

- Doelstellingen  
ET 4,5, 16, 17, 18\*, 19, 20, 31, 31.4, 32
- Lesopbouw  
In deze les maken de leerlingen kennis met het proces van taalverwerking bij computers. Aan de leerlingen willen we vooral tonen dat veel complexe processen hierin een rol spelen, maar dat menselijke input cruciaal is om die taalverwerking te laten slagen. Computers kunnen taal niet op eenzelfde manier interpreteren als mensen en vergeten soms bij hun talige output dat menselijk taalgebruik zelden objectief is. Ze produceren ook vaak taal in een gelimiteerde context. Wanneer mensen taal produceren zijn ze zich bewust van de context waarin ze deze taal produceren (gesprekspartner, omstandigheden, voorgeschiedenis, etc.) zonder dat deze expliciet gesteld moet worden, dit maakt menselijke taal zo complex. Met als gevolg dat computers vaak moeite hebben om deze taal te interpreteren, denk bijvoorbeeld aan de vele verkeerde vertalingen van figuurlijk taalgebruik. Leerlingen bewust maken van die complexiteit en beperkingen van computertaalverwerking is hier het hoofddoel.
- Aanbevelingen  
Deze les is geschikt voor leerlingen met een bepaalde interesse in wetenschappen, technologie en ICT.  
Als je een leerlingengroep hebt die thuis is in de wereld van AI bestaan er veel mogelijkheden om deze les verder uit te breiden. Bijvoorbeeld door dieper in te gaan op de specifieke programma's en processen die schuilgaan achter NLP en ML. Of door leerlingen zelf een bot te laten programmeren die talige taken kan uitvoeren. Dit kan gedaan worden in een vakoverschrijdend project met ICT en wiskunde bijvoorbeeld. De les zoals ze nu opgebouwd is, richt zich vooral tot leerlingen met een eerder beperkte voorkennis. Ze worden aan de hand van de opdrachten door het proces van computertaalverwerking geleid en hebben op het einde van de lessenreeks voldoende basiskennis over dit thema verworven.

### Oefening 1

- 1a Persoonlijk antwoord van de leerlingen
- 1b Tekst C. In deze tekst staat veel misinformatie, de tekst heeft een weinig logische opbouw en sommige zinsconstructies voelen raar aan. Het is wel degelijk bewezen waarom bananen krom zijn in tegenstelling tot wat in de tekst beweerd wordt (ze groeien namelijk tegen de zwaartekracht in). Bananen groeien aan een boomachtige plant en deze heet de Musa Basjoo. Ze hebben geen houtachtige stam zoals normale bomen, maar hebben een stam die bestaat uit de basis van de enorme bladstengels. 'Een ongezonde boom die zijn eigen gewicht niet kan dragen' is een vreemde zin, kan potentieel verwijzen naar het feit dat de trossen naar beneden gaan hangen door de zwaartekracht. 'of zelfs gebogen bananen', krom en gebogen zijn synoniemen van elkaar.

1c Deze tekst werd geschreven door een AI-bot. Maar, ook tekst A is geschreven door een AI-bot.

\* Je kan met de leerlingen een klasdiscussie aangaan waarom deze twee teksten zo verschillend zijn terwijl ze allebei door een bot werden geschreven.

### Oefening 2

2a Computers hebben enorm veel input nodig voordat ze een patroon, een gezicht, een dier, etc. kunnen herkennen. Mensen in tegenstelling hebben vaak maar 1 voorbeeld nodig. Kinderen zien voor de eerste keer een krokodil in de dierentuin en onthouden voor altijd hoe een krokodil eruitziet. Zelfs al komen ze in aanraking met krokodillen die afwijken van die eerste referentie.

2b Een computer moet alle verschillende versies aangeleerd worden voor het met zekerheid kan zeggen dat hetgeen het ziet overeenkomt met hetgeen het al kent. Zoals met het tekenen van de cijfers. Als een computer geen exacte match kan vinden in zijn geheugen dan gaat het er van uit dat het dat dier/cijfer/gezicht/voorwerp nog niet eerder gezien heeft.

2c Computers zullen veel talige input nodig hebben voor ze zelf taal kunnen produceren, en voordat ze alle menselijke taaluitingen als zodanig kunnen herkennen.

