingen aan ELIZA?](#)





Een neuraal netwerk probeert de menselijke cognitieve structuur na te bootsen.

Machine learning

Misschien wel de belangrijkste stap in de ontwikkeling van Generatieve AI was '*machine learning*'. De laatste twee decennia van de vorige eeuw begon men in te zien dat je met data hele complexe statistische modellen kon ontwikkelen die veel meer in hun mars hadden dan modellen die volledig door mensen waren geprogrammeerd. Bij machine learning worden data aangereikt en wordt de logica achter die data als het ware aan het computerprogramma uitgelegd. Het vernuft van machine learning ligt in het zelflerend vermogen van het programma: na deze eerste stap, tracht het programma niet slechts te reageren op nieuwe input, maar analyseert het die data ook door ze te vergelijken met de data in de set. De analyse van nieuwe data leidt tot een bijstelling van de interne parameters. Je zou kunnen stellen dat bij machine learning het model zichzelf voortdurend bijstuurt.

Tijdens de jaren '80 en '90 was het wel bekend dat je met deze zogenaamde *datagedreven* benadering zeer indrukwekkende resultaten kon boeken, maar de echte opmars van Generatieve AI bleef uit. Waarom werd er niet doorgepakt? Wie tijdens de jaren '80 en '90 gebruikmaakte van een computer kan het antwoord al raden. Computers konden nog niet zoveel: het procesvermogen en de opslagruimte waren beperkt. Daarnaast stond het internet in de kinderschoenen, waardoor de beschikbare data te beperkt waren om echt resultaat te boeken.

Neurale netwerken

Pas rond 2010 laaide het enthousiasme weer echt op. Enkele computertaalkundigen lieten zien dat er met technieken die onder het stof waren beland eigenlijk veel meer mogelijk was dan je zou denken. Doordat er met grotere datasets kon worden geëxperimenteerd en die data door

computers met veel vermogen konden worden geanalyseerd, daalde het besef weer in dat zogenaamde *neurale netwerken* de sleutel vormden tot nieuwe taaltechnologie. Een neurale netwerk is een complex statistisch model. Ook bij deze netwerken worden wiskundige en statistische principes gebruikt om relaties en patronen in taal- en andere data te modelleren. Je zou kunnen stellen dat er bij oude taalmodellen simpelweg een getalletje aan een woord werd gekoppeld. Bij een neurale netwerk is er niet slechts één getal. Bij neurale netwerken worden er op meerdere niveaus 'getalletjes' of 'waarden' aan data gekoppeld. Het resultaat? Het model weet een verfijndere voorstelling te vormen van de 'betekenis' van data.

Word embeddings

'*Word embeddings*' zijn typische voorbeelden van die verfijndere voorstelling. Door, na intensieve training, op verschillende niveaus 'waarden' aan woorden toe te kennen, krijgt elk woord als het ware een uniek coördinaat. Laten we 'hutspot' als voorbeeld nemen: door de verschillende contexten waarin dit woord wordt gebruikt uitvoerig te analyseren, verbindt het programma een coördinaat aan 'hutspot'. Aangezien wij het woord 'hutspot' veel in de context van 'eten' gebruiken, bevindt het coördinaat van 'hutspot' zich op een relatief kleine afstand van het coördinaat van 'eten'. Het coördinaat van 'wortel' zal zich ook op een relatief kleine afstand van 'hutspot' bevinden. Datzelfde geldt voor 'koken', 'ui', 'rookworst' etc. Op basis van vele voorbeelden weet het programma dus in te schatten dat woorden relatief vaak een context delen. De link tussen 'hutspot' en 'eten' zal bijvoorbeeld sterker zijn, dan die tussen 'hutspot' en 'kookboek'. Dit heeft puur te maken met het feit dat de woorden 'hutspot' en 'eten' vaker in elkaars nabijheid worden aangetroffen in een dataset. Maar de link tussen 'hutspot' en 'kookboek' zal weer veel sterker zijn dan die tussen 'hutspot' en 'trui'. En dat om dezelfde reden: de woorden 'hutspot' en 'trui' tref je relatief zelden in dezelfde zin aan.

In de loop van het vorige decennium zijn computerwetenschappers de statistische taalmodellen langzaam gaan verfijnen. Eigenlijk hebben ze dit vooral gedaan door te kijken naar de verschillende manieren waarop deze modellen taal- en andere data zouden kunnen lezen. Een heel interessante ontdekking werd in 2017 gedaan. Toen werd ontdekt dat data-analyse met veel ruis gepaard ging: de programma's kenden relatief veel 'waarde' toe aan onzinnige elementjes in de taal: een lidwoord (de, het, een) werd op vergelijkbare manier geanalyseerd als een zelfstandig of bijvoeglijk naamwoord (stampot, gaar). 'Attention is all you need' luidde vanaf dat moment het credo. Met de juiste focus kon je veel efficiënter te werk gaan: analyses konden efficiënter en vooral beter worden uitgevoerd. De standaardwerkwijze, waarbij je hele sequenties, bijvoorbeeld hele zinnen, van begin tot einde doorloopt, werd losgelaten.

BERT (Bidirectional Encoder Representations from Transformers)

In die periode ontstonden de eerste programma's die steunden op de zogenaamde 'Transformer-architectuur'. Met deze spits technologie werden dus veel betere betekenisverbanden gelegd en was de volgorde waarin de data werden verwerkt ineens een stuk minder belangrijk. Google ontwikkelde een reeks taalmodellen onder de naam '**BERT**' ('*Bidirectional Encoder Representations from Transformers*'). Ook deze modellen steunden op de bekende word embeddings, maar de voorstelling van de betekenis van woorden werd beter doordat er zowel op de context voor als achter het woord werd gelet. Als onderdeel van de training kreeg het programma eigenlijk steeds oefeningen in begrijpend lezen voorgelegd. Neem een zin als: 'Ik ga vanavond hutspot ... bij mijn ouders'. BERT werd geïnstrueerd om, op basis van alle voorbeeldzinnen die het programma al had verwerkt, te voorspellen welk woord er op de stippelijntje moet staan. Het antwoord laat zich voor ons mensen raden: 'eten' prijkt bovenaan ons lijstje van meest waarschijnlijke oplossingen; 'koken' is een tweede optie. Voor taalmodellen is deze oefening lange tijd geen abc'tje geweest. BERT kwam wel vaak met rake suggesties: 'eten' luidde de voorspelling. En ook de nodige nuances wist het programma aardig te treffen. Bij de zin 'Ik ga vanavond hutspot ... voor mijn ouders', zou 'koken' als meest waarschijnlijke oplossing worden aangewezen.

GPT (Generative Pre-Trained Transformer)

Met ChatGPT – GPT staat voor '*Generative Pre-Trained Transformer*' – is de bi-directionaliteit komen te vervallen. Dat wil zeggen: er wordt niet langer gekeken naar wat er voor en na een woord staat. Zoals de experts het steeds opnieuw zeggen, als ze de vraag krijgen wat het programma nu precies doet: ChatGPT voorspelt gewoon het volgende woord. Je zou het programma kunnen vergelijken met een autocomplete-functie: als je een zin typt, zie je dat de voorspelling van autocomplete steeds iets accurater wordt. Zo is het programma in staat om teksten te genereren. Dat het programma lijkt te begrijpen wat je in een prompt vraagt, heeft dan weer te maken met dat aandachtsprincipe: het programma formuleert een hypothese rond de betekenis van jouw input op basis van woorden (die bepaalde coördinaten in het model hebben). Daarom is iets als 'prompt engineering' in korte tijd ook zo belangrijk geworden: het programma begrijpt taal niet als een 'echt' mens – die eigenlijk ontzettend veel tussen de regels weet te lezen. De hypothese rond de afgeleide betekenis van jouw verzoek wordt enkel accurater naarmate je preciezer weet te formuleren. Wat veel gebruikers dus niet beseffen: de input is de basis van een bruikbare output.



Het trainen van AI vergt enorm veel data!

Reinforcement learning

GPT-modellen hebben in een paar jaar tijd overigens ook een aardig traject afgelegd. In 2018 kwam OpenAI met het eerste model op de proppen: GPT-1. Hoewel het model direct al hoge ogen gooide, werd een jaar later GPT-2 al gelanceerd. Bij elk nieuw model zagen we dat vooral de omvang van de gebruikte data toenam en het aantal parameters waar het systeem op werd getraind. Voor de training van GPT-2 is ongeveer 40 Gigabyte aan tekst gebruikt, terwijl het taalmodel GPT-3 niet minder dan 570 Gigabyte aan tekst kreeg aangereikt. Ook het aantal *parameters* nam hand over hand toe. Parameters zijn de aanpasbare gewichten en instellingen binnen het neurale netwerk. Ze bepalen hoe het model informatie uit input verwerkt en hoe het output genereert. Parameters worden aangeleerd als het model wordt getraind. Ze maken het mogelijk om patronen en structuren (woordgebruik, zinsbouw etc.) in taal te doorgronden en uiteindelijk te reproduceren. Kijken we naar GPT-1 dan zien we dat dat model 'slechts' 117 miljoen parameters hanteerde. Zijn opvolger, GPT-2, kon beschikken over 1,5 miljard parameters om taal te analyseren en te produceren.

GPT-1 presteerde al beter dan andere taalmodellen. GPT-2 zorgde ervoor dat de mensen bij OpenAI echt begonnen te geloven dat er toekomstmuziek in GPT-modellen zat. In 2019 wist het systeem al samenhangende teksten te produceren en had een enkeling al het gevoel dat er met een mens werd gecommuniceerd. Toch waren er nog veel verbeterpuntjes. Soms spuwde het programma onzin of feitelijk incorrecte informatie. Zorgwekkender nog was het feit dat GPT-2 ook kwetsende teksten kon produceren. Het grote succes van GPT-3 (en de bètaversie GPT-3.5, die in de gratis versie van ChatGPT wordt gebruikt) is dan ook te danken aan *'reinforcement learning'*.

