

Scholieren met dyslexie van de derde graad middelbaar onderwijs ondersteunen bij het lezen van wetenschappelijke papers via tekstvereenvoudiging.

De opbouw van een prototype en een vergelijkende studie.

Dylan Cluyse.

Scriptie voorgedragen tot het bekomen van de graad van
Professionele bachelor in de toegepaste informatica

Promotor: Mevr. L. De Mol

Co-promotor: J. Decorte; J. Van Damme; M. Dhondt

Academiejaar: 2022–2023

Eerste examenperiode

Departement IT en Digitale Innovatie .

**HO
GENT**

Woord vooraf

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Samenvatting

Ingewikkelde woordenschat en zinsbouw hinderen scholieren met dyslexie in het derde graad middelbaar onderwijs bij het lezen van wetenschappelijke artikelen. Adaptieve tekstvereenvoudiging helpt deze scholieren bij hun lees- en verwerkingsnelheid. Daarnaast kan artificiële intelligentie (AI) dit proces automatiseren om de werkdruk bij leraren en scholieren te verminderen. Dit onderzoek achterhaalt met welke technologische en logopedische aspecten AI-ontwikkelaars rekening moeten houden bij de ontwikkeling van een AI-toepassing voor adaptieve en geautomatiseerde tekstvereenvoudiging. Hiervoor is de volgende onderzoeksvraag opgesteld: "Hoe kan een wetenschappelijk artikel automatisch worden vereenvoudigd, gericht op de unieke noden van scholieren met dyslexie in het derde graad middelbaar onderwijs?". Een vergelijkende studie beantwoordt deze onderzoeksvraag en is uitgevoerd met bestaande toepassingen en een prototype voor adaptieve en geautomatiseerde tekstvereenvoudiging. Uit de vergelijkende studie blijkt dat toepassingen om wetenschappelijke artikelen te vereenvoudigen, gemaakt zijn voor een centrale doelgroep en geen rekening houden met de unieke noden van een scholier met dyslexie in het derde graad middelbaar onderwijs. Adaptieve software voor geautomatiseerde tekstvereenvoudiging is mogelijk, maar ontwikkelaars moeten meer inzetten op de unieke noden van deze scholieren.

Inhoudsopgave

Lijst van figuren

viii

1	Inleiding	1
1.1	Probleemstelling	2
1.2	Onderzoeksvraag	3
1.3	Onderzoeksdoelstelling	4
1.4	Opzet van deze bachelorproef	4
2	Stand van zaken	6
2.1	Onderzoeken rond dyslexie	6
2.1.1	Centraal zicht op dyslexie	6
2.1.2	Mogelijke drempels voor mensen met fonologische dyslexie.	6
2.1.3	Bewezen effecten van tekstvereenvoudiging -en aanpassing bij scholieren met dyslexie.	7
2.1.4	Conclusie	9
2.2	Wetenschappelijke artikelen	10
2.2.1	Trends rond wetenschappelijke artikelen	10
2.2.2	Woordenschat en vakjargon	12
2.2.3	Aanpak voor het lezen van wetenschappelijke artikelen	12
2.2.4	Conclusie	13
2.3	Tekstvereenvoudiging	13
2.3.1	Manuele tekstvereenvoudiging	14
2.3.2	Natural Language Processing	15
2.3.3	Prompt engineering	17
2.4	De verschillende soorten tekstvereenvoudiging	18
2.4.1	Lexicale vereenvoudiging	18
2.4.2	Syntactische vereenvoudiging	19
2.4.3	Tekstvereenvoudiging automatiseren	20
2.5	Samenvatten	20
2.5.1	Extraherend samenvatten	20
2.5.2	Abstraherend samenvatten	22
2.5.3	Hybride samenvatten	22
2.5.4	Evaluatie	22
2.5.5	Tekstvereenvoudigingstechnieken voor scholieren met dyslexie.	23
2.5.6	Conclusie	24

2.6	Valkuilen en uitdagingen voor AI-ontwikkelaars bij tekstvereenvoudiging met AI	24
2.6.1	Uitdagingen voor softwarebedrijven	25
2.6.2	Ambigüiteit, synoniemen en homoniemen	25
2.6.3	Paternalisme en ethische overwegingen	25
2.6.4	Valkuilen bij prompt engineering	26
2.6.5	Evaluatie en interpretatie	26
2.7	Beschikbare software voor tekstvereenvoudiging	26
2.7.1	Momenteel ingezet in het onderwijs	26
2.7.2	Proof-of-concepts en online webapplicaties	27
2.8	GPT-3	28
2.8.1	Bing AI	32
2.8.2	Huggingface en taalmodellen via API	34
2.8.3	Conclusie	34
2.9	Conclusie	34
3	Methodologie	36
4	Requirementsanalyse	37
4.1	Tekstanalyse	37
4.2	Lexicale vereenvoudiging	38
4.3	Syntactische vereenvoudiging	39
4.4	Samenvatten	39
4.5	Personalisatie en verdere functionaliteiten	40
4.6	Voor ontwikkelaars	40
4.7	Requirements	40
4.8	Conclusie	41
5	Prototype voor tekstvereenvoudiging	42
5.1	Opbouw van een prototype	42
5.2	Tekstvereenvoudiging met API	43
5.2.1	Annotaties van woordenschat	43
5.2.2	Samenvatting	43
5.3	Tekstinhoud extraheren en uitschrijven naar PDF/DOCX	44
5.4	Personaliseerbaarheid aanreiken	44
5.5	Lokale opzetomgeving	44
5.5.1	Taalmodellen	45
5.6	Conclusie	45
6	Vergelijkende studie	46
6.1	Vergelijking met referentieteksten	46
6.1.1	Methodologie	46

6.2	Resultaten.	47
6.2.1	Lexicale vereenvoudiging	47
6.2.2	Syntactische vereenvoudiging	48
6.2.3	Samenvatten	48
6.3	Vergelijking zonder referentieteksten.	49
6.4	Conclusie	49
7	Conclusie	50
A	Onderzoeksvoorstel	54
A.1	Introductie	54
A.2	State-of-the-art	56
A.2.1	Tekstvereenvoudiging	56
A.2.2	Noden van scholieren met dyslexie	56
A.2.3	Huidige toepassingen.	58
A.2.4	Ontwikkelen met AI	58
A.3	Methodologie	59
A.4	Verwacht resultaat, conclusie	61
	Bibliografie	62

Lijst van figuren

1.1	Figuur van De Meyer e.a. (2019). Uit een enquête blijkt dat Vlaamse 15-jarigen zeer negatief zijn over leesplezier. Bijna de helft van de scholieren beschouwt begrijpend lezen als tijdverspilling en slechts 17% noemt lezen een van hun favoriete hobby's.	3
2.1	Afbeelding van DuBay (2004)	7
2.2	Afbeelding van Rello, Baeza-Yates, Dempere-Marco e.a. (2013). Volgens de richting van de pijl wordt de ideale situatie benaderd, gekenmerkt door doelwaarden. Deze waarden worden bereikt door mensen zonder dyslexie onder optimale omstandigheden. Het gebruik van vaak voorkomende woorden vermindert de decodeertijd en verbetert de leesbaarheid voor mensen met dyslexie.	8
2.3	Afbeelding uit Rello en Baeza-Yates (2013). Verticaal wordt de gemiddelde mening van de bevroegden weergegeven. Horizontaal worden de lettertypes gerangschikt op gemiddelde leestijd van alle bevroegden. Dit onderzoek wijst uit dat Arial, CMU, Helvetica en Times de populaire keuzes zijn. Arial en CMU behoren hierbij tot de drie best scorende lettertypes rond gemiddeld leestempo.	9
2.4	Afbeelding uit Plavén-Sigray e.a. (2017). Links wordt de evolutie per FRE-score getoond. Hoe hoger de score, hoe hoger de gemiddelde complexiteit van een tekst. Rechts wordt de evolutie volgens de NDC-score getoond. Hoe hoger de score, hoe lager de gemiddelde complexiteit van een tekst. Het onderzoek schat dat nu een kwart van alle wetenschappelijke artikelen gebruik maken van Engels op het niveau van een masterstudent, ofwel een FRE onder nul.	11
2.5	Afbeelding van Murdos e.a. (2014) Volgens deze grafiek scoren de wetenschappelijke artikels rond fysica gemiddeld het best op de FRE-score. Al scoren de wetenschappelijke artikels rond microbiologie gemiddeld het zwakst op de FRE-score, ze scoren gemiddeld beter op de FOG-score.	12
2.7	Afbeelding uit McFarland (2023). Een illustratie over de werking en begeleiding van prompt engineering bij een taalmodel.	17
2.6	Voorbeeld van PoS-labeling op de Engelstalige zin "She sells seashells on the seashore". Afbeelding van Bilici (2021)	17

2.8	Voorbeeld van tekstvereenvoudiging. Oorspronkelijke tekst uit Historia 5 bron toe te voegen	18
2.9	Afbeelding van Althunayyan en Azmi (2021). Deze pipeline wordt in meerdere onderzoeken rond lexicale vereenvoudiging toegepast, zoals Bingel e.a. (2018), Bulté e.a. (2018) en Paetzold en Specia (2016)	19
2.10	Afbeelding uit Bulté e.a. (2018). Deze pipeline omvat de stappen die de toepassing aflegt.	24
2.11	Afbeelding uit Bulté e.a. (2018). TODO	24
2.12	Schermafbeelding van SciSpace	27
2.13	Afbeelding van Gooding 2022. De invoertekst is een paragraaf uit een niet-vermeld boek van de Russische schrijver Dostoevsky. Het resultaat van de meegegeven prompt is een transformatie dat iedere vorm van vooraf aangehaalde vereenvoudiging weergeeft. Lexicale, conceptuele en syntactische vereenvoudiging worden op de invoertekst toegepast.	28
2.14	Afbeelding van Gooding 2022. Gooding haalt verder aan dat modellen zoals ChatGPT op twee vlakken de leesbaarheid van een tekst kan bevorderen. Allereerst door het verlenen van verschillende mogelijke versies van een vereenvoudigingstaak.	29
2.15	Afbeelding van Gooding 2022.	29
2.16	Afbeelding van Binz en Schulz (2023). Dit toont de <i>mean regret</i> aan tussen de vier engines en de menselijke antwoorden.	30
2.17	Afbeelding van Simon (2021). De evolutie van pre-trained taalmodellen wordt hier weergegeven tot eind 2022. De performantie van de modellen ten opzichte van de grootte volgt een lineaire functie.	31
2.18	Afbeelding van Ribas (2023).	33
2.19	In deze afbeelding wordt er een online wetenschappelijk artikel meegegeven. Er wordt geen titel of onderwerp meegegeven, maar de Bing AI chatbot is in staat om een abstraherende samenvatting te maken van het artikel. Daarna geeft de chatbot verder uitleg over een bepaald onderwerp en geeft het extra referenties mee.	33
A.1	(Readable, 2021)	60

1

Inleiding

Het middelbaar onderwijs staat op springen. Dagelijks sneuvelen leerkrachten en scholieren van het middelbaar onderwijs onder de te harde werkdruk (Minnen e.a., 2018). Lezen is ingebed in ons dagelijks leven in de vorm van Nederlandstalige nieuwsartikelen tot de ondertiteling van televisieseries. Mensen van eender welke leeftijdsgroep kunnen lezen niet ontsnappen en dit moet van jongs af aan geprikkeld worden (Daoud, 2023).

Vakmiddelen -en didactiek in het onderwijs

Lerarenopleidingen benadrukken nu het gebruik van verschillende bronnen in lessen. De leesgraad van deze bronnen verandert echter niet, want de noodzaak aan bronnen met diverse leesgraden is bedoeld om scholieren uit te kunnen dagen (Surma e.a., 2019). Het Amerikaanse onderwijs stampte C.R.E.A.T.E.¹ uit de grond. Dit initiatief zet scholieren tussen 12 en 18 jaar aan om wetenschappelijke artikelen te lezen in plaats van enkel boeken. Scholieren komen zo in direct contact met wetenschappelijk onderzoek. Ze begrijpen hoe wetenschappers experimenten uitvoeren, plannen en resultaten analyseren en interpreteren. Vlaamse STEM-leerkrachten in de derde graad middelbaar onderwijs moeten volgens het M-decreet en de leerplannen van zowel het katholiek² als het gemeenschapsonderwijs³ hun theorielessen op een toegankelijke manier aanbieden, zodat alle scholieren worden meegenomen in het verhaal.