### Oefening 3

\* Voor deze oefening kan je als leerkracht hen de volgende websites aanreiken:

<https://datascience.eu/nl/natuurlijke-taalverwerking/natuurlijke-taalverwerking-nlp/>

<https://www.godisageek.com/2022/01/how-do-computers-learn-to-understand-natural-language/>

<https://www.scriptorij.be/blog/verschil-machine-learning-deep-learning-en-ai>

<https://trendskout.com/nl/general/ai-vs-machine-learning-vs-deep-learning/#:~:text=Machine%20Learning%20is%20een%20onderdeel,waarbij%20algoritmes%20zichzelf%20slimmer%20maken.>

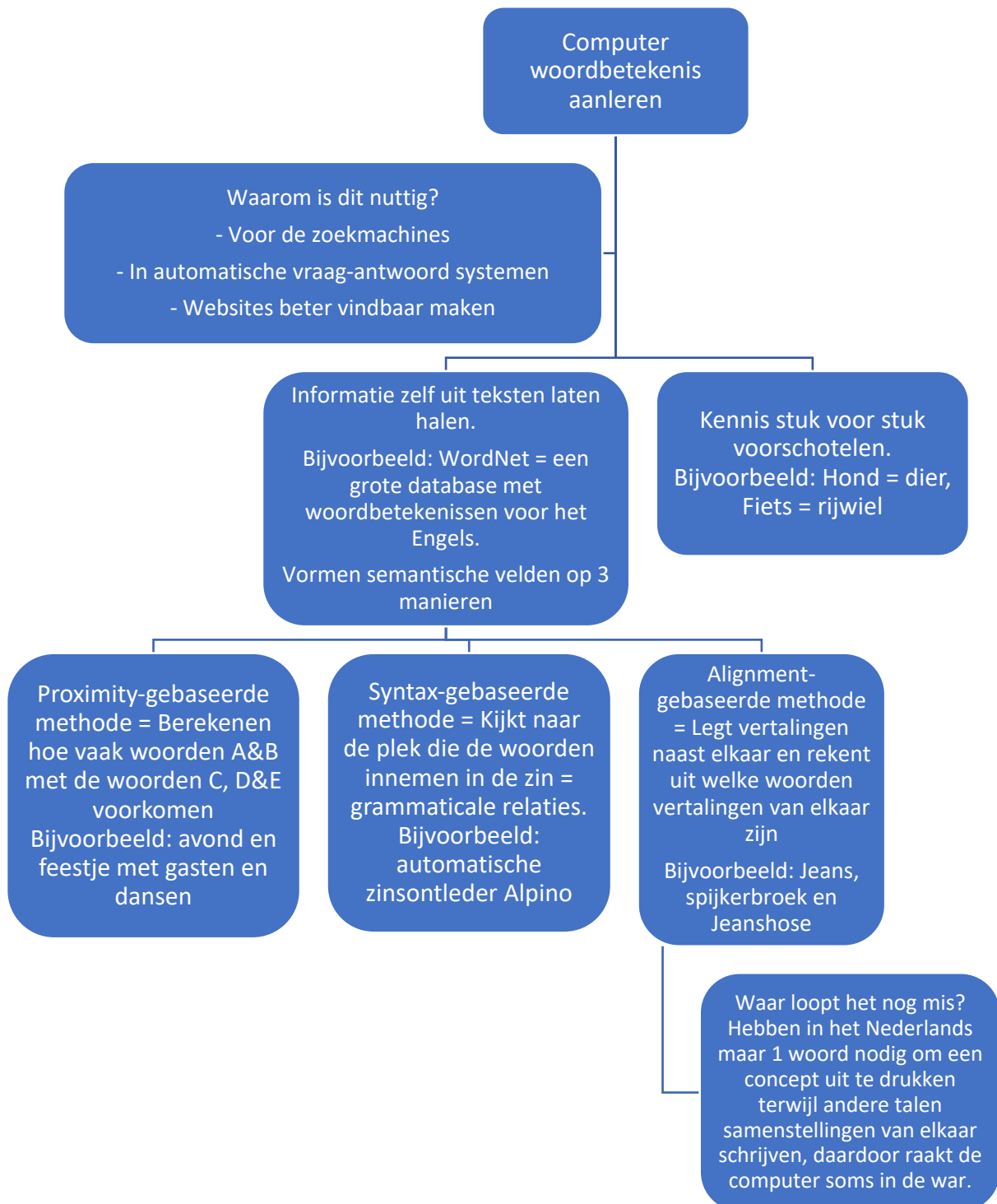
\* Differentiatie: bij een leerlingengroep met meer gevorderde onderzoeksvaardigheden kan je ervoor kiezen om deze vooraf vastgelegde bronnen niet aan te bieden en hen zelfstandig naar betrouwbare bronnen te laten zoeken. Lopen ze vast, dan kan je ervoor kiezen om hen verder op weg te helpen met deze bronnen.

\* Verdieping: bij deze opdracht kan je ervoor kiezen om dieper in te gaan op de achterliggende processen bij ML en NLP (vraag b). bv. lemmatisering, part-of-speech tagging, Named Entity Recognition (NER), Kunstmatige neurale netwerken (ANN), etc.