Bij reinforcement learning leert een taalmodel een onderscheid te maken tussen gewenste en ongewenste output. Eenvoudig gesteld ontvangt het

systeem een beloning of wordt het gestraft nadat het een antwoord heeft geproduceerd. Bekend is dat OpenAI veel onderbetaalde Kenianen heeft ingezet om resultaten op dit vlak te boeken. Bij OpenAI is er sprake van reinforcement learning 'with human feedback'. Menselijke beoordelaars hebben daarbij steeds tal van reacties op een prompt doorgenomen en hebben vervolgens de reacties gerangschikt, van goed naar slecht. Het taalmodel kon zo worden gefinetuned: GPT kreeg een beter idee van de kwaliteit van de eigen output zonder dat het model helemaal opnieuw hoefde te worden getraind. De output van GPT-3 is feitelijk correcter, minder kwetsend en beter afgestemd op de verwachtingen van gebruikers.

Inmiddels is ook GPT-4 gelanceerd. Dit model is geïntegreerd in de betaalde versie van ChatGPT en werkt op de achtergrond bij Bing Chat en bepaalde apps.

Overigens moeten er bij die GPT-modellen wel de nodige kanttekeningen worden geplaatst. ChatGPT lijkt taalafhankelijk te werken, maar gooit eigenlijk alle taal (dus ook beeldtaal) op een hoopje. Een netwerk van woorden is dus echt een allegaartje. Die woorden moeten ook ergens vandaan komen. Waar komen de woorden vandaan? Het antwoord is: van het hele internet. Het is dan ook geen wonder dat ChatGPT vooral in het Engels hoge ogen gooit: ongeveer 60 procent van het hele internet wordt gevuld met Engelstalige teksten. In andere talen zijn de prestaties van ChatGPT al een stuk minder indrukwekkend. Zelfs in het Nederlands merk je soms dat er zinnen worden geproduceerd die wat Engels aandoen. De output zal in minder goed vertegenwoordigde talen of 'less-resourced languages' nog een stuk minder goed presteren.

MOGELIJKHEDEN EN BEPERKINGEN VAN GENERATIEVE AI

Mogelijkheden

Er zijn dus steeds meer programma's en tools beschikbaar die zijn gemaakt met Generatieve AI, zoals ChatGPT en Bard, voor het genereren van tekst, en DALL-E voor het maken van beeld.

ChatGPT kan worden gebruikt als denkpartner of schrijfhulp. De bot kan om ideeën worden gevraagd, kan een tekst corrigeren op taal- of spelfouten of alinea's beter op elkaar doen aansluiten. Ook lukt het de bot vaak om concepten uit wollige teksten helder uit te leggen. De bot kan als zoekmachine worden gebruikt, kan teksten vertalen, samenvatten, herschrijven in eenvoudiger taal en feedback op teksten geven. Ook kan ChatGPT codes voor computerprogramma's schrijven. Naast deze inmiddels wel bekende tools, zijn er nog tal van minder bekende tools, zoals Boomy, die liedjes schrijft op basis van de omschrijving van een bepaalde muziekstijl en het tempo.

Beperkingen en risico's

Om een kritisch gebruik van de tools te stimuleren, willen we hier bepaalde misverstanden wegnemen en ook de belangrijkste beperkingen en risico's kort benoemen. Zo hebben de tools nog altijd weinig zicht op context en is er geen sprake van bewustzijn of identiteit. Dit klinkt logisch, maar instinctief zijn wij geneigd om betekenis en logica aan taal te hechten, ook als deze voortkomt uit levenloze objecten. Denk maar aan de secretaresse van Joseph Weizenbaum. De tool kan de gegeven informatie niet 'controleren' door terug te gaan naar de bron; zelfs bij toepassingen waarbij wel naar bronnen wordt verwezen, zoals bij de tool Perplexity, is de informatie gekoppeld aan losse fragmenten in die bronnen. Als er niet genoeg data zijn, heeft de bot de neiging om zelf iets te verzinnen. Dit noemen we 'hallucineren'. Met bepaalde prompts kun je hallucinatie zelfs

De kleine zeemeermin in Kopenhagen, Denemarken



aanmoedigen (zie praktische oefening 2 aan het eind van deze module). Het is niet altijd duidelijk waarom de machine hallucineert, zoals duidelijk wordt in dit voorbeeld:

Op de vraag aan Perplexity 'Does the little mermaid feel pain when walking in the disney version?' komt een verkeerd (en onsamenhangend) antwoord: 'Yes, in the Disney version of "The Little Mermaid," the Little Mermaid feels pain when walking on her new legs.'

Als je de bronnen raadpleegt, zie je dat deze het antwoord van Perplexity tegenspreken ('In Disney's adaptation, the magic used to transform Ariel into a human is quick and painless, so there is no real cost other than her gorgeous voice... which is also taken painlessly.') Saillante 'bijvangst': de vergelijking tussen Disney en Andersen lijkt overtuigend, maar zelfs die analyse klopt niet (in Den Lille Hafvrue: 'er som du traadte paa en skarp Kniv, saa dit Blod maatte flyde,' '... is het alsof je op een scherp mes stapt, waardoor er bloed vloeit.' [eigen vertaling]). Benadrukt moet worden dat het hier nog een enigszins subtiele hallucinatie betreft, waarin passende bronnen aangehaald worden – zij het op volstrekt onnauwkeurige wijze. Hallucinaties zijn vaak nog flagranter als er überhaupt geen bronnen zijn.

Does the little mermaid feel pain when walking in the disney version?

⚡ Quick Search

The Little Mermaid - Disneyfication: The...
weebly · 1

How Disney's 'The Little Mermaid' Changed the...
collider · 2

The Little Mermaid Wisely Changed The Original...
screenrant · 3

10 Ways Disney's 'The Little Mermaid' Differs...
collider · 4

What the Real (not Disney) Little Mermaid Has to Sa...
medium · 5

10 Dark Storylines That Disney Movies Left Out -...
whatculture · 6

☰ Answer

Yes, in the Disney version of "The Little Mermaid," the Little Mermaid feels pain when walking on her new legs. According to the search results, the transformation and every step the Little Mermaid takes on land is painful, like walking across knives ¹ ² ³ ⁴ ⁶. This is a significant difference from the original story by Hans Christian Andersen, where the Little Mermaid's transformation is also painful, but the pain is described as being like being cut in two rather than walking on knives ¹ ³ ⁵. Therefore, in the Disney version of "The Little Mermaid," the Little Mermaid feels pain when walking on her new legs.

🚩 [] 🗑️ ✎️ ...

Door zulke beperkingen gaat de opkomst van Generatieve AI gepaard met risico's, bijvoorbeeld op het gebied van *desinformatie*, *discriminatie*, *auteursrecht*, en *privacy*.

- **Desinformatie:** omdat AI-tools verstrekte informatie niet adequaat kunnen controleren, kunnen ze desinformatie verspreiden. Denk maar aan die hallucinaties. Bovendien is het met behulp van Generatieve AI veel eenvoudiger om desinformatie met opzet te fabriceren. Sommige systemen kunnen bijvoorbeeld iemands stem overtuigend nabootsen of een afbeelding produceren waarvan het moeilijk is om vast te stellen dat deze door een machine is gegenereerd.
- **Discriminatie:** gegenereerde teksten kunnen discriminerend zijn, soms op subtiele wijze. Bij grote taalmodellen zijn namelijk trainingsdata gebruikt die afkomstig zijn van (vrijwel) het gehele internet. Bij de voorgangers van het huidige ChatGPT waren bijvoorbeeld ook bepaalde 'subreddits' – een soort fora – verwerkt, die bekend stonden om openlijk discriminerend taalgebruik. Bovendien is het goed mogelijk dat sommige maatschappelijke groepen juist ondervertegenwoordigd zijn in de trainingsdata. AI-tools zouden daardoor ook onbedoeld kunnen discrimineren, door minder rekening te houden met de ervaringen van deze groepen.
- **Auteursrecht:** door de verwerking en het hergebruik van teksten waarop rechten rusten, zijn er grote zorgen over de schending van auteursrechten bij het werven van trainingsdata.
- **Privacy:** de input van gebruikers, inclusief persoonsgegevens, wordt voor taalmodellen gebruikt. Hierbij ligt ook het risico op fraude op de loer.

Beoordeel dus altijd kritisch de output van AI-tools.

Wat kan je doen om de risico's te verminderen wanneer je met AI-tools aan de slag gaat?

- Wees alert op **onjuiste informatie**. Hiervoor is (enige) kennis van het onderwerp nodig.
- Houd rekening met **vooroordelen of discriminatie**: zo gaat een chatbot er in een zakelijke brief mogelijk van uit dat deze aan een man gericht is.
- Controleer of er geen persoonsgegevens in de tekst staan om **schending van privacy** te voorkomen. Deel nooit vertrouwelijke informatie met openbare AI-tools.
- Het risico op **schending van auteursrechten** valt helaas moeilijk in te dammen, omdat het voor de 'gewone gebruiker' ondoenlijk is om ChatGPT hierop te controleren.