Kunstmatige intelligentie in Vlaanderen

België is met een jaarlijks budget van 32 miljoen een pionier op het gebied van artificiële intelligentie (AI) op de werkvloer (Crevits, 2022). Zo stampte de Vlaamse overheid verschillende AI-projecten uit de grond, om Vlaamse AI-ontwikkelingen te

¹<https://teachcreate.org/>

²<https://pro.katholiekonderwijs.vlaanderen/basisoptie-stem/ondersteunend-materiaal>

³<https://g-o.be/stem/>

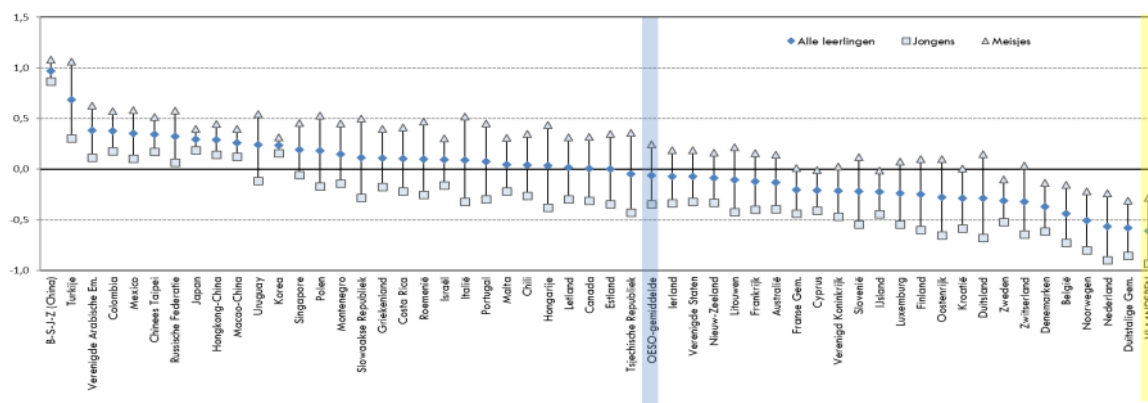
ondersteunen en om AI-softwarebedrijven te inspireren. Het amai!-project⁴ brengt AI-softwarebedrijven uit diverse domeinen samen en leidt tot het ontstaan van AI-toepassingen die processen automatiseren en de werkdruk verminderen, zoals real-time ondertiteling in de klas en een taalassistent voor leerkrachten in meertalige klasgroepen.

1.1. Probleemstelling

Vlaamse en Nederlandse onderzoeken van Desoete (2017) en Wentink e.a. (2008) wijzen uit dat gemiddeld 4% van de Vlaamse en Nederlandse bevolking de diagnose van dyslexie heeft. Het aantal scholieren met dyslexie in het lager en middelbaar onderwijs wereldwijd loopt op tot 15% ingeschat (Bonte, 2020; van der Meer, 2022). Lissens e.a. (2020) benadrukt dat de impact van leerstoornissen niet stopt na het middelbaar onderwijs. Scholieren met dyslexie in het middelbaar onderwijs kampen met unieke uitdagingen, waaronder een moeizame en stroeve automatisering bij het lezen en spellen. Deze doelgroep kunnen rekenen op ondersteuning van coaches en beschikbare hulpmiddelen om hun leesachterstand te beperken. Het leerplan voor STEM-vakken stimuleert het gebruik van wetenschappelijke artikelen, maar houdt niet altijd rekening met de complexe leesgraad. De ingewikkelde woordenschat en syntax in wetenschappelijke artikelen kunnen een hinderenis vormen voor de begrijpelijkheid van een tekst, waardoor scholieren met dyslexie de kerninhoud moeilijk kunnen doorgronden. Teksten vereenvoudigen, zonder de kern- en bijzaken te verliezen, kan een oplossing bieden.

Leesvaardigheid is cruciaal voor succes op school en in het werkveld. Mensen met dyslexie kunnen problemen hebben met spelling, wat kan leiden tot onzekerheid en stress. Vooroordelen zijn nog steeds een probleem en kunnen leiden tot stigmatisering. Onderzoek toont echter aan dat mensen met dyslexie doorzettingsvermogen hebben en goede probleemoplossers zijn (Bonte, 2020; Ghesquière, 2018; Lissens e.a., 2020). Ondersteuning op scholen en werkplekken is belangrijk, omdat dyslexie uitdagingen kan veroorzaken bij het betreden van de arbeidsmarkt. Onderzoek richt zich vooral op kinderen, maar ook jongvolwassenen en ouderen hebben ondersteuning nodig. Een gepersonaliseerde analyse is nodig om specifieke begeleiding te kunnen bieden bij begrijpend lezen (Lissens e.a., 2020; Van Vreckem & Desoete, 2015). De driejaarlijkse PISA-test meet de wiskundige en wetenschappelijke geletterdheid van 15-jarigen in ongeveer 79 geïndustrialiseerde landen. In 2018 namen 4822 Vlaamse scholieren van vijftien jaar deel aan deze test. Hoewel leer- en leerstoornissen niet worden meegenomen in de test, geven de resultaten een idee van de leesvaardigheid en wetenschappelijke geletterdheid van scholieren van die leeftijd.

⁴<https://amai.vlaanderen/>



Figuur (1.1)

Figuur van De Meyer e.a. (2019). Uit een enquête blijkt dat Vlaamse 15-jarigen zeer negatief zijn over leesplezier. Bijna de helft van de scholieren beschouwt begrijpend lezen als tijdverspilling en slechts 17% noemt lezen een van hun favoriete hobby's.

Wetenschappelijke artikelen vereenvoudigen kan tijd en energie van docenten in de derde graad middelbaar onderwijs opsorpen. Het Vlaamse middelbaar onderwijs staat onder druk en docenten hebben moeite om met deze werkdruk boven water te blijven. Daarom is er nood aan software die wetenschappelijke artikelen automatisch kan vereenvoudigen, specifiek gericht op de unieke noden van scholieren met dyslexie. Een dergelijke toepassing kan scholieren met dyslexie in de derde graad middelbare onderwijs ondersteunen bij het lezen van een wetenschappelijk artikel en kan het routinematige werk van STEM-docenten verminderen.

1.2. Onderzoeksvraag

De volgende onderzoeksvraag is opgesteld: "Hoe kan een wetenschappelijke artikel automatisch vereenvoudigd worden, gericht op de unieke noden van scholieren met dyslexie in de derde graad middelbaar onderwijs?". Daarnaast worden de volgende deelvragen beantwoord.

- Welke aanpakken zijn er voor geautomatiseerde tekstvereenvoudiging? Aansluitende vraag: "Hoe worden teksten handmatig vereenvoudigd voor scholieren met dyslexie?"
- Welke specifieke noden hebben scholieren met dyslexie van de derde graad middelbaar onderwijs bij het begrijpen van complexere teksten?
- Wat zijn de specifieke kenmerken van wetenschappelijke artikelen?
- Met welke valkuilen bij taalverwerking met AI moeten ontwikkelaars rekening houden?

- Welke toepassingen, tools en modellen zijn er beschikbaar om Nederlandse geautomatiseerde tekstvereenvoudiging met AI mogelijk te maken?
- Welke functies ontbreken AI-toepassingen om geautomatiseerde tekstvereenvoudiging mogelijk te maken voor scholieren met dyslexie in de derde graad middelbaar onderwijs? Aansluitende vraag: "Welke manuele methoden voor tekstvereenvoudiging komen niet in deze tools voor?"

1.3. Onderzoeksdoelstelling

Het doel van dit onderzoek is om te achterhalen met welke technologische en logopedische aspecten AI-ontwikkelaars rekening moeten houden bij de ontwikkeling van een adaptieve AI-toepassing voor geautomatiseerde tekstvereenvoudiging. Het resultaat van dit onderzoek is een prototype voor een toepassing die de tekstinhoud van een wetenschappelijke paper zal vereenvoudigen, naargelang de specifieke noden van een scholier met dyslexie in de derde graad middelbaar onderwijs. Het prototype houdt rekening met de transformatie van het bronbestand, bijvoorbeeld een PDF of een afbeelding, naar de tekstinhoud. Hiervoor bestaan er kant-en-klare pakketten dat de transformaties al voor de ontwikkelaar doen. De invoer van dit prototype is een wetenschappelijk artikel van minstens 500 woorden lang.

1.4. Opzet van deze bachelorproef

De rest van deze bachelorproef is als volgt opgebouwd:

In Hoofdstuk 2 wordt een overzicht gegeven van de stand van zaken binnen het onderzoeksdomein, op basis van een literatuurstudie.

In Hoofdstuk 3 wordt de methodologie toegelicht en worden de gebruikte onderzoekstechnieken besproken om een antwoord te kunnen formuleren op de onderzoeksvragen.

- Welke aanpakken zijn er voor geautomatiseerde tekstvereenvoudiging?
 - Aansluitende vraag: "Hoe worden teksten handmatig vereenvoudigd voor scholieren met dyslexie?"
- Welke specifieke noden hebben scholieren met dyslexie van de derde graad middelbaar onderwijs bij het begrijpen van complexere teksten?
- Wat zijn de specifieke kenmerken van wetenschappelijke artikelen?
- Met welke valkuilen bij taalverwerking met AI moeten ontwikkelaars rekening houden?
- Welke toepassingen, tools en modellen zijn er beschikbaar om Nederlandstalige geautomatiseerde tekstvereenvoudiging met AI mogelijk te maken?

- Hoe kan een intuïtieve lokale webtoepassing worden ontwikkeld die zowel scholieren met dyslexie als docenten helpt bij het vereenvoudigen van wetenschappelijke artikelen met behoud van semantiek, jargon en zinsstructuren.
- Welke functies ontbreken AI-toepassingen om geautomatiseerde én adaptieve tekstvereenvoudiging mogelijk te maken voor scholieren met dyslexie in de derde graad middelbaar onderwijs?
 - "Welke manuele methoden voor tekstvereenvoudiging komen niet in deze tools voor?"

In Hoofdstuk ?? wordt de requirementsanalyses gegeven om een antwoord op de volgende onderzoeksvraag te geven: Welke functies ontbreken AI-toepassingen om geautomatiseerde én adaptieve tekstvereenvoudiging mogelijk te maken voor scholieren met dyslexie in de derde graad middelbaar onderwijs? Aansluitende vraag: "Welke manuele methoden voor tekstvereenvoudiging komen niet in deze tools voor?"

In Hoofdstuk ?? wordt een stappenplan voor de ontwikkeling van een prototype voor tekstvereenvoudiging gegeven.

In Hoofdstuk ?? wordt de shortlist van tools en het ontwikkelde prototype vergeleken op basis van bestaande leesgraadmetrieken.

In Hoofdstuk 7, tenslotte, wordt de conclusie gegeven en een antwoord geformuleerd op de onderzoeksvragen. Daarbij wordt ook een aanzet gegeven voor toekomstig onderzoek binnen dit domein.

2

Stand van zaken

2.1. Onderzoeken rond dyslexie

Lezen is een essentieel onderdeel van ons dagelijks leven en speelt een belangrijke rol in onze communicatie en begrip. Dyslexie kan het functioneren in het dagelijks leven belemmeren. Het begrijpen van de noden en hindernissen voor een scholier met dyslexie is van belang om deze doelgroep te ondersteunen en hun kwaliteit van lezen te verbeteren. Deze sectie zal ingaan op de unieke noden en bespreken hoe mensen met dyslexie kunnen worden geholpen bij het lezen. De volgende onderzoeksvraag wordt in deze sectie beantwoordt: "Welke specifieke noden hebben scholieren met dyslexie van de derde graad middelbaar onderwijs bij het begrijpen van complexere teksten?"