		Proces 1	Proces 2
a.	Naam proces?	Machinaal leren/machine learning (ML)	Natuurlijke taalverwerking/Natural Language Processing (NLP)
b.	Wat houdt het proces in?	Leert de machine zelf op basis van data die hij inleest. Door één neural netwerk vast te leggen en daar vervolgens continu trainingsdata door te sturen om de parameters van het netwerk te verbeteren. Na enorm veel data herkent hij zo patronen en gelijkenissen. De machine leert dus op basis van alle data die hem wordt voorgeschoteld en zoekt daarin patronen.	Richt zich op de interactie tussen computers en mensen. Bedenken een gemeenschappelijke taal die bekend staat als een natuurlijke taal. Ze kunnen menselijke taal lezen, interpreteren en begrijpen in een proces dat doorgaans 7 stappen omvat. (De technische details moeten de leerlingen niet kennen)
c.	Welke beperkingen zijn er?	Je moet veel trainingsdata aanbieden alvorens de machine kan leren.	Complexe aard van menselijke communicatie. Niet makkelijk om de context van de gediceerde informatie te begrijpen. Natuurlijke talen zijn gebaseerd op dubbelzinnige en onnauwkeurige kenmerken.
d.	Voorbeeld?	Amazon: gebruikers automatisch producten	Hulpprogramma's zoals Siri en Alexa. Vertalingstools zoals

		aanbeveelt op basis van hun eerdere aankopen. Netflix: abonnees series en films voorstelt op basis van eerder kijkgedrag.	Google Translate Programma's die grammatica controleren.
e.	Verhouding tot menselijke taalverwerving?	Kinderen hebben ook bepaalde vorm van input nodig om taal te kunnen aanleren, maar wel niet zo uitgebreid als een computer.	Leunt het dichts aan bij menselijke taalverwerving door de interactie-gerelateerde aanpak.

## Oefening 4



Oefening 5

## 5a Persoonlijk antwoord leerlingen

Vertalingen zijn een correcte weergave van de letterlijke betekenis van de woorden, maar de figuurlijke betekenis kan de computer niet achterhalen

Zoals besproken in oefening 4, heeft de computer moeite met het begrijpen van de context van dit soort taaluitingen. Zowel de culturele context als de lokale context speelt hierbij een rol. Daarbovenop komen de vele nuances en ambiguïteiten die gesproken en geschreven tekst kleur geven. Taal is heel erg vatbaar voor interpretatie.  
<https://vvn.ugent.be/blog/taalles-voor-computers/>

\* Differentiatie: voor leerlingen die minder kennis hebben van spreekwoorden en gezegden kan je altijd zelf een aantal spreekwoorden selecteren.

\* Differentiatie: de taal waarin ze de spreekwoorden moeten vertalen werd niet expliciet vermeld in de opdracht zodat dit aangepast kan worden naar de sterktes/interesse van de leerlingen/klasgroep. Bij leerlingen met een niet-Nederlandse thuistaal kan je aan hen ook vragen om spreekwoorden uit hun moedertaal te gebruiken en die te vertalen naar het Nederlands (of vice versa).

\* Uitbreiding: bij de bron van deze oefening staat nog een interessant filmpje over de mogelijkheid van het creëren van een universele vertaler die kan dienen als uitbreiding bij de les. → <https://www.yuqo.nl/hoe-vertalen-computers-menselijke-taal/>

## 5b Bij de command gegeven in de Standaardtaal zouden de leerlingen normaalgezien de juiste respons moeten krijgen.

Het merendeel van de chatbots en spraakassistenten zijn getraind op het herkennen van taaluitingen in de opgegeven Standaardtaal. De corpussen die aan de basis liggen van de training van deze AI, zijn corpussen die voor het grootste deel gebaseerd zijn op Nederlands Standaardtaal.

## 5c Grote kans dat de opdracht nu verkeerd begrepen wordt

Je spreekt onduidelijker, gebruikt woordenschat die afwijkt van de Standaardtaal, de bot interpreteert de klanken verkeerd, etc.

Oefening 6

## 6a Proximity-gebaseerde methode

## 6b Kiest de meest voorkomende woorden eruit en plaatst woorden samen die vaak in elkaars buurt voorkomen.

## 6c Syntax-gebaseerde methode. Je krijgt veel ongrammaticale zinnen dus de bot snapt niet dat sommige woorden of zinsdelen niet na elkaar kunnen geplaatst worden. Of bepaalde constructies staan op de verkeerde plaats in de zin bv. He is not going to be asking to it for the world.

## Oefening 7

- 7a Ongecontroleerd machinaal leren in combinatie met waarschijnlijk te weinig input. Het aantal batmanfilms en -series zijn beperkt tot ongeveer 50 uur aan beeldmateriaal, wat betekent dat de AI-bot dit 20 maal moest bekijken. Waarschijnlijk ook problemen met context bv. give birth to Robin/flips Alfred like a coin en zelfstandige naamwoorden verwarren met eigennamen bv. Goth ham (Gotham = stad, the Pinguin = superschurk die Batman bevecht). Is gebaseerd op spraak, sommige acteurs hebben misschien een afwijkende uitspraak ten opzichte van de Engelse Standaardtaal (RP) waardoor de bot sommige zinnen verkeerd interpreteert.