Het is bovendien belangrijk om stil te staan bij de grote ecologische sporen die AI-tools achterlaten. De opslag en vooral de verwerking van gegevens om algoritmes te trainen in datacenters of in de cloud slurpen energie. Zo moet een algoritme dat wordt getraind om bijvoorbeeld een kat te herkennen, miljoenen afbeeldingen van katten verwerken. Sommige 'datafarms', vooral in Noord-Europa en Canada verbruiken net zoveel energie als een kleine stad. De ICT, waaronder datacentra, die voor AI wordt gebruikt, is qua brandstofuitstoot vergelijkbaar met de luchtvaart. Ten slotte creëren we een nieuwe afhankelijkheid van grote technologiebedrijven. Het vraagt namelijk een aanzienlijk kapitaal om Generatieve AI te maken, te onderhouden en beschikbaar te stellen.



1B. Machinevertaling

WAT IS MACHINEVERTALING?

Een *machinevertaling* is, zoals de naam al aangeeft, een vertaling die in zijn geheel door een vertaalmachine is gemaakt. Vertaalmachines, zoals Google Translate en DeepL, steunen op technologie die vergelijkbaar is met die van Generatieve AI. Ze berusten op neurale taal- en vertaalmodellen die erop gericht zijn de relaties tussen woorden en zinnen in verschillende talen te analyseren. Ook vertaalmachines zijn afhankelijk van een opdracht, al is de opdracht doorgaans een hele concrete: een brontekst dient te worden vertaald. Vertaalmachines dienen dus een specifiek doel: de machines moeten een correcte en vloeiende weergave van de brontekst in een doeltaal maken.

Met Generatieve AI is het overigens ook mogelijk om teksten te vertalen. De taalmodellen waar Generatieve AI op steunt, maken geen duidelijk onderscheid tussen talen. Echte vertaalmachines zijn vaak getraind met grote hoeveelheden meertalige data, bijvoorbeeld originele teksten en hun vertaling, en kunnen daardoor talen en taalspecifieke patronen beter onderscheiden. De vertalingen die door vertaalmachines worden gemaakt, zijn daardoor vaak beter dan vertalingen die door Generatieve AI zijn geproduceerd.

EEN KORTE TERUGBLIK OP HET ONTSTAAN VAN MACHINEVERTALING

Vertaalmachines delen een groot deel van hun geschiedenis met Generatieve AI. Honderden jaren geleden werd er al door filosofen nagedacht over technieken (of technologieën) waarmee 'machinevertaling' mogelijk zou zijn. Filosofen die zich hiermee bezighielden waren: Ramon Llull (13e eeuw), Blaise Pascal en Gottfried Leibniz (17e eeuw). Echter kwam de technologie in de twintigste eeuw pas echt van de grond. Zo vroeg de Frans-Georgische Georges Artsrouni in 1933 een patent aan voor een

'cerveau mécanique', een mechanisch brein met ingebouwd mechanisch woordenboek, en vroeg de Rus Petr Trojanskij in datzelfde jaar eveneens een patent aan voor een heuse 'vertaalmachine'. Toch staat het voor iedereen buiten kijf dat er pas sinds de jaren '50 echt werk is gemaakt van vertaalmachines.

Tijdens de Tweede Wereldoorlog merkten overheden hoe belangrijk het was om informatie van de vijand te onderscheppen en snel te vertalen. Na de Tweede Wereldoorlog bleef dat belang overeind: hoewel ze tijdens de oorlog nog zij aan zij hadden gevochten tegen de zogenaamde 'asmogendheden' (Duitsland, Italië en Japan), kwamen de Russen en Amerikanen al vrij snel tegenover elkaar te staan. De Russische en de Amerikaanse overheden investeerden daardoor lange tijd flink in vertaaltechnologie. Er heerste optimisme: lange tijd werd geloofd dat vertalers binnen vijf jaar overbodig zouden zijn (zie link aan het einde van deze module).

'Rule-based' systemen

Waarom heerste er zoveel optimisme? Er werd geëxperimenteerd met zogenaamde '*rule-based*' systemen. De machines werden gevuld met meertalige woordenlijstjes en kregen ook nog de grammaticaregels van verschillende talen toegestopt. Stel je voor dat je een zinnetje hebt als: 'I am eating porridge'. Als je deze zin als Nederlander echt wil begrijpen, volstaat het niet om de betekenis van de losse woorden op te zoeken ('ik', 'ben', 'etend', 'pap'). Daarom bedachten de knappe koppen dat je goede resultaten kon boeken door taalregels toe te voegen. Als het systeem weet dat de zin in een zogenaamd 'present continuous' (met een onvoltooid deelwoord) is gevat en het weet hoe die werkwoordsvorm in het Nederlands woord weergegeven, dan is het mogelijk om een goede vertaling te krijgen, namelijk: 'ik ben pap aan het eten'. Computerwetenschappers deden dus hun stinkende best om alle woordjes en grammaticaregels van talen in computers te proppen.

In 1966 kwam het rapport van de ALPAC-commissie uit. Het rapport was vernietigend: ondanks alle investeringen maakten vertaalmachines nog altijd de domste fouten. De taal bleek ongrijpbaarder dan gedacht.



Statistische vertaalmachines

Pas vier decennia later gloorde er weer licht aan de horizon. In de jaren '80 en '90 werd er in de achterkamertjes van IBM gespeeld met statistische vertaalmachines ('*Statistical Machine Translation*' of SMT). Het idee was eigenlijk vrij simpel: door een computerprogramma te voeden met heel veel meertalige data, met name bronteksten en vertalingen, kon je heel gericht waarschijnlijkheidsberekeningen laten uitvoeren. Vandaar dus ook het woord: 'statistisch'. We gaan terug naar ons voorbeeld: 'I am eating porridge'.

Een SMT-systeem zal eerst de zin analyseren en kijkt daarbij naar de structuur van de zin, de woorden en de grammaticale elementen. Bij een ietwat geavanceerder SMT-systeem zullen 'I am eating' en 'porridge' worden herkend. Dan kijkt het systeem als het ware in de data waarmee het gevoed is en zoekt het naar vertalingen van 'I am eating' en 'porridge'. De vertaling die uit het systeem kwam gerold, was een combinatie van de meest waarschijnlijke vertaling van 'I am eating' ('ik ben aan het eten') en 'porridge' ('pap'). Rond 2006 zwierde Google het bekende 'Google Translate' online. De machine steunde op de principes van SMT. De kwaliteit van Google Translate-vertalingen leek de eerste jaren steeds beter te worden. Toch verstomde het optimisme: rond 2010 werden de vertalingen niet meer beter en bleef het systeem nog te vaak worstelen met betekenisweergave, grammatica en samenhang.

Neurale vertaalmachines

Maar ook de ontwikkelaars van vertaalmachines waren inmiddels op de hoogte van de nieuwste ontwikkelingen op het vlak van machine learning en neurale netwerken. Het is zelfs zo dat het grote publiek pas echt beseftte wat er met die nieuwste AI mogelijk was op het moment dat Google in november 2016 met een nieuwe versie van Google Translate kwam. Volgens de specialisten van Google kon een machinevertaling nog nauwelijks van een ('menselijke') vertaling worden onderscheiden, dankzij neurale vertaalmachines ('*Neural machine translation*' of NMT). Heel simpel gesteld zou je kunnen zeggen dat een SMT-systeem op één manier naar een tekst kijkt en vervolgens een voorspelling doet. Een

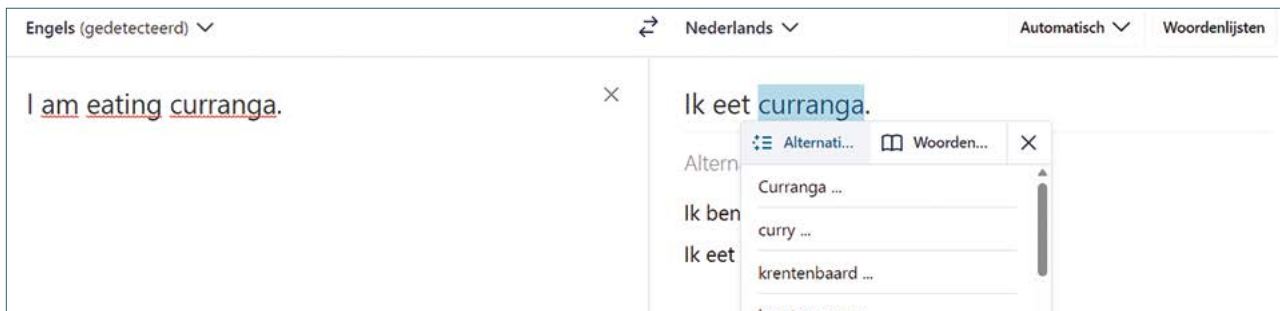
NMT analyseert een tekst op velerlei manieren, waardoor de kans op een correcte vertaling groter is. We leggen het verschil uit aan de hand van een voorbeeld. Stel, het woord 'porridge' is verkeerd gespeld in je brontekst ('porrigde') of het woord 'porridge' komt niet eens voor in je dataset. Bij SMT is de kans in deze situaties groot dat je een waardeloze vertaling krijgt: 'Ik ben *porrigde* aan het eten' of 'ik ben *porridge* aan het eten'. Je krijgt het woord gewoon onvertaald terug, omdat de waarschijnlijkheidsberekening niets oplevert. Bij NMT is de kans, zoals je hieronder ziet, groot dat je een goede vertaling krijgt. Het systeem bekijkt de brontekst op verschillende manieren en houdt bijvoorbeeld rekening met vormelijke overeenkomsten tussen 'porrigde' en 'porridge' (een woord dat wél in de dataset voorkomt). Als het systeem ook nog merkt dat 'porridge' vaak in de context van 'eating' wordt gebruikt, zal het NMT-systeem beslissen dat 'porrigde' wel 'porridge' zal zijn.