2.1.1. Centraal zicht op dyslexie

Lezen is onnatuurlijk en volgens de geschiedenis van de mens een recent begrip. Pas 5000 jaar geleden werd de geschreven taal bedacht. Mensen worden niet met leesvaardigheid geboren, maar leren dit zelf aan en daarvoor moet het brein heringericht worden (Bonte, 2020; van der Meer, 2022). Dyslexie betekent letterlijk 'beperkt lezen'. Het voorlezen kan radend, langzaam en letter-voor-letter verlopen. Goede woordenschat ontwikkeling of vaak voorlezen is een beschermende factor tegen dyslexie Bonte (2020) en Vellutino e.a. (2004). Onderzoeken halen drie verschillende types van dyslexie aan, namelijk fonologische dyslexie, *surface dyslexia* en *deep dyslexia*. Dezelfde onderzoeken wijzen erop dat een overlap van kenmerken over de drie types heen mogelijk is (Rello e.a., 2012; Vellutino e.a., 2004).

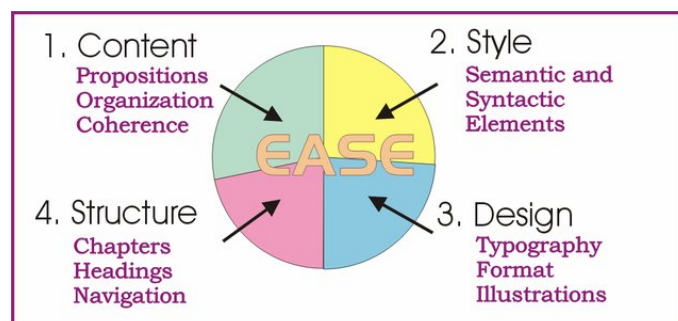
2.1.2. Mogelijke drempels voor mensen met fonologische dyslexie.

Mensen met fonologische dyslexie kunnen verschillende drempels ervaren, waaronder trage woordbenoeming, hardnekkig letter-voor-letter lezen, problemen met

woordherkenning en -herinnering, letter- en klankvorming, homofonische of pseudo-homofonische woordenschat en begripsproblemen (Bonte, 2020; Rivero-Contreras e.a., 2021; M. Zhang e.a., 2021). Het oefenen van pseudowoorden en het herkennen van woorden kan volgens Filipiak (2020) helpen, evenals het gebruik van educatieve apps en software, e-books en luisterboeken, woordspelletjes en puzzels en tekst-naar-spraak technologie. Visuele ondersteuning met film en afbeeldingen kan het leesbegrip verbeteren en het gebruik van een grote woordenschat en redeneervermogen kan nuttig zijn. Schriftelijke expressie kan echter problematisch zijn.

2.1.3. Bewezen effecten van tekstvereenvoudiging -en aanpassing bij scholieren met dyslexie

Dyslexie kan zich op verschillende manieren uiten bij elke leeftijdsgroep. Een ondersteunende toepassing moet met een individuele analyse van de specifieke behoeften en uitdagingen van elke leerling in gedachten worden ontworpen (Gooding, 2022). Instructies moeten op een begrijpelijke en geïndividualiseerde manier worden gepresenteerd om de leerlingen te helpen bij het begrijpen en toepassen van de informatie. Het is belangrijk om te erkennen dat dyslexie zich bij verschillende kinderen op verschillende manieren kan uiten. Een bijkomende stoornis heeft bijvoorbeeld geen impact op de spellingprestaties van een kind. Het is daarom belangrijk om een toepassing te ontwerpen met de diversiteit van dyslexie in het achterhoofd.

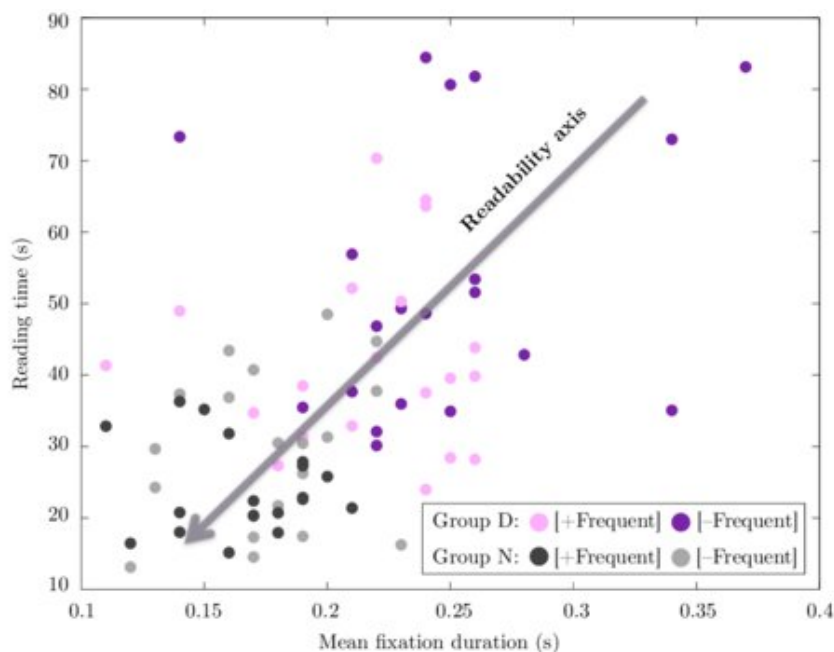


Figuur (2.1)

Afbeelding van DuBay (2004)

Lexicale wijzigingen

Het onderzoek van Rivero-Contreras e.a. (2021) laat zien dat vereenvoudigde teksten de leessnelheid en woordherkenning van kinderen met visuele dysfunctie significant kunnen verbeteren. Uit het experiment van Rello, Baeza-Yates, Dempere-Marco e.a. (2013) blijkt dat frequent woordgebruik de ontcijfertijd bij mensen met dyslexie significant vermindert en dat bevroagden met dyslexie minder leesfouten maken bij teksten met een verminderde lexicale complexiteit, volgens Gala en Ziegler (2016). Het experiment benadrukt ook de moeilijkheden van kinderen met



Figuur (2.2)

Afbeelding van Rello, Baeza-Yates, Dempere-Marco e.a. (2013). Volgens de richting van de pijl wordt de ideale situatie benaderd, gekenmerkt door doelwaarden. Deze waarden worden bereikt door mensen zonder dyslexie onder optimale omstandigheden. Het gebruik van vaak voorkomende woorden vermindert de decodeertijd en verbetert de leesbaarheid voor mensen met dyslexie.

dyslexie bij het lezen van woorden met meer dan zeven karakters en onregelmatige en infrequente lettergreepcombinaties.

Grammaticische wijzigingen

Onderzoek rond de effecten op syntactische vereenvoudiging bij kinderen en scholieren met dyslexie zijn in schaarse hoeveelheid. Het aanpassen causale structuren bij kinderen en jongeren met een lage leesgraad had een significant effect op het leestempo en de foutenmarge van de bevroagden uit het experiment van Lindnerholm e.a. (2000). Bij de revisies werden coherentieonderbrekingen werden hersteld door extra uitleg te voorzien, alsook door tekstgebeurtenissen in een temporele of tijdsafhankelijke volgorde te plaatsen. Zowel vaardige als minder vaardige lezers hadden baat bij de revisies. Verbale parafrasering heeft geen significant effect op lezers met dyslexie volgens Rello, Baeza-Yates en Saggion (2013). De bevroagden waren 13 tot en met 37 jaar oud met een gemiddelde leeftijd van 21 jaar. Het tekstformaat bleef ongewijzigd, maar lettertypes werden wel aangepast.

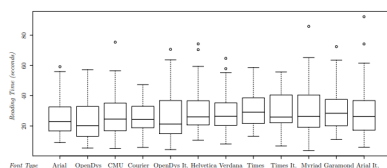
Formaat- en structurele wijzigingen

Geassisteerd samenvatten bevoordeelt de leesbaarheid van een scholier met dyslexie volgens het experiment van Nandhini en Balasundaram (2013). De geassisteerde samenvatting is gebaseerd op onaangepaste zinnen afkomstig uit de oorspronkelijke tekst. Het ontwerp bij dit experiment haalt de belangrijkste zinnen on-

aangepast uit de oorspronkelijke tekst, herorganiseerd deze volgens de structuur van de oorspronkelijke tekst en presenteert deze aan de lezer. Al werd de logische structuur van de gepresenteerde zinnen in vraag gesteld, de leesbaarheid van de bevroagden was significant beter dan bij de oorspronkelijke tekst zonder een nadelig effect op de verstaanbaarheid van de bevroagden.

Visuele presentatie

Experimenten tonen aan dat scholieren met dyslexie gevoeliger zijn voor veranderingen in visuele parameters zoals lettertype, karakterafstand, tekst- en achtergrondkleur en grijswaarden. Aanbevolen kleurencombinaties zijn een lichtgrijze achtergrond met zwart lettertype op een gele achtergrond, of zachtgele, -groene of lichtblauwe achtergrondkleuren (Bezem & Lugthart, 2016; Rello & Baeza-Yates, 2015; Rello & Bigham, 2017). Minimalistische ontwerpen met pictogrammen en afbeeldingen hebben een positief effect op de leesbaarheid van tekst. Lettergrootte groter dan 14pt en een sans-serif, *monospaced* of *roman* lettertype vergroten de leessnelheid. Het gebruik van lettertypes zoals OpenDys heeft geen effect op lezers met of zonder dyslexie, terwijl cursieve lettertypes worden afgeraden (Rello & Baeza-Yates, 2013; Rello & Baeza-Yates, 2015).



Figuur (2.3)

Afbeelding uit Rello en Baeza-Yates (2013). Verticaal wordt de gemiddelde mening van de bevroagden weergegeven. Horizontaal worden de lettertypes gerangschikt op gemiddelde leestijd van alle bevroagden. Dit onderzoek wijst uit dat Arial, CMU, Helvetica en Times de populaire keuzes zijn. Arial en CMU behoren hierbij tot de drie best scorende lettertypes rond gemiddeld leestempo.

2.1.4. Conclusie

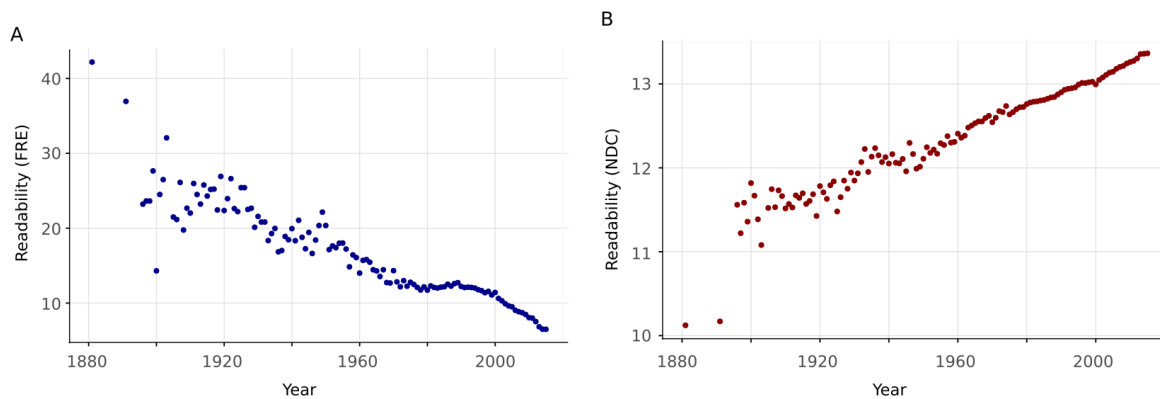
Op basis van de literatuurstudie kan besloten worden dat het aanpassen van teksten een significant effect heeft op de leessnelheid en woordherkenning van jongeren met dyslexie. Het vereenvoudigen van de lexicon en het toepassen van frequente woorden zorgt voor een verminderde decodeertijd en verbetert de leesbaarheid van teksten. Bevroagden met dyslexie hebben minder moeite met het lezen van teksten met verminderde lexicale complexiteit en geassisteerd samenvatten van teksten verbetert de leesbaarheid. Ook het aanpassen van de tekstweergave, zoals het gebruik van zachtgele, -groene of lichtblauwe achtergrondkleuren, kan de leeservaring van scholieren met dyslexie verbeteren. Het onderzoek rond de effecten op syntactische vereenvoudiging bij kinderen en jongeren met dyslexie is echter beperkt.