Je zou nu denken dat het systeem een ingebouwde 'autocorrectie' kent. Toch is het systeem verfijnder dan je zou vermoeden. Vervangen we 'porrigde' door een fantasiegerecht als 'curranga', dan zien we nog dat de vertaalmachine, in dit geval DeepL, een gokje doet. Het systeem gokt er in eerste instantie nog op dat 'curranga' een exotisch gerecht moet zijn, maar kijkt vervolgens op meerdere manieren naar overeenkomsten tussen het woord 'curranga' (in deze context) en andere soorten eten. Dat heeft alles te maken met de 'aandacht' ('attention') waar we het eerder over hadden. Het is heel moeilijk te achterhalen hoe NMT deze concrete oplossingen heeft berekend, maar blijkbaar zorgt de combinatie eating-curranga ervoor dat het systeem 'curry', 'krentenwegge' en verderop ook 'krentenbollen' en 'krentenbrood' als mogelijke vertalingen typeert.

Exotische krentenwegge?



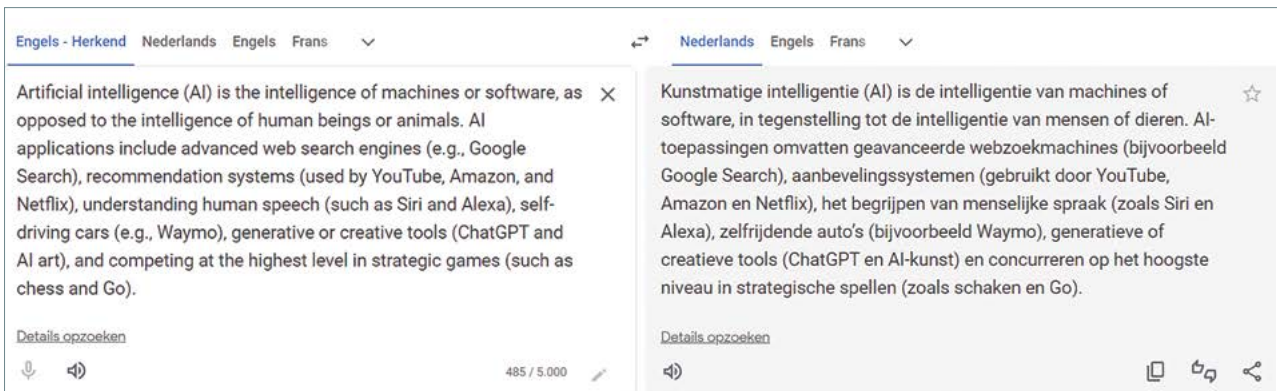


Een belangrijk verschil tussen klassieke NMT en Generatieve AI is dat vertaalmachines vaak getraind zijn met meertalige data, vaak originele teksten en hun vertalingen. Dat is ook niet gek: vertaalmachines dienen een welbepaald doel. Generatieve AI is getraind om verschillende taken uit te voeren. Het is dan ook logisch dat de kwaliteit van vertalingen die met Generatieve AI worden geproduceerd nog niet geweldig is. Toch mogen we verwachten dat de ontwikkelaars van vertaalmachines in de toekomst meer naar Generatieve AI kijken, vooral omdat vertaalmachines nog erg letterlijk vertalen en Generatieve AI veel beter in staat is om 'creatieve' oplossingen te vinden.

Een ding staat buiten kijf: vertaalmachines hebben veel gebruikers de laatste jaren versteld laten staan. Toch schuilt er ook in machinevertaling een gevaar. Onverantwoord gebruik kan ernstige gevolgen hebben.

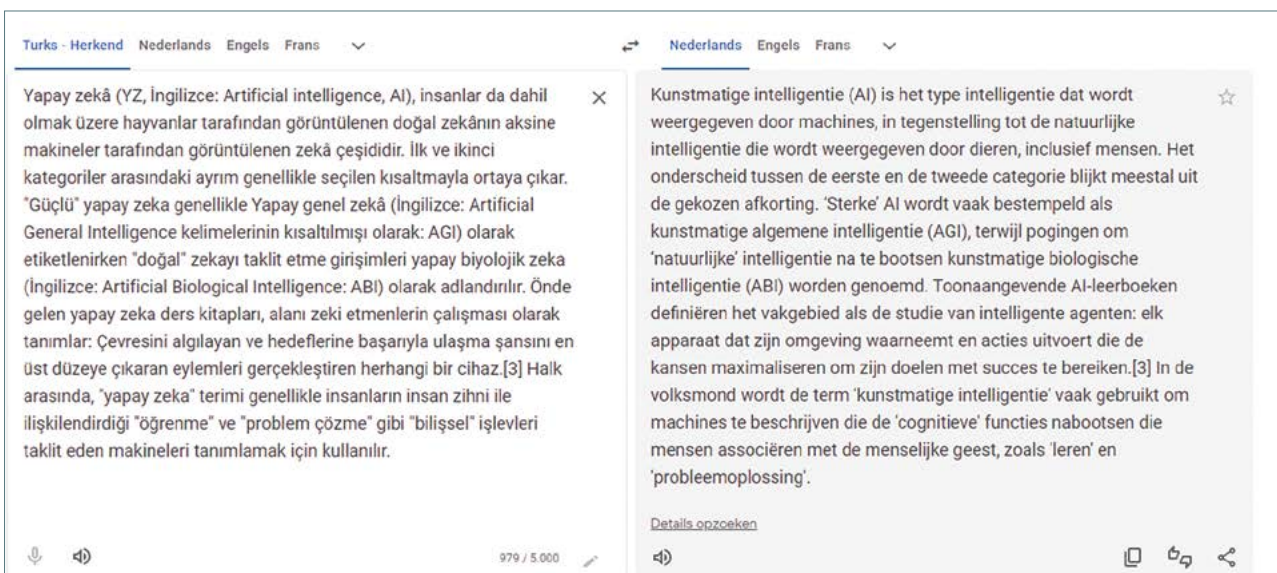
MOGELIJKHEDEN EN BEPERKINGEN VAN MACHINEVERTALING

Tegenwoordig is de kwaliteit van machinevertaling behoorlijk goed. Toch is de kwaliteit van de machinevertaling sterk afhankelijk van een aantal factoren. Teksten over complexe onderwerpen worden niet altijd even goed weergegeven. Hetzelfde geldt voor teksten met erg lange of juist erg korte zinnen. Ook met uitdrukkingen en gezegden weet de machine niet altijd goed raad. Verder speelt de taalcombinatie een belangrijke rol: een machinevertaling uit het Engels is doorgaans veel beter dan een vertaling uit het Turks of Armeens.



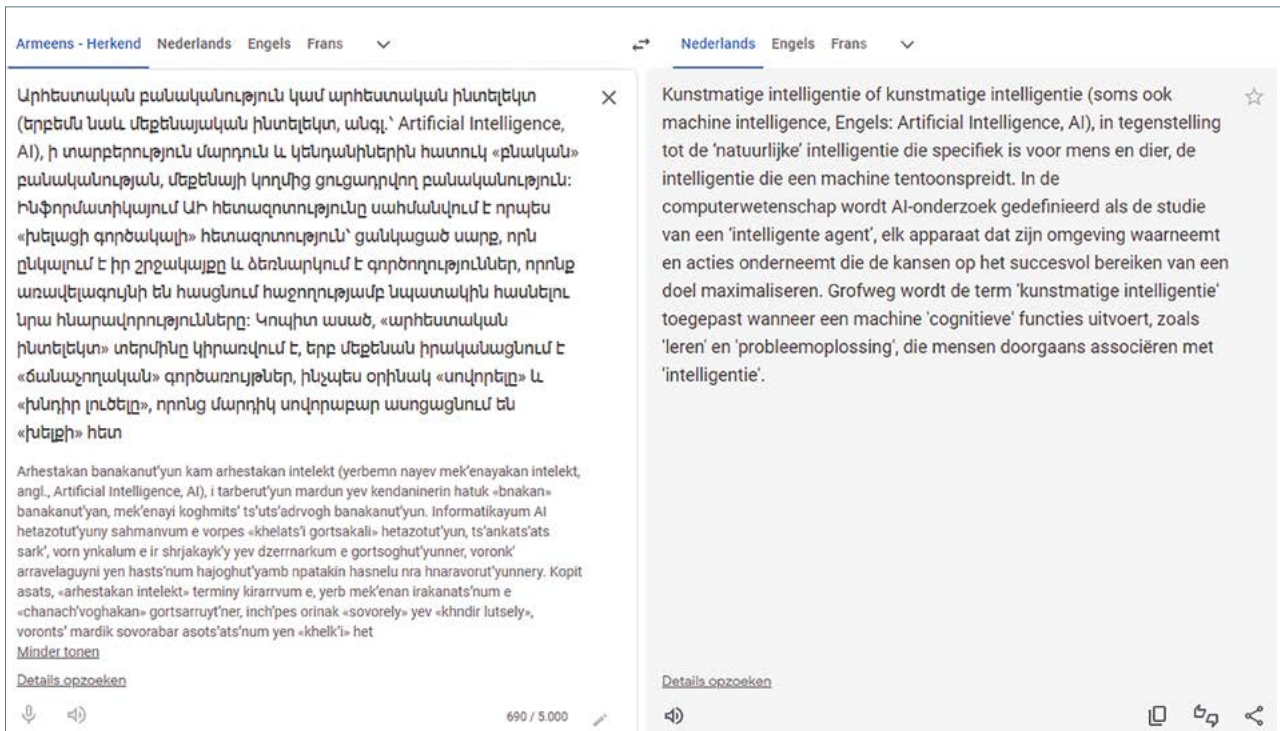
The screenshot shows the Google Translate interface. On the left, the source text in English is: "Artificial intelligence (AI) is the intelligence of machines or software, as opposed to the intelligence of human beings or animals. AI applications include advanced web search engines (e.g., Google Search), recommendation systems (used by YouTube, Amazon, and Netflix), understanding human speech (such as Siri and Alexa), self-driving cars (e.g., Waymo), generative or creative tools (ChatGPT and AI art), and competing at the highest level in strategic games (such as chess and Go)." On the right, the translated text in Dutch is: "Kunstmatige intelligentie (AI) is de intelligentie van machines of software, in tegenstelling tot de intelligentie van mensen of dieren. AI-toepassingen omvatten geavanceerde zoekmachines (bijvoorbeeld Google Search), aanbevelingssystemen (gebruikt door YouTube, Amazon en Netflix), het begrijpen van menselijke spraak (zoals Siri en Alexa), zelfrijdende auto's (bijvoorbeeld Waymo), generatieve of creatieve tools (ChatGPT en AI-kunst) en concurreren op het hoogste niveau in strategische spellen (zoals schaken en Go)." The interface also shows language selection (Engels - Herkend, Nederlands), search options, and character counts (485 / 5.000).