2.2. Wetenschappelijke artikelen

Wetenschappelijke artikelen volgen een uniform IMRAD-formaat en worden gebruikt als leermiddel voor jongeren in het middelbaar en hoger onderwijs. Het lezen van deze artikelen brengt uitdagingen met zich mee vanwege de verschillen in syntax en woordenschat. Docenten kunnen dit aanpakken in de derde graad van het middelbaar onderwijs door te benadrukken wat de specifieke kenmerken van wetenschappelijke artikelen zijn. In deze sectie wordt de volgende onderzoeksvraag beantwoordt: "Wat zijn de specifieke kenmerken van wetenschappelijke artikelen?"

2.2.1. Trends rond wetenschappelijke artikelen

De leesgraad van wetenschappelijke teksten volgt al sinds de tweede helft van de twintigste eeuw een stijgende trend (Hayes, 1992). Meerdere onderzoeken in de voorbije tien jaar besluiten dat de complexe woordenschat en zinsbouw deze wetenschappelijke artikelen ontoegankelijk maakt voor doelgroepen naast onderzoekers (Ball, 2017; Jones e.a., 2019; Plavén-Sigray e.a., 2017). Plavén-Sigray e.a. (2017) onderzoekt de verschillende trends waarom wetenschappelijke artikelen alsmaar moeilijker te lezen worden. De relatie tussen de leesbaarheid van een abstract werd vergeleken met het jaar waarin het wetenschappelijk artikel werd gepubliceerd. De *Flesch-Reading-Ease* of FRE score werd gebruikt om de leesgraad van een wetenschappelijk artikel te beoordelen. Om te bevestigen dat de relatie tussen de complexiteit van een abstract overeenstemt met die van de volledige tekstinhoud, werden er vergelijkingen gemaakt met zes verschillende wetenschappelijke journalen. De overeenkomst tussen de leesgraad van het abstract en de overige tekstinhoud in een wetenschappelijk artikel werd eerder bevestigd door Dronberger en Kowitz (1975). Dat onderzoek benadrukte dat een abstract complexer werd geschreven, vergeleken met de rest van een wetenschappelijk artikel.

**Figuur (2.4)**

Afbeelding uit Plavén-Sigraay e.a. (2017). Links wordt de evolutie per FRE-score getoond. Hoe hoger de score, hoe hoger de gemiddelde complexiteit van een tekst. Rechts wordt de evolutie volgens de NDC-score getoond. Hoe hoger de score, hoe lager de gemiddelde complexiteit van een tekst. Het onderzoek schat dat nu een kwart van alle wetenschappelijke artikelen gebruik maken van Engels op het niveau van een masterstudent, ofwel een FRE onder nul.

De hoge leesgraad van wetenschappelijke artikelen beperkt volgens Plavén-Sigraay e.a. (2017) twee aspecten: de toegankelijkheid en de herproduceerbaarheid.

Toegankelijkheid

Het belang van toegankelijke wetenschappelijke inhoud wordt ook benadrukt door Snow (2010), die stelt dat wetenschappelijke kennis een belangrijke rol speelt in het dagelijks leven en dat het daarom van groot belang is om wetenschappelijke informatie begrijpelijk te maken voor het bredere publiek. Dit kan de wetenschappelijke gemeenschap helpen om meer vertrouwen en steun te krijgen vanuit de samenleving en kan bijdragen aan het oplossen van problemen die voortkomen uit wetenschappelijke kwesties, zoals klimaatverandering. Om dit te bereiken, pleiten zowel F en Ennals (2010) als Snow (2010) voor het gebruik van eenvoudigere taal en duidelijkere zinsstructuren in wetenschappelijke communicatie. Door wetenschappelijke informatie toegankelijker te maken, kan de wetenschap beter bijdragen aan de maatschappij als geheel.

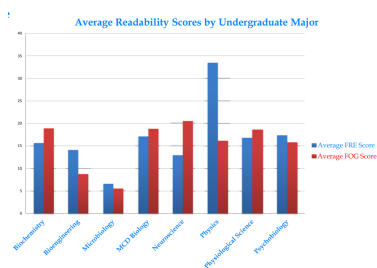
Reproduceerbaarheid

Wetenschappelijke artikelen kunnen zelfs voor vakexperten onbegrijpelijk zijn door complexe zinsstructuren en ontoegankelijke woordenschat. Om de begrijpbaarheid te vergroten, is het volgens onderzoekers als Hartley (1999) en Snow (2010) belangrijk om abstracten te herschrijven. Voor wetenschappers is het cruciaal dat hun onderzoeken reproduceerbaar zijn, wat betekent dat de inhoud van het wetenschappelijke artikel begrijpelijk moet zijn. Een lage leesgraad en duidelijke zinsbouw kunnen het aantal misopvattingen en verwarringen bij onderzoekers beperken. Experimenten van Hubbard en Dunbar (2017) tonen aan dat zowel de methodologie als de resultaten van wetenschappelijke artikelen een hoge leesgraad

vereisen, wat het belang onderstreept van begrijpelijke wetenschappelijke taal.

2.2.2. Woordenschat en vakjargon

Wetenschappelijke artikelen maken gebruik van complexe processen, methoden en ideeën, verwoord met academische taal die zich onderscheidt van taal van een lagere leesgraad. Hoewel de kenmerken van academische taal variëren afhankelijk van de discipline, het onderwerp en de vorm, zijn er gemeenschappelijke kenmerken die wetenschappelijke taal onderscheiden (F & Ennals, 2010; Snow, 2010). Volgens Plavén-Sigray e.a. (2017) dienen wetenschappelijke artikelen in eerste instantie als uitwisseling van kennis tussen vakexperten. Echter, de lengte van deze artikelen kan een nadelig effect hebben op de beschikbare uitleg van de terminologie. Daarom moet er in STEM-vakken of vakken waar deze wetenschappelijke artikelen aan bod komen, voldoende uitleg over de toegepaste grammatica en woordenschat worden voorzien tijdens de lessen, zoals benadrukt door Snow (2010).



Figuur (2.5)

Afbeelding van Murdos e.a. (2014) Volgens deze grafiek scoren de wetenschappelijke artikels rond fysica gemiddeld het best op de FRI-score. Al scoren de wetenschappelijke artikels rond microbiologie gemiddeld het zwakst op de FRI-score, ze scoren gemiddeld beter op de FOG-score.

2.2.3. Aanpak voor het lezen van wetenschappelijke artikelen

Als reactie op een satirisch artikel van Ruben (2016), bracht Pain (2016) het onderwerp bij wetenschappers aan het licht om zo verschillende tactieken te verzamelen om wetenschappelijke artikelen te begrijpen. Sommige wetenschappers zoeken direct onbekende woorden op of raadplegen extra informatiebronnen, terwijl andere wetenschappers hoofdstukken overslaan. Het is belangrijk om een balans te vinden tussen het begrijpen van de inhoud en het efficiënt gebruiken van de tijd. Sommige wetenschappers geven toe dat ze het soms opgeven als het te moeilijk wordt of als de literatuur net niet relevant is voor hun onderzoek. Pain (2016) bouwt verder op deze adviezen en bouwt een stappenplan op hoe lezers wetenschappelijke artikelen kunnen aanpakken.

1. Lees de samenvatting en conclusie om een idee te krijgen van het doel en de uitkomst van het onderzoek.
2. De figuren en tabellen in het artikel zijn cruciaal omdat deze een snelle en duidelijke weergave geven van de belangrijkste bevindingen.

3. Focus op de nodige informatie en ga vervolgens terug om de technische details te begrijpen.
4. Let op de beperkingen en interpretatie van de resultaten. Controleer of de onderzoeksvraag en -methode adequaat zijn.
5. Controleer of de referenties relevant zijn en zoek naar andere artikelen over hetzelfde onderwerp.
6. Overweeg welke stukken prikkelend, nieuw en relevant zijn voor eigen onderzoeksvragen en hypotheses.
7. Maak annotaties en schrijf tijdens het lezen, zodat de lezer actief betrokken is bij het lezen van het artikel.

Wetenschappelijke artikelen vereisen een selectieve en kritische leesstijl. Onderzoekers geven aan dat ze bepaalde delen van een artikel prioriteren, zoals de abstract en de discussie, om te bepalen of het artikel de moeite waard is om te lezen. Sommige onderzoekers adviseren om de methodologie over te slaan en direct naar de resultaten te gaan, terwijl anderen juist de hypothese willen kennen voordat ze verder gaan. Het artikel wordt vervolgens meermaals gelezen, waarbij de lezer steeds dieper ingaat op de details. Kritisch lezen is hierbij van groot belang, waarbij de conclusies worden beoordeeld en de data voor zichzelf spreekt. Er is geen standaardaanpak, maar er worden tactieken aanbevolen zoals selectief en met een specifiek doel voor ogen lezen. Het doel van deze selectieve leesstijl is om de wetenschappelijke inhoud beter te begrijpen en te beoordelen op relevantie voor het eigen onderzoek (Hubbard & Dunbar, 2017).

2.2.4. Conclusie

Het lezen van wetenschappelijke artikelen kan overweldigend zijn, vooral bij onbekende vakgebieden, lange artikelen en technisch vakjargon. Nieuwe versies van een wetenschappelijk artikel moeten meer doelgroepen toelaten om over voldoende achtergrondinformatie te beschikken. De gebruikte syntax, woordenschat en compact formaat sluiten aan bij de mogelijke struikelblokken voor een scholier met dyslexie in de derde graad van het middelbaar onderwijs.

2.3. Tekstvereenvoudiging

Vereenvoudigde teksten worden geschreven om leerlingen te ondersteunen bij het begrijpen van specifieke taalkenmerken, het beperken van de hoeveelheid nieuwe woordenschat en het beheersen van de complexiteit van de tekst. Deskundigen zijn van mening dat vereenvoudigde teksten nuttig zijn voor startende en gevorderde lezers (Louwerse e.a., 2007). Samenvattingen van teksten bieden een oplossing om een snel zicht te krijgen over (lange) documenten, of om de kerninhoud

van een tekst die al gelezen is opnieuw te prikkelen (McCombes, 2022). Vereenvoudigen kan handmatig door de docent gebeuren, maar recente technologische ontwikkelingen laten de automatisatie van dit proces toe met een gelijkwaardig eindresultaat. Deze sectie beantwoordt de volgende onderzoeksvraag: "Welke aanpakken zijn er voor geautomatiseerde tekstvereenvoudiging?". Aansluitend hierop wordt de volgende subvraag beantwoordt: "Hoe worden teksten handmatig vereenvoudigd voor scholieren met dyslexie?"

2.3.1. Manuele tekstvereenvoudiging

Wetenschappelijke artikelen moeten informatie begrijpelijk weergeven voor een breed publiek, waaronder de scholieren die deze artikelen voorgeschoteld krijgen. Teksten vereenvoudigen heeft volgens Crossley e.a. (2012) drie algemene doelen, namelijk het illustreren van een specifiek taalkenmerk, ongekende woordenschat voor een doelgroep aan te passen en de hoeveelheid gegeven informatie onder controle te houden. Crossley e.a. (2012) wijst op twee soorten van handmatige tekstvereenvoudiging. Intuïtieve tekstvereenvoudiging is een methode waar de auteur die de transformatie uitvoert, wordt beïnvloed door persoonlijke vermoedens over wat een tekst beter leesbaar maakt. Structurele vereenvoudiging daarentegen vervangt vermoedens door het gebruik van woordenlijsten en leesbaarheidsformules zoals Flesch Reading Ease (FRE), Gunning Fog (FOG), SMOG-Cro (SMOG) en de Coleman-Liau Index (CLI). Het onderzoek van Crossley e.a. (2012) wijst het bevorderend effect van de unieke aanpakken per docent uit. Iedere docent heeft een eigen intuïtie of afweging waarop teksten kunnen worden vereenvoudigd. Er is geen uniform formaat waarin een tekst kan worden vereenvoudigd.

Lengte en formaat

Een samenvatting verkort de tekst door kernzinnen en trefwoorden te markeren of te parafraseren. Dit vereist meerdere lezingen van de tekst. Trefwoorden achterhalen gebeurt gelijkaardig en deze zijn regelmatig af te leiden uit de inhoudstafel en titels. Er zijn verschillende soorten samenvattingen: informatief vervangt de tekst, indicatief bevat links naar andere bronnen, en kritisch bevat de kerninhoud en een opiniestuk (Hahn & Mani, 2000; Rijkhoff, 2022). Tekstvereenvoudiging vereenvoudigt complexe concepten en maakt de tekst begrijpelijker zonder betekenis of nauwkeurigheid te verliezen. Opsommingen benadrukken belangrijke punten en creëren structuur. Pragmatische vereenvoudiging zet metaforen, slang en idiomaten om naar duidelijke taal. Hahn en Mani (2000) onderscheidt generieke en gebruikersgerichte samenvattingen, waarbij het belang van gebruikersgerichte samenvattingen wordt benadrukt door technologieën zoals full-text-search en gepersonaliseerde informatiefiltering. De opbouw van een gebruikersgerichte samenvatting omvat analyse van de brontekst, het aanduiden van de kernpunten en het samenvoegen tot één uitvoertekst. Een wetenschappelijke samenvatting moet volgens Hollenkamp (2020) en McCombes (2022) drie vragen beantwoorden: waarom

werd het onderzoek verricht, wat werd er geëxperimenteerd en welke conclusies trekken de onderzoekers? Dit omvat de achtergrondinformatie, hypothesen, methoden, resultaten, implicaties, beperkingen en aanbevelingen. Alleen noodzakelijke kwalitatieve waarden moeten worden genoemd in de samenvatting. Vervolgens kan tekst naar een ander formaat worden omgezet, zoals *post-it*notes, tabelvorm of opsommingen om tekst begrijpelijker te maken (Rijkhoff, 2022).

2.3.2. Natural Language Processing

Tekstvereenvoudiging is het proces waarin het technisch leesniveau en/of woordgebruik van een geschreven tekst wordt verminderd. Het resultaat van deze fase is een tekst die korter en aangenamer is, zonder het verlies van de kerninhoud. Binnen machinaal leren (ML) is tekstvereenvoudiging een zijtak van natuurlijke taalverwerking. (Siddharthan, 2006) Volgens (Siddharthan, 2014) bestaat een complete en geautomatiseerde tekstvereenvoudiging uit vier verschillende vereenvoudigingen. *Natural Language Processing* (NLP) of natuurlijke taalverwerking is een brede term die zich richt op het verwerken en analyseren van menselijke taal door computers (Eisenstein, 2019). NLP omvat verschillende technieken, zoals tekstanalyse, taalherkenning en -generatie, spraakherkenning en -synthese, en semantische analyse. Computers zijn in staat om op een menselijke manier te communiceren en begrijpen wat er wordt gezegd. De volgende begrippen worden aangehaald in Eisenstein (2019) en Sohom (2019) en zijn fundamenteel voor de concepten die volgen.

Tokenisatie

Tokenisatie splitst tokens in een tekst en bouwt zo een woordenschat voor een taalmodel op. Dit kan volgens Menzli (2023) op woord-, karakter-, subwoord- en zinniveau gebeuren. Bij karakter-tokenisatie wordt de inputlengte groter en heeft daarmee volgens Ribeiro e.a. (2018) weinig betekenis. Zeldzame woorden worden opgesplitst in kleinere stukken om een woordenschat op te bouwen. Dit biedt voordelen ten opzichte van word-level tokenisatie (Iredale, 2022).

Lemmatiseren

Lemmatiseren in NLP bouwt verder op *stemming*, maar de betekenis van ieder woord wordt in acht genomen. Voor het lemmatiseren bestaan er Nederlandstalige modellen, waaronder JohnSnow¹. Bij **omgekeerd lemmatiseren** wordt er een afgeleide achterhaald vanuit de stam. Voor zelfstandige naamwoorden, zoals 'hond', is dit enkelvoud of meervoud (Eisenstein, 2019). Bij een **parsing**-fase wordt er een label aan ieder woord of zinsdeel toegekend. Voorbeelden van labels zijn zelfstandig naamwoord, bijwoord, werkwoord, bijzin of stopwoord. Het herkennen van zinsdelen wordt *chunking* genoemd. Parsing heeft een dubbelzinnigheidspro-

¹https://nlp.johnsnowlabs.com/2020/05/03/lemma_nl.html

bleem, want een 'plant' staat niet gelijk aan de vervoeging van werkwoord 'planten' (Eisenstein, 2019).

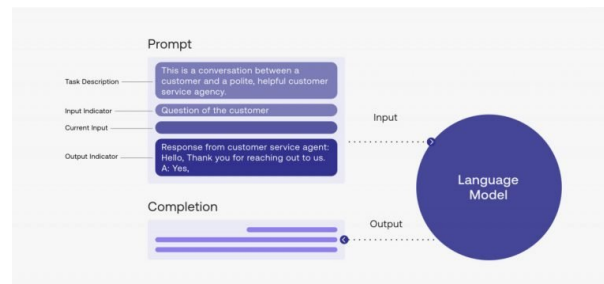
Sequence labeling en part-of-speech tagging

Een machine moet de betekenis achter ieder token kunnen vatten. Hier komt *sequence labeling* aan de pas volgens Eisenstein (2019). Elk woord in een tekst wordt gekoppeld aan een *Part-of-Speech* (PoS) of *Named-Entity-Recognition* (NER) label. Deze NLP-fase achterhaalt de structuur van een tekst. PoS-tagging richt zich op grammaticale categorieën van woorden, terwijl NER-labeling instaat voor het herkennen van specifieke entiteiten in een tekst. Bij PoS-tagging worden de woorden in een zin geanalyseerd. Elk woord wordt gekoppeld aan een grammaticale categorie, zoals een zelfstandig naamwoord, werkwoord, bijvoeglijk naamwoord of bijwoord. *PoS-tagging* helpt bij het achterhalen van de syntactische structuur van een zin. Deze taak komt van pas bij parsing en machinevertaling. *PoS-tagging* wordt aanschouwelijk gemaakt op 2.6. Namen van personen, organisaties en locaties worden herkend en geclassificeerd met NER-labeling. Met NER-labeling wordt volgens Jurafsky e.a. (2014) specifieke informatie uit tekst gehaald, zoals het identificeren van de namen van personen, plaatsen of bedrijven die in nieuwsartikelen worden genoemd, of het extraheren van belangrijke data of getallen uit financiële rapporten. Dit wordt aanschouwelijk gemaakt ?? J. Li e.a. (2018) benoemt vier vormen voor NER-labeling: *dictionary-based*, *rule-based*, *ML-based* en *deep learning-based*. De eerste twee gebruiken vooraf gedefinieerde woordenboeken en regels, terwijl de laatste twee gebruik maken van statistische of neurale netwerken om te leren hoe entiteiten te herkennen. Elke vorm gebruikt verschillende kenmerken en representaties om entiteiten te modelleren. Poel e.a. (2008) onderzocht *PoS-tagging* met een neuraal netwerk voor Nederlandstalige teksten. Het model behaalde een nauwkeurigheid van 97,88% voor bekende woorden en 41,67% voor onbekende woorden en gebruikte de Corpus Gesproken Nederlands (CGN) als trainingsdata.

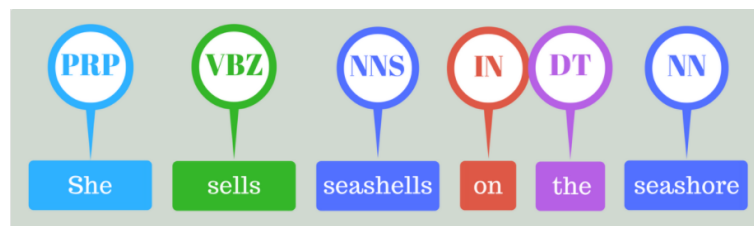
Word embeddings

NLP-systemen gebruiken embeddings om woorden numeriek te representeren en tekst te verwerken. Traditionele word embeddings bouwen een woordenschat op zonder de betekenis ervan op te volgen, terwijl contextual word embeddings wel de context van een woord begrijpen. BERT is een meertalig LLM dat contextual word embeddings gebruikt en getraind is op 110 miljoen parameters uit 104 verschillende talen, waaronder Nederlands. Voor de Nederlandse taal zijn er twee varianten van BERT², namelijk RobBERT en BERTje, waarvan RobBERT als krachtiger wordt beschouwd. Om de beste vervanging van woorden te bepalen, gebruikt het model de Substitution Ranking (SR) stap om substituties op basis van relevantie te rangschikken.

²<https://github.com/google-research/bert>

**Figuur (2.7)**

Afbeelding uit McFarland (2023). Een illustratie over de werking en begeleiding van prompt engineering bij een taalmodel.

**Figuur (2.6)**

Voorbeeld van PoS-labeling op de Engelstalige zin "She sells seashells on the seashore". Afbeelding van Bilici (2021)

2.3.3. Prompt engineering

Large Language Models of LLM's zoals GPT-3, BERT en T5 genereren tekst en karakters op basis van de probabiliteit of waarschijnlijke uitkomst van een gegeven input. Deze modellen maken gebruik van een neurale netwerk om patronen in de input te herkennen en deze patronen te gebruiken om voorspellingen te doen over de uitvoer (Liu e.a., 2020). Iedereen kan volgens McFarland (2023) een input of prompt schrijven. Deze tools zoals chatbots zijn ontworpen om zo intuïtief mogelijk te zijn voor een algemeen doelpubliek. Prompt engineering is een steeds belangrijkere vaardigheid die nodig is om effectief te communiceren met LLM's, zoals ChatGPT (Harwell, 2023).

Deze prompts kunnen volgens Liu e.a. (2020) gebruikt worden om werk te produceren dat is aangepast aan het doel. Een concrete en geoptimaliseerde prompt omvat een concrete scope, duidelijke vraagstelling, specifieke sleutelwoorden, de context en ten slotte gepersonaliseerde keuzes (McFarland, 2023). Bij een zoekopdracht moeten voldoende parameters in de prompt worden opgenomen. Zo niet zal het model te algemeen blijven en mogelijks afwijken van de intentie van de gebruiker. Effectieve AI prompt engineering leidt tot hoogwaardige trainingsgegevens die het AI-model in staat stellen om nauwkeurige voorspellingen en beslissingen te maken (Liu e.a., 2020).

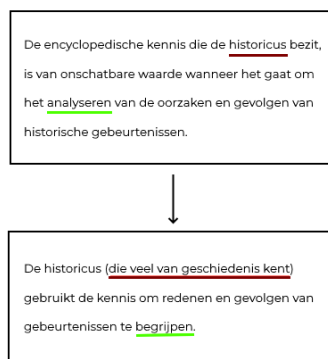
Prompt patterns

Prompt patterns is samen met prompt engineering naar boven gekomen en is vergelijkbaar met software patterns. Deze patronen zijn herbruikbare oplossingen voor veelvoorkomende problemen in een bepaalde context, waaronder vooral de interactie bij het werken met LLM's. White e.a. (2023) benoemt vier *prompt patterns*:

- Intent-prompts waarbij een LLM een instructie krijgt met een specifiek verwacht antwoord.
- Restriction-prompts die het antwoord van een LLM inperkt. Deze pattern is noodzakelijk om een LLM binnen de lijnen te houden.
- Contextualization-prompts verzekeren dat de output van een LLM relevant is. Een context wordt aan de LLM meegegeven.
- Expansion/reduction-prompts genereren een beknopte output met voldoende details.

2.4. De verschillende soorten tekstvereenvoudiging

Tekstvereenvoudiging bestaat volgens Siddharthan (2014) uit vier soorten transformaties: lexicale, syntactische en semantische vereenvoudiging en samenvatten.