Vertaling van Google Translate van de eerste alinea van de Wikipedia-pagina 'Artificial Intelligence' (Engels)



The screenshot shows the Google Translate interface. On the left, the source text in Turkish is: "Yapay zekâ (YZ, İngilizce: Artificial intelligence, AI), insanlar da dahil olmak üzere hayvanlar tarafından görüntülenen doğal zekânın aksine makineler tarafından görüntülenen zekâ çeşididir. İlk ve ikinci kategoriler arasındaki ayrım genellikle seçilen kısaltmayla ortaya çıkar. "Güçlü" yapay zeka genellikle Yapay genel zekâ (İngilizce: Artificial General Intelligence kelimelerinin kısaltılmışı olarak: AGI) olarak etiketlenirken "doğal" zekâyı taklit etme girişimleri yapay biyolojik zeka (İngilizce: Artificial Biological Intelligence: ABI) olarak adlandırılır. Önde gelen yapay zeka ders kitapları, alanı zeki etmenlerin çalışması olarak tanımlar: Çevresini algılayan ve hedeflerine başarıyla ulaşma şansını en üst düzeye çıkaran eylemleri gerçekleştiren herhangi bir cihaz.[3] Halk arasında, "yapay zekâ" terimi genellikle insanların insan zihni ile ilişkilendirdiği "öğrenme" ve "problem çözme" gibi "bilişsel" işlevleri taklit eden makineleri tanımlamak için kullanılır." On the right, the translated text in Dutch is: "Kunstmatige intelligentie (AI) is het type intelligentie dat wordt weergegeven door machines, in tegenstelling tot de natuurlijke intelligentie die wordt weergegeven door dieren, inclusief mensen. Het onderscheid tussen de eerste en de tweede categorie blijkt meestal uit de gekozen afkorting. 'Sterke' AI wordt vaak bestempeld als kunstmatige algemene intelligentie (AGI), terwijl pogingen om 'natuurlijke' intelligentie na te bootsen kunstmatige biologische intelligentie (ABI) worden genoemd. Toonaangevende AI-leerboeken definiëren het vakgebied als de studie van intelligente agenten: elk apparaat dat zijn omgeving waarneemt en acties uitvoert die de kansen maximaliseren om zijn doelen met succes te bereiken.[3] In de volksmond wordt de term 'kunstmatige intelligentie' vaak gebruikt om machines te beschrijven die de 'cognitieve' functies nabootsen die mensen associëren met de menselijke geest, zoals 'leren' en 'probleemoplossing'." The interface also shows language selection (Turks - Herkend, Nederlands), search options, and character counts (979 / 5.000).

Vertaling van Google Translate van de eerste alinea van de Wikipedia-pagina 'Yapay zekâ' (Turks)



Vertaling van Google Translate van de eerste alinea van de Wikipedia-pagina 'Արհեստական բանականություն' (Armeens)

De machine kan er ook 'met de tekst vandoor gaan' en produceert dan een vertaling die niets met het origineel te maken heeft. Dit is ook een vorm van hallucinatie.

Verder is het belangrijk te beseffen dat de teksten die in vertaalmachines worden ingevoerd vaak worden gebruikt om de onderliggende (ver)taalmodellen te trainen, zonder dat de gebruiker dit weet. Deel dus nooit vertrouwelijke informatie met een vertaalmachine en wees je ervan bewust dat de eigenaar van het vertaalsysteem auteursrechten verkrijgt over de vertaalde teksten. Dit geldt overigens niet voor het online vertaalsysteem van de Europese Commissie, eTranslation, dat in het volgende hoofdstuk nader aan de orde komt (zie ook de link aan het einde van deze module).

De ongrijpbaarheid en schoonheid van taal

Ook ruim een halve eeuw na het verschijnen van het rapport van de ALPAC-commissie brengt de output van AI-taaltools nog altijd de ongrijpbaarheid en schoonheid van natuurlijke taal aan het licht. Intuïtieve verbanden die mensen in gesprekken of teksten leggen, ontbreken in AI-tekstuitingen, en AI-tools missen de context waarin natuurlijke taal ontstaat. Kunstmatige intelligentie is misschien slim, maar niet intelligent. En de gevolgen van 'dom' gebruik van slimme tools zijn soms moeilijk te overzien...

DEEL 2. WAT DOET DE EU OP HET GEBIED VAN GENERATIEVE AI EN MACHINEVERTALING?



AI biedt vele mogelijkheden, maar kan ook bijdragen aan de verspreiding van desinformatie en ongewenste uniformiteit, zoals we gezien hebben in het vorige hoofdstuk. Verder weerspiegelt de output van de machine de input, d.w.z. de trainingsdata, afkomstig van vrijwel het gehele internet. Data uit het Engelse taalgebied zijn daarin dominant en bepaalde taal- en cultuurgroepen zijn dus sterker vertegenwoordigd dan andere. Volgens [Statista](#) is bijna 60 procent van het tekstmateriaal online in het Engels gesteld. De top drie wordt vervolledigd door Russisch en Spaans. Beide talen zijn met vijf procent echter nog beperkt vertegenwoordigd.

Wanneer een AI-tool op basis van statistiek tot een oplossing komt, kan de output daarom ongenueanceerd zijn, talig onnatuurlijk ogen door de invloed van het Engels en vooroordelen of zelfs discriminatie bevatten. De precieze inhoud van de machines blijft ondoorzichtig en dubieuze bronnen zijn niet uit te sluiten. Alleen de grote techbedrijven die de tools ontwikkelen, kennen de exacte werking van hun systemen en daarom zijn de gebruikers afhankelijk van keuzes waarin zij zelf geen inzicht hebben. Wel is overigens bekend dat techreuzen inspanningen leveren om ongewenste output tegen te gaan: hierbij wordt gebruik gemaakt van een feedbacksysteem dat reinforcement learning heet, zoals uitgelegd in deel 1 van deze module.

Om de hierboven beschreven risico's te beperken en zo desinformatie en ongewenste invloeden van commerciële bedrijven en landen buiten Europa te voorkomen en de democratie te beschermen, is de EU actief op verschillende fronten. Hieronder leest u daarover meer.

2A. Europese AI-wetgeving

In 2021 kwam de Europese Commissie als eerste wereldwijd met een ontwerpwet om kunstmatige intelligentie op allerlei manieren te reguleren. Toen eind 2022 ChatGPT werd gelanceerd, moest deze '**EU AI Act**' grondig op de schop en op bepaalde punten worden aangevuld. Het moeilijke van het reguleren van AI zit 'm namelijk juist in het feit dat de techniek die met regels moet worden omkaderd, nog volop in ontwikkeling is. Denk maar aan zelfrijdende auto's. Deze door AI gestuurde auto's hebben de potentie om grote maatschappelijke veranderingen teweeg te brengen: wegen kunnen drukker worden, taxi- en vrachtwagenchauffeurs kunnen zonder werk komen te zitten en terroristen zouden auto's kunnen inzetten om bommen af te leveren. Om nog maar niet te spreken over de gevaren voor de verkeersveiligheid. Maar toen de auto's op de markt kwamen, was de wetgeving nog verre van toereikend. Een centrale vraag hierbij draait om de aansprakelijkheid bij ongelukken.



Ook toen zelfrijdende auto's op de markt kwamen, was de wetgeving nog verre van toereikend.

In juni 2023 heeft het Europees Parlement een aangepaste versie van de EU AI Act aangenomen. Na goedkeuring door de lidstaten zou de wet eind 2023 definitief moeten zijn, waarmee Europa mondiaal voorop komt te lopen in de regulering van AI. De effectieve uitvoering van de wet kan dan nog wel twee jaar op zich laten wachten. En de ontwikkelingen staan niet stil... Daarom werkt de Commissie momenteel ook aan een snel in te zetten 'AI-pact', waarin techreuzen als Google en Microsoft vrijwillig afspraken met de EU maken.

AI-toepassingen zijn in de nieuwe EU AI Act in vier risicocategorieën ingedeeld en op basis daarvan kunnen er voor elk AI-systeem transparantie- en controleverplichtingen worden vastgesteld. Ook zijn er regels gemaakt over het gebruik van persoonlijke gegevens en energieverbruik. Over het algemeen:

- moeten AI-systemen **veilig en niet-discriminerend** zijn.
- moet de mens de **controle** over AI houden.
- mogen **grondrechten** niet in gevaar komen.

Bepaalde systemen kunnen volledig worden verboden, zoals real-time gezichtsherkenning.