Figuur (2.8)

Voorbeeld van tekstvereenvoudiging. Oorspronkelijke tekst uit Historia 5 bron toe te voegen

2.4.1. Lexicale vereenvoudiging

Bij *lexical simplification* (LS) of lexicale vereenvoudiging worden complexe woorden vervangen door eenvoudigere synoniemen. Bijvoorbeeld, het woord 'adhesief' wordt vervangen door 'klevend'. Kandula e.a. (2010) haalt twee manieren aan om lexicale vereenvoudiging mogelijk te maken, namelijk het vervangen door een synoniem en het aanmaken of genereren van extra uitleg. De zinsstructuur verandert niet en er is garantie dat de kerninhoud en benadrukking in een tekst iden-

tiel blijft. Het doel van lexicale vereenvoudiging is om de moeilijkheidsgraad van de woordenschat in een zin of tekst te verlagen.

Complex Word Identification

Complex word identification (CWI) is een gesuperviseerde NLP-taak. In een pipeline voor lexicale tekstvereenvoudiging is CWI de eerste stap. Moeilijke woorden of *multi-word expressions* (MWE) in een tekst worden achterhaald (Gooding & Kochmar, 2019; Shardlow, 2013). Na CWI kan LS gebruikt worden om deze woorden te vervangen door eenvoudigere synoniemen of om verdere elaboratie te voorzien met behulp van voorbeelden of definities (Kandula e.a., 2010; Zeng e.a., 2005). CWI is volgens Shardlow (2013) een cruciale stap, want een lage *recall* van dit component zal een uitvoertekst geven waar moeilijke woorden niet worden vereenvoudigd. Het model zal moeilijke woorden laten staan.

Substitutiegeneratie en ranking

Substitutiegeneratie wordt gedaan door synoniemen te zoeken voor een doelwoord in lexicale databanken zoals WordNet, BERT, context2vec, nPIC of OOC.



Figuur (2.9)

Afbeelding van Althunayyan en Azmi (2021). Deze pipeline wordt in meerdere onderzoeken rond lexicale vereenvoudiging toegepast, zoals Bingel e.a. (2018), Bulté e.a. (2018) en Paetzold en Specia (2016).

2.4.2. Syntactische vereenvoudiging

Syntactische vereenvoudiging vermindert de complexiteit van een tekst door de grammatica en zinsstructuur aan te passen. Dit kan door het combineren van twee zinnen tot één eenvoudigere zin of door de syntax te vereenvoudigen. Deze transformaties verbeteren de toegankelijkheid van de tekst zonder de inhoud te verliezen. Kandula e.a. (2010) ontwikkelden een model om medische informatie te vereenvoudigen. Dit model omvat drie modules, die zinnen met meer dan tien woorden vereenvoudigen en eventueel vervangen door kortere zinnen. De architectuur omvat een PoS Tagger, een Grammar Simplifier en een Output Validator.

- Voor de *PoS Tagger*-fase gebruikten Kandula e.a. (2010) beschikbare functies uit het open-source pakket OpenNLP³.
- De *Grammar Simplifier* module splitst de lange zin in twee of meer kortere zinnen door POS-patronen te identificeren en een set transformatieregels toe te passen.

³<https://opennlp.apache.org/>

- De *Output Validator* module controleert de output van de Grammar Simplifier op grammatica en leesbaarheid.

2.4.3. Tekstvereenvoudiging automatiseren

Geautomatiseerde tekstvereenvoudiging is geen nieuw concept. Volgens onderzoeken van Canning e.a. (2000) en Siddharthan (2006) waren de eerste aanpakken op geautomatiseerde tekstvereenvoudiging gebouwd op rule-based modellen. Deze modellen bewerken de syntax door zinnen te splitsen, te verwijderen of de volgorde van de zinnen in een tekst aan te passen. Lexicale vereenvoudiging kwam hier niet aan de pas. Enkel bij recentere onderzoeken van Bulté e.a. (2018) en Coster en Kauchak (2011) werd het duidelijk hoe lexicale en syntactische vereenvoudiging gecombineerd kon worden.

2.5. Samenvatten

Lexicale, conceptuele en/of syntactische vereenvoudiging van teksten leidt niet altijd tot een kortere tekst. Technologieën zoals full-text-search en gepersonaliseerde informatiefiltering benadrukken het belang van gebruikersgerichte samenvatting. De architectuur van een samenvattingssysteem omvat drie fases: analyse van de brontekst, identificatie van kernpunten en het samenvoegen van de punten tot één uitvoertekst. Teksten machinaal samenvatten is geen nieuw concept en kan op twee manieren gebeuren: extraherend en abstraherend (DuBay, 2004; Hahn & Mani, 2000).

2.5.1. Extraherend samenvatten

Bij extraherende samenvatting worden de belangrijkste zinnen gemarkeerd en opnieuw neergeschreven, maar dit kan leiden tot onsamenhangende uitvoertekst. Om de kernzinnen te achterhalen, zijn zes kenmerken volgens Khan (2014) essentieel, namelijk woordfrequentie, de plaats van een zin in de tekst, de *cue method*, titels, de lengte van de zin, de gelijkenissen tussen de zin en de rest van het document, *proper nouns* woordgebruik en ten slotte de afstand tussen *text units* waarin entiteiten voorkomen. Verma en Verma (2020) onderzocht verschillende manieren om een tekst extraherend samen te vatten, waaronder graafgebaseerd extraherend samenvatten, maximal marginal relevance en meta-heuristisch-gebaseerd samenvatten.

Graafgebaseerd extraherend samenvatten

De graafgebaseerde techniek van extraherend samenvatten vertegenwoordigt een document als een graaf van zinnen en gebruikt algoritmen om de belangrijkste zinnen te bepalen en redundantie te vermijden (Parveen & Strube, 2015). Dit kan volgens Parveen en Strube (2015) zowel voor lange wetenschappelijke artikelen als korte nieuwsartikelen goede resultaten opleveren, vooral als coherentie en posi-

tionele informatie worden opgenomen. Het compacte SqueezeBERT-model kan ook worden ingezet voor real-time samenvatting, als een interessant alternatief op het grotere BERT-model, met bijna de helft van de grootte en minimale afbreuk in prestaties. Beide methoden kunnen de prestaties van NLP-downstream taken verbeteren (Abdel-Salam & Rafea, 2022).

Maximal Marginal Relevance

Traditionele samenvattingssystemen zijn gebaseerd op de architectuur van Carbonell en Goldstein (1998), die gebruik maakt van de maximaal marginale relevantiescore (MMR) om de relevantie en diversiteit van gemarkeerde zinnen te bepalen. MMR zorgt ervoor dat de geselecteerde zinnen niet te veel overlappen in inhoud en relevantie. Onderzoekers hebben voorgesteld om het gulzige zoekalgoritme van MMR te vervangen door een globaal optimale formulering, wat een betere samenvatting oplevert, maar wel meer rekenkracht en tijd vereist. Evaluaties van deze methode tonen echter significant betere resultaten H. Lin en Bilmes (2010) en McDonald (2007).

Metaheuristiek-gebaseerd

Metaheuristische samenvatting is een benadering die gebruik maakt van optimalisatie-algoritmen zoals genetische algoritmen, simulated annealing en zwermoptimalisatie om de belangrijkste zinnen in een tekst te vinden. Volgens Premjith e.a. (2015) en Verma en Verma (2020) zoeken deze algoritmen naar de beste combinatie van zinnen om de kerninformatie in de tekst te bevatten. De evaluatiefunctie kan verschillende criteria bevatten, zoals zinslengte, relevantie en verbanden, aldus Rani en Kaur (2021). Een beperking van deze methode is dat deze vaak vastloopt in een lokaal optimum en geen extremen of steilere hellingen op een zoekruimte aan geeft. Om de convergentie te versnellen, is het nodig om een optimalisatiestrategie te gebruiken die gebaseerd is op gradiënten.

Experimenten over extraherend samenvatten

McKeown e.a. (1999) voerden experimenten uit op extraherende samenvattingen van nieuwsartikelen. De resultaten wijzen erop dat deze vorm vatbaar is op vooroordelen of *bias* van de auteur. De zinnen worden genomen zoals ze zijn. Hahn en Mani (2000) bouwde verder op dit experiment. Zij voerden een experiment uit met een mix van *knowledge-rich* en *knowledge-poor* methoden, met significant positieve resultaten tot gevolg. De nadruk bij extraherend samenvatten ligt in het kiezen van de *salient text units*. Deze punten zijn typisch in de vorm van zinnen. Er is nood aan een manier om de lexicale en statistische relevantie van een zin te kunnen aanduiden. Hiervoor haalt Hahn en Mani (2000) twee manieren aan:

- Een lineair gewicht model. Iedere teksteenheid wordt gewogen op factoren zoals de *location weight* en het aantal voorkomens.

- Een gewicht model op basis van de statistische opvallendheid van een eenheid. Zo wordt er rekening gehouden met de aanwezigheid van een woord in (sub)titels.

Nallapati e.a. (2017) wilden de nauwkeurigheid van deze modellen overbruggen. Dit doen ze met *SummaRuNNer*⁴, een oplossing voor het extraherend samenvatten van teksten met een neurale netwerk. De toepassing werd opgebouwd met PyTorch in en bestaat uit een combinatie van drie modellen: een recurrent neurale netwerk, een convolutioneel recurrent neurale netwerk en een *hiërarchische attentie netwerk*.

2.5.2. Abstraherend samenvatten

Er zijn twee manieren om een tekst abstraherend samen te vatten: semantisch en structuurgebaseerd. De structuurgebaseerde benadering gebruikt regels om belangrijke informatie in de tekst te vinden en kan leiden tot samengevatte zinnen met lage linguïstische kwaliteit en grammaticale fouten. De semantisch-gebaseerde benadering gebruikt de betekenis van de tekst om korte en duidelijke samenvattingen te maken met minder redundante zinnen en betere linguïstische kwaliteit, hoewel een extra parsingfase nodig kan zijn. Cao (2022) heeft verder onderzoek gedaan naar deep learning methoden om automatisch abstraherende samenvattingen te genereren. Deep learning-modellen zoals RNN's, CNN's en Seq2Seq kunnen worden gebruikt voor abstraherend samenvatten door de betekenis van de tekst te begrijpen en belangrijke informatie over te brengen (Suleiman & Awajan, 2020). Het Pegasus-model, beschreven in J. Zhang e.a. (2020), maakt gebruik van pre-trained modellen voor samenvatting met NLP en handelt gap-zinnen af, en is getraind en beoordeeld op verschillende soorten samenvattingstaken. LED of Long-former Encoder-Decoder is specifiek ontworpen om lange documenten te verwerken, waardoor het geschikt is voor het samenvatten van langere wetenschappelijke artikelen.

2.5.3. Hybride samenvatten

In het best denkbare geval wordt abstraherende en extraherende samenvatting gecombineerd volgens Hsu e.a. (2018) en Huang e.a. (2019). Zo omvat een pipeline voor hybride samenvatting twee onderdelen: een *content selection* fase waarin de kernzinnen met extraherende samenvatting worden opgehaald en *paraphrasing*-fase waarbij de gemarkeerde kernzinnen abstraherend worden samengevat.

2.5.4. Evaluatie

Samenvattingen van lange documenten handmatig beoordelen vergt tijd en voldoende planning van een mens (Nenkova & Passonneau, 2004). Met behulp van

⁴<https://github.com/hpzhao/SummaRuNNer>

een vooraf geschreven samenvatting als referentietekst is het mogelijk om een samenvatting automatisch te laten beoordelen. Samenvattingen beoordelen kan ook zonder referentietekst, al moeten verschillende factoren worden opgevolgd.