Generatieve AI-tools, zoals ChatGPT, komen in een aparte categorie te vallen. Ontwikkelaars moeten ervoor zorgen dat:

- de tools **geen illegale inhoud** voortbrengen.
- nadrukkelijk wordt **vermeld dat teksten met behulp van AI** tot stand zijn gekomen, bijvoorbeeld met behulp van een soort watermerk.
- er een **risicoanalyse** wordt gemaakt van de toepassing in de praktijk en inzicht geven in de **data** die zij gebruiken.

2B. eTranslation en een Europese ChatGPT?

Vanaf de oprichting van de Europese Unie staat vertaling centraal bij alles wat zij onderneemt. De Europese Unie is immers een *meertalige democratie*, met 24 officiële talen. EU-burgers kunnen niet alleen in hun eigen taal de Europese wetgeving lezen, maar hebben ook het recht om in hun eigen taal vragen te stellen en informatie te krijgen over onderwerpen die zij belangrijk vinden.

Om deze reden sleutelde de Europese Commissie al in de vroege jaren '70 van de vorige eeuw aan een vertaalsysteem voor eigen gebruik. In eerste instantie bracht een interne onderzoeksgroep van de Europese Commissie een rule-based systeem uit. In de jaren '90 bouwde de Europese Commissie dit rule-based systeem om tot een statistisch vertaalsysteem. Eind vorige eeuw werd dit systeem ook toegankelijk voor gebruikers buiten de Commissie. Het huidige machinevertaalsysteem, **eTranslation**, uitgaande van statistiek en neurale netwerken, wordt nu op grote schaal bij alle EU-instellingen gebruikt. Ook overheden, universiteiten, freelancers, kleine en middelgrote bedrijven en ngo's in de EU-lidstaten kunnen eTranslation gebruiken (zie link aan het einde van deze module).

eTranslation biedt belangrijke veiligheids garanties die commerciële systemen als Google Translate vaak niet bieden, zeker niet bij gratis gebruik. Om deze reden mogen EU-vertalers uitsluitend met eTranslation werken. Zo worden gebruikersgegevens en teksten niet bewaard of als trainingsdata gebruikt en behouden gebruikers de auteursrechten over hun teksten. De trainingsdata worden zorgvuldig 'opgeschoond' en bevatten dus minder 'ruis' dan die van commerciële aanbieders.

De EU wil ook een eigen open *GPTx-systeem* (een openbaar toegankelijke Europese variant van ChatGPT) bouwen, om de afhankelijkheid van de (vaak niet-Europese) techreuzen te beperken.

2C. Europese projecten tegen desinformatie

De Europese Unie steunt projecten om *mediawijsheid* te vergroten, met name onder jongeren, en hen erbij te helpen desinformatie te herkennen en te bestrijden. Een voorbeeld hiervan is het SMILES-project, in het kader waarvan workshops worden aangeboden aan scholen en bibliotheken in Nederland, België en Spanje (zie link aan het einde van deze module).





1

- _____
- _____
- _____
- _____
- _____

A B C D

2

- _____
- _____
- _____
- _____
- _____

A B

3

- _____
- _____
- _____
- _____
- _____

B C D

4

- _____
- _____
- _____
- _____
- _____

A B

DEEL 3. OEFENINGEN

3A. Quizvragen: test je kennis van AI-taaltools!

1. ChatGPT is ontwikkeld door...
 - de EU.
 - de Amerikaanse overheid.
 - een Amerikaans bedrijf.
 - een Amerikaanse non-profitorganisatie.
2. Google Translate gebruikt jouw gegevens...
 - nooit.
 - om het systeem verder mee te trainen.
 - om de onderhoudskosten te financieren.
 - om veiligheidsredenen.
3. eTranslation is...
 - een betaalde en dus veiliger versie van Google Translate.
 - de vertaalmachine van de Europese Commissie.
 - voor iedereen toegankelijk.
 - het vertaalsysteem van de Amerikaanse overheid.
4. AI-wetgeving (meerdere antwoorden mogelijk)...
 - is in de maak in de EU.
 - is zeer complex vanwege de zich snel ontwikkelende technologie.
 - bestaat in China en de VS.
 - is een illusie.
5. AI-tools kunnen leiden tot (meerdere antwoorden mogelijk)...
 - nepnieuws.
 - schending van privacy.
 - schending van auteursrechten.
 - creatieve vondsten.
6. Gebruik AI-tools...
 - nooit.
 - alleen met kritische blik en eventuele aanpassing van de output.
 - alleen voor exacte vakken.
 - alleen wanneer je geen enkele kennis hebt van het onderwerp of de taal waarnaar je vertaalt.

3B. Praktische oefeningen met AI

Legenda:



Makkelijke opdracht



Gemiddelde opdracht



Moeilijke opdracht

Opdrachten met ChatGPT:



1. Opdracht rond de autocomplete-functie van ChatGPT

Opdracht:

Test de 'autocomplete-functie' van ChatGPT: tik een willekeurige zin in en laat daarbij een belangrijk woord open. Vraag ChatGPT om het woord op de stippellijn te voorspellen. Verander je prompt een paar keer of stel aanvullende vragen en kijk hoe dit het antwoord beïnvloedt.¹



2. Laat met opzet ChatGPT 'hallucineren'

Groepsopdracht:

Verzin een naam en een achternaam. De volgende leerling verzint een beroep. Weer een andere leerling verzint twee hobby's. Vraag dan ChatGPT om een lovende biografische tekst te schrijven over [verzonnen naam en achternaam] en vraag daarbij om nadrukkelijk in te gaan op het verzonnen beroepsleven en de hobby's. Geeft ChatGPT gehoor aan je opdracht? Verzint ChatGPT zomaar wat of zegt hij geen kennis te hebben van de verzonnen persoon?

¹ Naar het voorbeeld op blz. 7 van deze module.



Tip: Je kan dit ook doen met bekendere mensen. Begin bijvoorbeeld met mensen die niet heel bekend zijn en werk zo toe naar prompts over hele bekende figuren. Ga na in hoeverre de tool jouw informatie verwerkt en in welke mate de tool zelf in staat is informatie te verzinnen.



3. Brainstormoefening met ChatGPT

Opdracht:

Vraag ChatGPT vijf argumenten voor en vijf argumenten tegen een volumebeperking op muziekfestivals te bedenken. Wat vind je van het resultaat? Zijn de argumenten evenwichtig?



4. Laat ChatGPT jouw tekst corrigeren

Opdracht:

Schrijf 5 minuten aan een stuk aan een tekst. Corrigeer jezelf niet en herzie je tekst niet. Stop abrupt na 5 minuten. Vraag ChatGPT om een samenhangende tekst van jouw verhaal te maken. Doe dat in stapjes. Laat eerst de spelfouten wegwerken. Laat ChatGPT vervolgens de tekst structureren of herstructureren. Vraag ChatGPT ten slotte ook nog om er een samenhangende en kwalitatief goede tekst van te maken. Wat vind je van het resultaat?



5. Laat ChatGPT jouw ideeën ordenen

Opdracht:

Produceer in groepjes een aantal associaties bij een onderwerp. Noteer die een voor een, zonder structuur aan te brengen. Vraag ChatGPT vervolgens om puntsgewijs een opzette voor een tijdschriftartikel over dit onderwerp te schrijven. Is het resultaat bruikbaar? Is er sprake van een logische ordening?

6. Schrijfopdracht met ChatGPT

Opdracht:



A. Vraag ChatGPT een gedicht of haiku te schrijven over een van je hobby's, bijv. een sport, instrument of iets wat je heel leuk of boeiend vindt. Ga na of de inhoud klopt. Is het gedicht goed gelukt? Verbeter dan het gedicht en gebruik hierbij je eigen woorden.



B. Vraag ChatGPT een gedicht of haiku te schrijven over jouw favoriete schoolvak. Ga na of de inhoud klopt. Is het gedicht goed gelukt? Verbeter dan het gedicht en gebruik hierbij je eigen woorden.



C. Vraag ChatGPT een gedicht of haiku te schrijven over de studie die je wilt gaan doen. Ga na of de inhoud klopt. Is het gedicht goed gelukt? Verbeter dan het gedicht en gebruik hierbij je eigen woorden.

Tips:

Geef ChatGPT een duidelijke opdracht en verfijn je opdracht zo nodig met nieuwe vragen. Je kan de bot bijvoorbeeld naar de geschiedenis of oorsprong van jouw hobby/schoolvak/studie vragen.

Bekijk het resultaat kritisch, check eventuele feiten en herschrijf dan de tekst waar nodig. Gebruik je creativiteit om het gedicht mooier, vloeiender of grappiger te maken.

Vragen:

Klopt de inhoud van de door ChatGPT geschreven tekst?

Heb je ChatGPT om verduidelijking gevraagd of op een andere manier je vraag bijgesteld?

Wat vind je van het taalgebruik van ChatGPT? Zou jij het ook zo formuleren?

7. Herschrijfopdracht met ChatGPT

Opdracht:



A. Vraag ChatGPT een moeilijke tekst over social media voor een jong doelpubliek te herschrijven. Lees het resultaat kritisch en verander de tekst waar nodig.



B. Vraag ChatGPT een moeilijke tekst over de Europese Unie voor een jong doelpubliek te herschrijven. Lees het resultaat kritisch en verander de tekst waar nodig.



C. Vraag ChatGPT een moeilijke tekst over artsen of juristen voor een jong doelpubliek te herschrijven. Lees het resultaat kritisch en verander de tekst waar nodig.

Tips:

Haal een moeilijke tekst uit een vaktijdschrift over het door jou gekozen onderwerp.

Kies een duidelijke leeftijdscategorie en vermeld deze expliciet in je opdracht (prompt) aan ChatGPT.