Evaluatie met referentieteksten

Bij het vergelijken van teksttransformaties worden vaak BLEU en ROUGE gebruikt, twee metrieken die de gelijkenis tussen machine-gegenereerde en referentieteksten meten gebaseerd op exacte *token matches*. ROUGE volgens C.-Y. Lin (2004) is recall-gebaseerd en houdt geen rekening met synoniemen, maar er zijn ROUGE-modellen die dit wel doen. ROUGE-2 van Ganesan (2018) voorziet dictionaries van synoniemen, zodat er rekening wordt gehouden met synonieme zinnen. ROUGE-G van ShafieiBavani e.a. (2018) gebruikt graafalgoritmen om lexicale en semantische matching mogelijk te maken. BLEU is precision-gebaseerd en een *brevity penalty* introduceert om te korte teksten te voorkomen. Er zijn Python-bibliotheken beschikbaar voor zowel ROUGE⁵ als BLEU⁶.

Evaluatie zonder referentieteksten

Een samengevatte tekst beoordelen zonder een referentietekst vereist volgens Steinberger en Jezek (2009) meer subjectiviteit en menselijke betrokkenheid dan met een referentietekst. Deze soort kan handmatig gebeuren, maar ook semi-automatisch. De type tekst, de lengte en de complexiteit van de oorspronkelijke tekst zijn factoren die in acht moeten worden genomen bij het beoordelen van de samengevatte tekst. Daar moet er worden gekeken naar het doelpubliek en het formaat. De tekst- en inhoudskwaliteit van de samengevatte tekst moet worden beoordeeld. De tekstkwaliteit is de grammaticale correctheid, niet-redundantie van zinnen en woordenschat en coherente structuur (McCombes, 2022). De inhoudskwaliteit wijst op de informatie dat wordt opgenomen in de samengevatte tekst. Dit omvat de relevantie met de doelgroep bij kern- en bijzaken of misleidende informatie door een misinterpretatie van het systeem (McCombes, 2022).

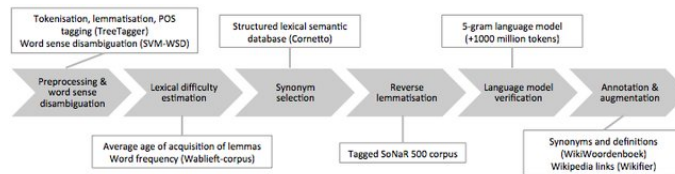
2.5.5. Tekstvereenvoudigingstechnieken voor scholieren met dyslexie.

In het onderzoek van Bingel e.a. (2018) wordt een systeem beschreven voor lexicale tekstvereenvoudiging, waarbij gebruik wordt gemaakt van een embeddings-gebaseerde aanpak voor substitution generation en een gesuperviseerd SR-model voor de selectie van synoniemen. Het systeem kan gepersonaliseerd worden op basis van gebruikersfeedback en maakt gebruik van een seed-dataset van complex-simple overeenkomsten. Het onderzoek van De Belder (2010) richt zich op tekstvereenvoudiging voor kinderen van acht tot twaalf jaar en maakt gebruik van een methode voor lexicale en syntactische vereenvoudiging, terwijl Bulté e.a. (2018) een

⁵<https://github.com/pltrdy/rouge>

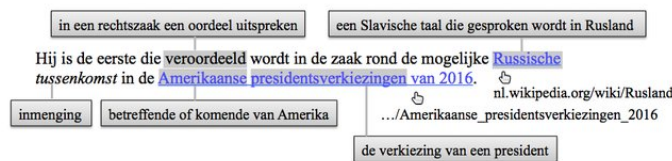
⁶<https://github.com/neural-dialogue-metrics/BLEU>

pipeline ontwikkelt om moeilijke woorden naar simpele synoniemen te vervangen, met behulp van een lexicale databank en LLM's.



Figuur (2.10)

Afbeelding uit Bulté e.a. (2018). Deze pipeline omvat de stappen die de toepassing aflegt.



Figuur (2.11)

Afbeelding uit Bulté e.a. (2018). TODO

Al zijn er onderzoeken over lexicale, syntactische en semantische vereenvoudiging voor kinderen en scholieren met dyslexie, het aantal onderzoeken over samenvatten voor deze doelgroep is schaars. Zoals eerder aangehaald is er wel onderzoek gedaan naar de verschillende manieren om een tekst samen te vatten, maar er is geen toepassing of onderzoek dat dit concreet uitwerkt.

2.5.6. Conclusie

Wetenschappelijke artikelen volgen een gelijke structuur. De inhoud in PDF- of afbeeldingvorm vergt voldoende cleaning-fasen. Het beoordelen van de samenvatting op basis van een referentietekst met de ROUGE-metriek wordt aangeraden, al kan deze beoordeling niet enkel machinaal gebeuren. Daarnaast is er input en bijsturing nodig van de mens omtrent een samenvatting op maat en de grammaticale, lexicale en semantische correctheid. Tools gericht op het lexicaal en adaptief vereenvoudigen van teksten voor kinderen en scholieren met dyslexie zijn reeds uitgewerkt. Methoden om menselijke en grammatisch correcte samenvatting op te bouwen zijn reeds beschikbaar.

2.6. Valkuilen en uitdagingen voor AI-ontwikkelaars bij tekst-vereenvoudiging met AI

AI en ML zijn volop in groei. NLP gebruikt AI en ML om menselijke taal te verwerken, terwijl NLU deze technologieën gebruikt om menselijke taal te begrijpen. Hoewel deze technologieën veelbelovend zijn, moeten AI-ontwikkelaars rekening houden

met veelvoorkomende en genegeleerde uitdagingen en valkuilen (Khurana e.a., 2022; Roldós, 2020; Sciforce, 2020). Deze sectie beantwoordt de volgende onderzoeksvraag: "Met welke valkuilen bij taalverwerking met AI moeten ontwikkelaars rekening houden?"

2.6.1. Uitdagingen voor softwarebedrijven

NLP- en NLU-toepassingen behoren tot de duurste om te ontwikkelen, wat een obstakel kan vormen voor veel IT-professionals. Het gebrek aan NLP-expertise, de kwaliteit en kwantiteit van data, de integratie en deployment van modellen en de transparantie van modellen zijn allemaal factoren die bijdragen aan deze hoge kosten (IBM, 2022). Software-ontwikkelaars verkiezen volgens voor *black-box* modellen bij de ontwikkeling en finetuning van een NLP-toepassing met AI. AI is het verschil qua nauwkeurigheid minimaal, de afweging wordt gemaakt bij de transparantie van het model. Na een transformatie wordt er niet aangegeven waarom specifieke transformaties werden uitgevoerd, bijvoorbeeld het vervangen van een woord door een eenvoudiger synoniem. White-box taalmodellen zijn er in schaarse hoeveelheden (Sikka & Mago, 2020).

2.6.2. Ambigüiteit, synoniemen en homoniemen

Homoniemen kunnen volgens Roldós (2020) problemen veroorzaken bij sequence labeling of het labelen van tokens in een doorlopende tekst. Bijvoorbeeld bij het woord 'bank' is het niet duidelijk voor de machine of het gaat over de geldinstelling of het meubel. Word Sense Disambiguation (WSD), PoS-tagging en contextual embeddings kunnen de betekenis van een woord achterhalen op basis van de context (Eisenstein, 2019; Liu e.a., 2020). Het gebruik van synoniemen en antoniemen in NLP-systemen kan verbeterd worden door het gebruik van candidate generation en synonym detection, en meertalige transformers zoals BERT bieden een oplossing voor het gebrek aan niet-Engelstalige toepassingen (Dandekar, 2016; Roldós, 2020).

2.6.3. Paternalisme en ethische overwegingen

Tekstvereenvoudiging is bedoeld om gelijke kansen te bieden aan iedereen, maar ethische overwegingen en bewustzijn van de behoeften van de eindgebruiker zijn belangrijk bij het ontwikkelen van adaptieve tekstvereenvoudigingstoepassingen, zoals beschreven in onderzoeken van Gooding (2022), Niemeijer e.a. (2010) en Xu e.a. (2015). De eindgebruiker moet de keuze hebben om te kiezen welke delen van de tekst vereenvoudigd moeten worden, wat kan worden bereikt door synoniemen te kiezen of zinnen te markeren die moeilijk te begrijpen zijn.

2.6.4. Valkuilen bij prompt engineering

Het bouwen van een conversatie met een chatbot is voor iedereen mogelijk, maar het vereist doordachte input en planning bij de ontwikkelaar om kwalitatief hoogwaardige antwoorden te krijgen. Een onnauwkeurige prompt of gebrek aan trainingsdata kan leiden tot onjuiste output, terwijl het gebruik van conditionele expressies of finetunen van hyperparameters kan helpen de betrouwbaarheid van het antwoord te vergroten (Jiang, 2023; Miszczak, 2023).

2.6.5. Evaluatie en interpretatie

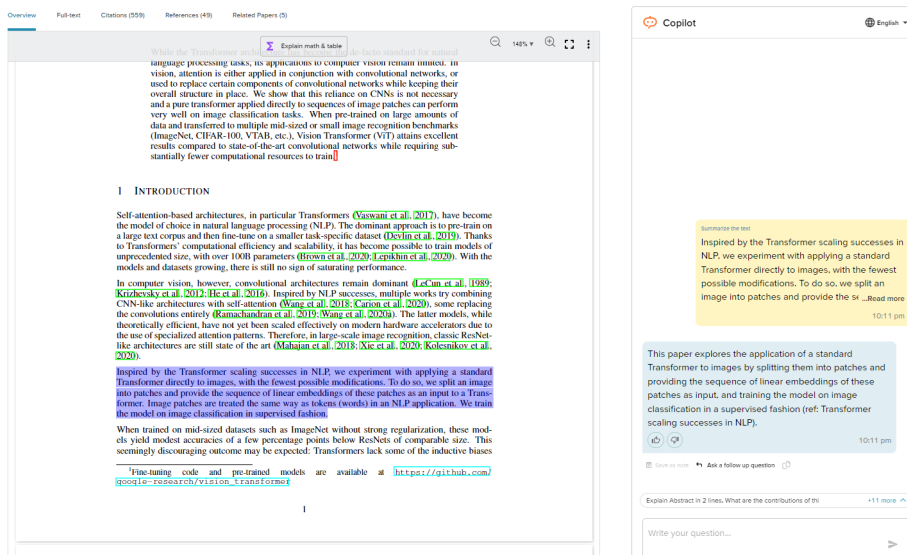
ROUGE en BLEU zijn beperkt omdat ze geen rekening houden met semantiek, maar ROUGE-L, ROUGE-SU en METEOR wel (Raj, 2017; Tatman, 2019). Menselijke evaluatie moet worden overwogen bij het onderzoeken van samenvattingsmethoden, en een mix van machinale en menselijke evaluatie is nodig volgens Fabbri e.a. (2020). De onderzoekers stimuleren verder onderzoek naar nieuwe standaarden en best practices voor betrouwbare menselijke beoordeling op extraherende en abstraherende samenvattingen. De doelgroep waarvoor een tekst wordt samengevat, moeten nauw in het proces worden opgenomen (Iskender e.a., 2021).

2.7. Beschikbare software voor tekstvereenvoudiging

Dyslexie is een veelvoorkomende aandoening die de lees- en schrijfvaardigheden van scholieren kan belemmeren. Om deze scholieren te ondersteunen, worden er verschillende softwareprogramma's en tools ontwikkeld. In dit hoofdstuk zal worden gekeken naar mogelijke nationale en internationale software die specifiek is ontworpen om scholieren met dyslexie te helpen bij het lezen van teksten. Er zal met name worden gekeken naar de beschikbare software in Vlaamse middelbare scholen, chatbots, zoals Bing AI en ChatGPT, en software die speciaal is ontwikkeld om dyslexie te ondersteunen bij het lezen. Deze sectie beantwoordt de volgende onderzoeksvraag: "Welke toepassingen, tools en modellen zijn er beschikbaar om Nederlandstalige geautomatiseerde tekstvereenvoudiging met AI mogelijk te maken?"