Beoordeel of je ook zelf de door ChatGPT geschreven tekst beter begrijpt dan de vaktekst die als basis diende van jouw opdracht.

Vragen:

Weet de tool raad met moeilijke woorden en vaktermen?

Heb je genoeg kennis om te controleren of de inhoud van de vereenvoudigde tekst klopt? Zo niet, hoe los je dit op?

Is dit een vaardigheid die je op school kunt gebruiken? Zo ja, bij welk vak?



8. Begrippenopdracht: ChatGPT en je tekstboek

Opdracht:

Vraag ChatGPT twee begrippen uit je tekstboek Nederlands of een ander vak uit te leggen. Komt de uitleg van ChatGPT overeen met de uitleg in je tekstboek? Zo nee, wat is het verschil?

Tip:

Kies begrippen die je lastig vindt. Kijk of je de uitleg van ChatGPT kunt begrijpen.

Vragen:

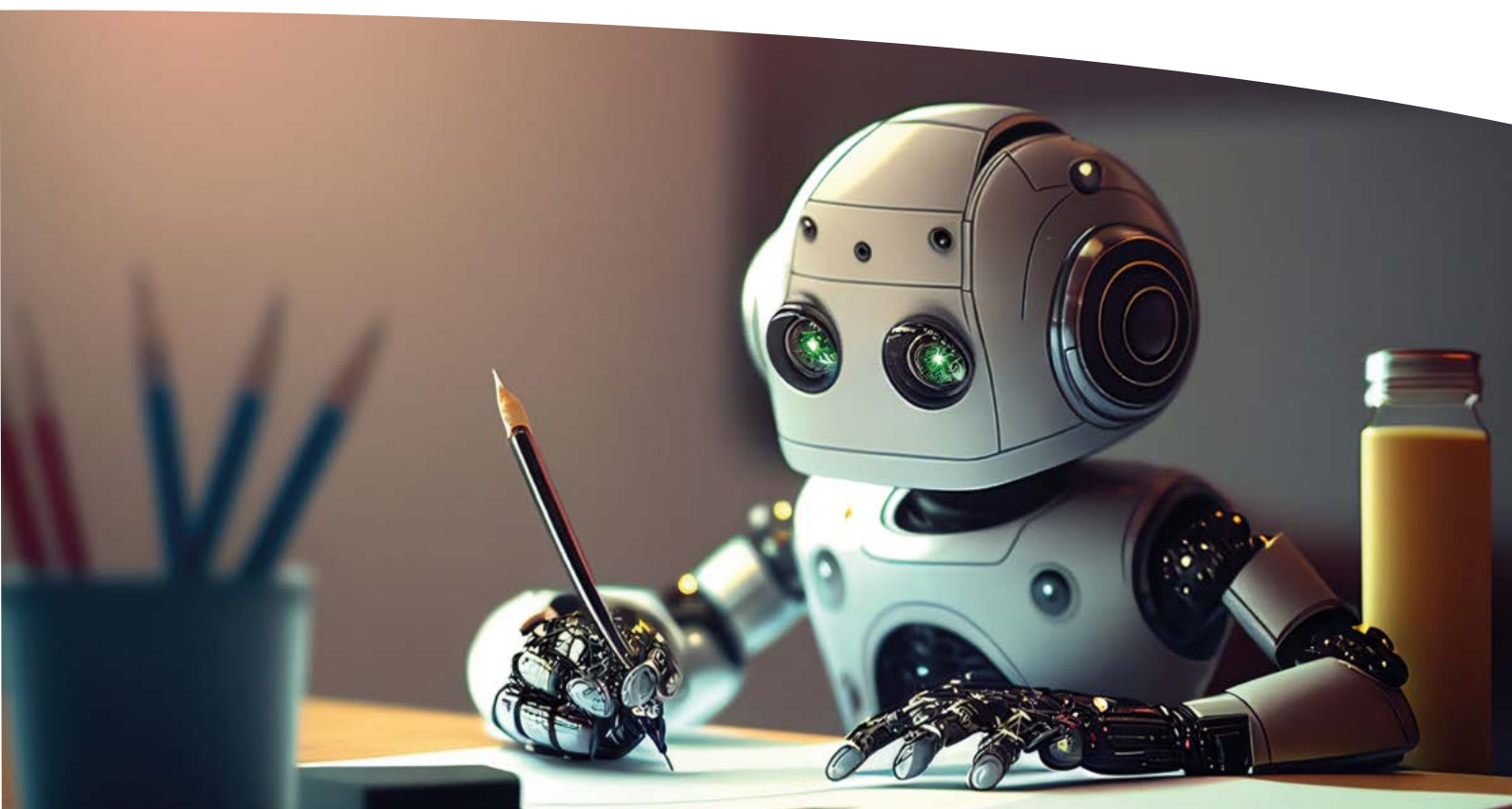
Is de uitleg van ChatGPT duidelijker dan de uitleg van het begrip in je tekstboek? Zo ja, waarom? Zo nee, waarom niet?



9. Pleeg plagiaat met ChatGPT

Opdracht:

Vraag ChatGPT om een tekst van 500 woorden over een onderwerp uit de lesstof te produceren. Neem de tekst goed door. Ga na of je uit de tekst zelf kan afleiden dat die door een computerprogramma is gegenereerd. Waar leid je dit uit af?



3C. Vertaalopdrachten



10. Vertaalopdracht: 'de taaltelefoon'

Opdracht:

Laat een woord of korte zin door een vertaalmachine als Google Translate of DeepL achtereenvolgens naar een aantal verschillende talen vertalen. Vergelijk het verschil tussen het oorspronkelijke woord of de oorspronkelijke zin en de laatste vertaling. Hoe is het eindresultaat? Komt de laatste vertaling nog overeen met het oorspronkelijke woord of de oorspronkelijke zin?



11. Vertaalopdracht: hoe goed 'begrijpt' de vertaalmachine jouw moedertaal/tweede taal?

Opdracht:

Spreek je thuis een andere taal dan Nederlands? Ben je met twee talen opgegroeid?

Test dan hoe goed Google Translate of DeepL die taal 'kent'.

Kies of schrijf zelf een tekst in jouw moedertaal of tweede taal en laat deze door de vertaalmachine naar het Nederlands vertalen. Hoe goed is het resultaat?

Heeft de machine de tekst goed begrepen en in goed Nederlands vertaald?

Verbeter de vertaling en corrigeer eventuele fouten.

Tips:

Kies een tekst op Wikipedia. Wikipedia biedt vaak verschillende taalversies van dezelfde of vergelijkbare teksten aan. Zo kun je de door jouw gekozen tekst ook uit het Engels naar het Nederlands laten vertalen. Meestal vertalen vertaalmachines beter vanuit het Engels dan vanuit andere talen naar het Nederlands. Vergelijk de resultaten. Is er een kwaliteitsverschil tussen de vertaling uit de Engels en de vertaling uit jouw moedertaal of tweede taal?

(Discussie)vragen:

Wat zegt het kwaliteitsverschil tussen de twee vertalingen over de data waarmee de vertaalmachine is gevoed? Kunnen dergelijke kwaliteitsverschillen tot ongelijkheden tussen landen of mensen leiden? Hoe kan dit in de toekomst worden opgelost?

12. Vertaalopdracht: Google Translate en DeepL (opdracht 1) vs ChatGPT (opdracht 2)

Opdracht 1:



Laat Google Translate en DeepL jouw lievelingsliedje vertalen naar het Nederlands of een andere taal die je goed begrijpt. Vergelijk de twee vertalingen en maak op basis hiervan een eindvertaling.



Laat Google Translate en DeepL een tekst uit je tekstboek van het vak Engels vertalen naar het Nederlands of een andere taal die je goed begrijpt.

Vergelijk de twee vertalingen en maak op basis hiervan een eindvertaling.



Laat Google Translate en DeepL een krantenartikel vertalen naar het Nederlands of een andere taal die je goed begrijpt. Vergelijk de twee vertalingen en maak op basis hiervan een eindvertaling.

Tips:



Corrigeer duidelijke taal- of begripsfouten, maar denk ook aan het ritme van je liedje. Klinkt het vertaalde liedje net zo goed als het origineel? Kan je het nog zingen op dezelfde muziek?

Let ook op de vertaling van eigennamen, kijk of er woorden of tekstdelen ontbreken, of de machine nieuwe woorden of vertalingen heeft verzonnen of bepaalde woorden of uitdrukkingen juist te letterlijk heeft vertaald.



Corrigeer duidelijke taal- of begripsfouten. Blijft de kern van de tekst hetzelfde? Komt de boodschap van de tekst/het artikel nog over?

Let ook op de vertaling van eigennamen, kijk of er woorden of tekstdelen ontbreken, of de machine nieuwe woorden of vertalingen heeft verzonnen of bepaalde woorden of uitdrukkingen juist te letterlijk heeft vertaald.

Vragen:

Wat viel je het meest op in de machinevertalingen van Google en DeepL? Waren ze even goed of was een van de machines beter? Waar lag dat aan?

Opdracht 2:



Laat nu ook ChatGPT jouw liedje vertalen.



Laat nu ook ChatGPT de tekst uit je tekstboek Engels vertalen.



Laat nu ook ChatGPT het krantenartikel vertalen.

Vragen:

Wat heb je precies aan ChatGPT gevraagd? Heb je de bot één enkele of verschillende vragen gesteld?

Staan er vooroordelen of rechtstreekse onzin in ChatGPT's vertaling? Wat werkt beter voor jouw liedje/tekst/artikel, een vertaalmachine (Google/DeepL) of ChatGPT?



13. Vertaalopdracht: machinevertaling vs ChatGPT

Opdracht:

Vertaal jouw lievelingsgrap eerst met een vertaalmachine naar keuze en dan met ChatGPT. Pas de vertaling aan zodat de grap goed overkomt.

Tip:

Zorg ervoor dat je de taal waarnaar je vertaalt (redelijk) goed beheerst.

Vragen:

Zijn de vertalingen correct of zelfs goed, en heb je zelf nog iets moeten aanpassen?

Zo ja, wat?

Welke tool geeft de betekenis van jouw grap het beste weer in de door jou gekozen doeltaal?

14. Vertaalopdracht: Jow eigen vertaling vs die van ChatGPT

Opdracht:



Vertaal het gedicht 'Snowball' van Shel Silverstein zelf naar het Nederlands. Laat het gedicht daarna door ChatGPT naar het Nederlands vertalen.



Vertaal het gedicht 'Fire and Ice' van Robert Frost zelf naar het Nederlands. Laat het gedicht daarna door ChatGPT naar het Nederlands vertalen.



Vertaal de eerste tien regels van het gedicht 'The Hill We Climb' van Amanda Gorman zelf naar het Nederlands. Laat het gedicht daarna door ChatGPT naar het Nederlands vertalen.

Tip:

Lees het gedicht voor je gaat vertalen goed door. Wat is de kern van het gedicht/waar gaat het over? Laat het ritme ook goed op je inwerken.

Vragen:

Welke verschillen zie je tussen de vertaling die je zelf gemaakt hebt en de vertaling die ChatGPT heeft gemaakt? Welke vertaling vind je beter?

Heeft de machine rekening gehouden met het ritme van het gedicht?

Geeft de machinevertaling goede ideeën voor jouw eigen vertaling?

3D. Discussieonderwerpen

Over de kwaliteit van de output van de tools, op basis van de praktische oefeningen hierboven:

Wat zijn de zwakke en sterke punten van vertaalmachines/chatbots als ChatGPT?

Waar is 'de machine' goed in en waar gaat 'ie onderuit?

Wat doet het gebruik van een taaltool met jouw creatieve schrijfproces?

Heb je een voorkeur voor een van de programma's?

Over de praktische risico's en beperkingen van gratis online AI-tools:

Wanneer je gratis online tools als ChatGPT en Google Translate gebruikt, wat zijn dan de beperkingen/risico's?

Zou jij voor je volgende schrijfpdracht Frans een gratis vertaalsysteem gebruiken? Of ChatGPT om hulp vragen bij het schrijven van je profielwerkstuk Nederlands? (Hoe) kunnen deze tools je in jouw schoolwerk helpen? Wanneer is er sprake van plagiaat?

Over de toekomst van Generatieve AI in het onderwijs:

Welke mogelijkheden en bedreigingen zien jullie voor Generatieve AI in het onderwijs? Waar kan het gebruik van Generatieve AI toe leiden in het onderwijs? Zullen de schoolvakken of lessen ingrijpend gaan veranderen?

Over de maatschappelijke risico's van Generatieve AI van het gebruik van Generatieve AI:

Wat zijn de kansen en gevaren van Generatieve AI voor de maatschappij? Is er behoefte aan regelgeving en hoe moet deze eruit zien? Hoe moeten de belangrijkste regels eruitzien?

Is het mogelijk om met regelgeving grip te houden op de snelle ontwikkelingen?

Over vervuiling door grote AI-systemen:

De opslag en verwerking van gegevens om AI-tools mee te trainen slurpen energie. Sommige 'datafarms' verbruiken net zoveel energie als een kleine stad. Wegen de maatschappelijk voordelen van de tools tegen de ecologische nadelen op?

Over het overnemen van menselijke werkzaamheden door AI:

Als machines meer creatief en schrijfwerk van de mens gaan overnemen, is er dan een risico dat we het verleren om zelf na te denken en creatief te zijn? Als je het niet meer gewend bent om zelf te schrijven, komt er dan nog wel iets uit je vingers als de computer het niet doet? Merk je nu al een verschil in jouw schrijf- en denkwerk?

Over AI-wetgeving:

Waarom het is zo moeilijk om wetgeving op het gebied van kunstmatige intelligentie vast te stellen. Is dit überhaupt mogelijk of is het een illusie?

Over de bestrijding van desinformatie:

Bekijk de website: [About - EUvsDisinfo](#). Op deze website kun je zien hoe de EU desinformatie - in dit geval specifiek desinformatie uit Rusland - probeert te bestrijden.

Wat vind je van deze website? Hoe maakt de opkomst van AI zulk soort 'fact-checking' moeilijker? Kan de opkomst van AI hier juist ook bij helpen?

Nawoord

Taaltechnologie op basis van AI ontwikkelt zich razendsnel en jongeren maken massaal gebruik van de beschikbare online tools, vaak al in het basisonderwijs. Ook lezen jongeren veel teksten die door AI zijn gegenereerd, vaak zonder dat zij dit doorhebben. Dit gaat gepaard met risico's: teksten kunnen onjuiste informatie, vooroordelen of discriminerende uitingen bevatten. Ook kunnen auteursrechten of privacy worden geschonden en loopt de gebruiker van de tools ook zelf het risico om de controle over de verstrekte informatie te verliezen. Maar AI biedt ook kansen: vertaalmachines produceren doorgaans een prima eerste versie van een vertaling. Programma's die steunen op Generatieve AI kunnen uitkomst bieden bij het schrijven van teksten, teksten samenvatten en analyseren.

Deze module geeft een aanzet tot discussie over het gebruik van AI-tools in de klas. Praktische oefeningen scherpen de vaardigheden aan die jongeren nodig hebben voor een kritisch, (doel)bewust en efficiënt gebruik van de tools, alsook hun taal- en analytische vaardigheden. Deze vaardigheden zullen zij in hun werkende leven goed kunnen gebruiken. Technologie zal in het Westen een oplossing bieden voor de krapte op de arbeidsmarkt en de nieuwe generaties zullen die technologie moeten kunnen gebruiken. In andere woorden, jongeren moeten 'AI-geletterd' zijn. Oude beroepen zullen verdwijnen of van aard veranderen. Daarbij wordt weleens aangenomen dat er minder expertise nodig zal zijn; de technologie zou het werk van de mens overnemen. Wij denken echter dat er meer nodig zal zijn dan een 'expert in the loop'. Naast AI-expertise zullen jongeren in hun beroepsleven ook taal- en communicatievaardigheden hard nodig blijven hebben. Dit wordt hopelijk duidelijk bij het doen van de oefeningen van deze module.

Veel plezier!

Diverse links



- [eTranslation](#), het machinevertaalsysteem van de Europese Commissie, toegankelijk voor Europese overheden, universiteiten, freelancers, kleine en middelgrote bedrijven en ngo's. Om eTranslation te gebruiken, heeft u een EU-account nodig. Via de link kunt u dit doen door op 'Create an account' of 'Account aanmaken' te klikken. Registratie bij eTranslation geeft ook toegang tot [andere taaltools](https://language-tools.ec.europa.eu/) (https://language-tools.ec.europa.eu/)



- [ELIZA](#)



- [Podcast: 'ChatGPT, een hallucinerende babbelaar'](#), van het Instituut voor de Nederlandse Taal



- [Optimisme over AI en machinevertaling](#) in de jaren '50 en '60 van de vorige eeuw (machinevertaling vanaf 2:03)

Links om digitale geletterdheid te bevorderen, met name onder jongeren:



- Persbericht van de Europese Commissie over de bestrijding van desinformatie en de bevordering van de [digitale geletterdheid van jongeren](#)



- **Brochure:** [Richtsnoeren voor leerkrachten en onderwijsactoren inzake het aanpakken van desinformatie en het bevorderen van digitale geletterdheid via onderwijs en opleiding](#) – Bureau voor publicaties van de Europese Unie



- [SMILES-project](#) om mediawijsheid onder jongeren te vergroten, met workshops voor scholen en bibliotheken in Nederland, België en Spanje



- [Education Kit on "Fake News and Disinformation"](#) – VPRO In Europe Schools: interactief onderwijspakket voor het maken van een korte documentaire in de klas

• Bronnen



- [Podcast: ChatGPT, een hallucinerende babbelaar](#) – Instituut voor de Nederlandse Taal.
- 'ChatGPT is omstreden op de universiteit', Joost Ingen-Housz, NRC Handelsblad, 7 juni 2023.
- 'Leve het notitieboekje', Christiaan Weijts, NRC Handelsblad, 22-23 april 2023.
- 'Mag je een chatbot je huiswerk laten doen?', Marc Hijink, NRC Handelsblad, 17-18 december 2022.
- 'De taal is van ons, niet van ChatGPT', Eric Sadin (vertaling Harry Bos), NRC Handelsblad, 18-19 februari 2023.
- 'Gaat Brussel ChatGPT de mond snoeren? De techwereld kijkt gespannen toe', Marc Hijink, Stijn Bronzwaer en Clara van de Wiel, NRC Handelsblad, 13 juni 2023.



- ['AI can help us fight climate change. But it has an energy problem, too'](#), Annette Ekin, Horizon, 12 september 2019.



- ['Generatieve AI, Wietse van Bruggen'](#), Kennisnet.nl, 8 juni 2023.



- ['Common languages used for web content 2023, by share of websites'](#), Ani Petrosyan, Statista, 24 februari 2023.



- ['New Modes of Learning Enabled by AI Chatbots: Three Methods and Assignments'](#), Ethan R. Mollick en Lilach Mollick, SSRN, 13 december 2022.



- ['Using AI to Implement Effective Teaching Strategies in Classrooms: Five Strategies, Including Prompts'](#), Ethan R. Mollick en Lilach Mollick, SSRN, 17 maart 2023.