2.7.1. Momenteel ingezet in het onderwijs

In het middelbaar onderwijs wordt lees- en studieondersteuning voor scholieren met dyslexie enkel in de vorm van voorleessoftware voorzien (De Craemer e.a., 2018; OnderwijsVlaanderen, 2023). OnderwijsVlaanderen (2023) leent licenties voor de volgende softwarepakketten uit SprintPlus, Kurzweil3000, Alinea Suite, IntoWords en TextAid. Naast luister- en schrijfopties kunnen scholieren deze toepassingen gebruiken om zinnen te markeren om deze zinnen vervolgens samen te vatten. Enkel de gemarkeerde zinnen worden betrokken in de samengevatte versie, dus de zinnen blijven lexicaal, syntactisch en semantisch identiek. Alle vermelde softwa-



Figuur (2.12)
Schermabbeelding van SciSpace

repakketten bieden echter geen onafhankelijke samenvat- of vereenvoudigfunctie aan. Tops e.a. (2018) benadrukt de handige aspecten van deze software, maar deze software moet zo vroeg mogelijk in een schoolcarrière worden ingezet. Zo raken de scholieren snel vertrouwd met het gebruik, wat kan leiden tot een optimaal gebruik in verdere studies. Volgens Tops e.a. (2018) is het te laat om deze software pas in het hoger onderwijs te introduceren.

2.7.2. Proof-of-concepts en online webapplicaties

Online zijn er tools beschikbaar om teksten generiek samen te vatten. Resoomer, Paraphraser en Scholarcy zijn oorspronkelijk Engelstalige tools, met ondertussen de mogelijkheid om een abstraherende samenvatting te maken van Nederlandstalige teksten. De taalmodellen waar deze applicaties op werken, is niet gekend. Daarnaast zijn er ook geen API's beschikbaar om mee te werken. Gepersonaliseerde toepassingen zijn er in mindere mate. Bingel e.a. (2018) omschrijft een proof-of-concept voor een webtoepassing dat teksten vereenvoudigd, met oog op mensen met dyslexie. Deze software noemt nu Hero en bevindt zich in betafase. Toepassingen om wetenschappelijke artikelen te vereenvoudigen zijn schaars, maar er zijn enkele gratis en betalende toepassingen beschikbaar. SciSpace⁷ is gratis. Scholarcy⁸ is betalend.

⁷<https://typeset.io/>

⁸<https://www.scholarcy.com/?ref=theresanaiforthat>

2.8. GPT-3

Generative Pretrained Transformer 3 of GPT-3 is een taalmodel ontworpen door OpenAI. Dit taalmodel gebruikt een tweestapsleerparadigma waarbij het eerst ongesuperviseerd wordt getraind met een taalmodelleringsdoel en daarna gesuperviseerd wordt gefinetuned. Over drie versies heen is het model aanzienlijk vergroot, van anderhalf miljard parameters bij GPT-2 naar 175 miljard parameters bij GPT-3. Het model is getraind op niet-gecategoriseerde data van het internet en gebruikt datasets waaronder Common Crawl, WebText2, Books1, Books2, and Wikipedia (C. Li, 2022; Radford e.a., 2019).



Can you simplify the following paragraph:

At some ideas you stand perplexed, especially at the sight of sin. And you must ask yourself whether to combat it with force or with humble love. Always decide to combat it with humble love. If you decide this once and for all, you can conquer the whole world! Humble love is a frightening force, of which there is nothing stronger.



Here is a simpler version of the paragraph:

When you see something bad, like sin, you might be confused. Ask yourself if you should use force or love to fight it. Always choose love. If you do, you can win against anything. Love is the strongest power.

Figuur (2.13)

Afbeelding van Gooding 2022. De invoertekst is een paragraaf uit een niet-vermeld boek van de Russische schrijver Dostoevsky. Het resultaat van de meegegeven prompt is een transformatie dat iedere vorm van vooraf aangehaalde vereenvoudiging weergeeft. Lexicale, conceptuele en syntactische vereenvoudiging worden op de invoertekst toegepast.



Can you simplify the following paragraph in three different versions:

At some ideas you stand perplexed, especially at the sight of sin. And you must ask yourself whether to combat it with force or with humble love. Always decide to combat it with humble love. If you decide this once and for all, you can conquer the whole world! Humble love is a frightening force, of which there is nothing stronger.



Here are three simplified versions of the paragraph:

1. When you see something bad, like sin, you might be confused. Ask yourself if you should use force or love to fight it. Always choose love. If you do, you can win against anything. Love is the strongest power.
2. When you encounter ideas that confuse you, especially when it comes to sin, you may be unsure how to react. You must ask yourself if it is better to fight back with force or with humble love. It is always better to choose humble love. If you make this decision once and for all, you can overcome anything. Humble love is a powerful and intimidating force that is stronger than anything else.
3. If you see something bad, like sin, and you don't know what to do, remember that love is always the answer. Choose love over force. If you do, you will be able to conquer anything. Humble love is the most powerful force in the world.

Figuur (2.14)

Afbeelding van Gooding 2022. Gooding haalt verder aan dat modellen zoals ChatGPT op twee vlakken de leesbaarheid van een tekst kan bevorderen. Allereerst door het verlenen van verschillende mogelijke versies van een vereenvoudigingstaak.



Which of the three simplified versions is the simplest?

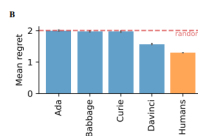


Of the three simplified versions, the first one is the simplest. It uses the shortest and simplest words and phrases to convey the main idea of the original paragraph, which is that humble love is the strongest force and can be used to overcome anything. The other two versions add more details and examples, which make them slightly more complex than the first version. However, all three versions are simpler than the original paragraph and effectively convey the same idea.

Figuur (2.15)

Afbeelding van Gooding 2022.

Lisowski (2023) vergelijkt de twee OpenAI taalmodellen met een *mixed-methods* onderzoek. Al blijken de twee heel gelijkaardig, het experiment benadrukt dat het ChatGPT-model gericht is op conversationele doeleinden met voorkeur als chatbot, terwijl GPT-3 een ML-model is bedoeld om met hoogstens één prompt te werken. De grootte van het GPT-3 model met 175 miljard parameters imposanter dan ChatGPT. Daarnaast is de limiet bij het meest recente GPT-3 model is 4000 tokens. Verder haalt Lisowski aan dat de kwaliteit bij beide modellen sterk afhankelijk is van de invoer. De prompts moeten concreet genoeg zijn, om zo niet af te wijken van wat de gebruiker wilt (Lisowski, 2023). Deze twee API's zijn nu vrij beschikbaar voor ontwikkelaars als betalende API (Greg e.a., 2023).



Figuur (2.16)

Afbeelding van Binz en Schulz (2023). Dit toont de *mean regret* aan tussen de vier engines en de menselijke antwoorden.

Beschikbare GPT-3 engines

De documentatie van OpenAI⁹ reikt vier verschillende engines voor het GPT-3 taalmodel aan, namelijk Davinci, Curie, Babbage en Ada. In Maart 2023 voegde een vijfde engine zich toe, namelijk GPT-3 Turbo wat de basis is achter Chat-GPT. Davinci-003 is het meest geavanceerde model dat alles kan wat de andere engines ook kunnen, met de meest menselijke antwoorden en geschikt voor taken zoals essays schrijven en code genereren. Curie is goed voor nuance maar minder menselijk dan Davinci, terwijl Ada en Babbage minder krachtig zijn en aangeraden worden voor eenvoudige taken zoals tekst aanvullen en sentiment analyse (Greg e.a., 2023).

Tools met GPT-3

De mogelijkheden van OpenAI's ChatGPT en GPT-3 modellen zijn nog volop in ontwikkeling, maar er zijn al enkele vergelijkende onderzoeken uitgevoerd. Uit het experiment van Tanya Goyal (2022) blijkt dat *zero-shot* samenvattingen met GPT-3 beter presteren dan *fine-tuned* modellen. Daarnaast haalt Mottes (2023) verschillende tools aan die gebruik maken van de GPT-3 API, waaronder Jasper AI en ChatSonic. Ook voor het onderwijs zijn er mogelijkheden, zoals de hoge toegankelijkheid en granulaire personalisatie van het GPT-3 model (Garg, 2022; Roose, 2023). Echter, GPT-3 is niet geschikt voor alle taken, zoals sentimentanalyse en -classificatie, waarvoor een kleinschaliger taalmodel beter presteert (C. Li, 2022). Bovendien is er aandacht voor de ecologische effecten van de grote omvang van deze modellen, waarvoor alternatieve oplossingen zoals het gebruik van Cloud-infrastructuur en geschikte model finetuning worden voorgesteld (Simon, 2021; Strubell e.a., 2019).

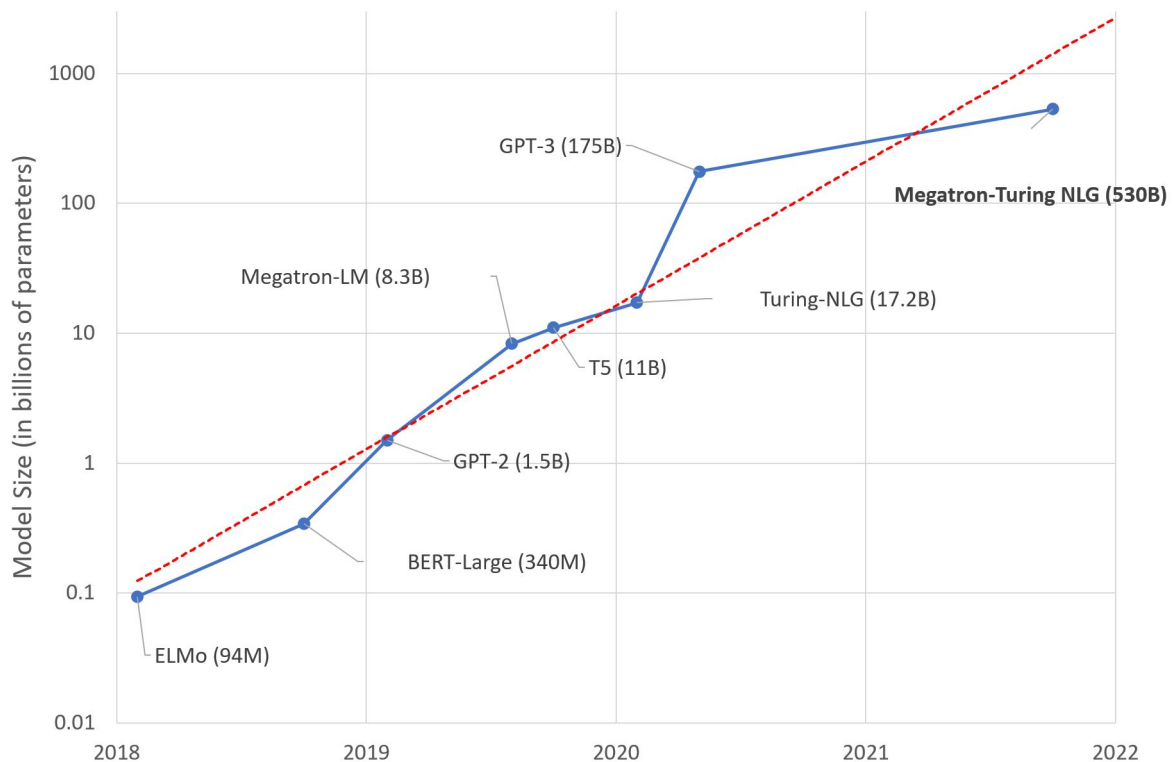
Vergelijking met andere taalmodellen

De architectuur tussen GPT-3 en BERT is volgens Mottes (2023) het meest opvallende verschil. GPT-3 is een autoregressief model en houdt daarmee enkel rekening met de linkercontext bij het voorspellen of genereren van tekst. BERT daarentegen is bidirectioneel en neemt zowel de linker- als de rechtercontext in overweging. De bidirectionele werking is geschikt voor sentimentanalyse waarbij begrip van de volledige zincontext noodzakelijk is. GPT-3 heeft toegang tot meer informatie (45TB) dan BERT (3TB), wat het een voordeel kan geven bij het samenvatten of

⁹<https://platform.openai.com/docs/>

het vertalen. Ten slotte zijn er ook verschillen in grootte. Hoewel beide modellen erg groot zijn, GPT-3 is aanzienlijk groter dan de voorganger vanwege de uitgebreide trainingsdatasetgrootte (Brown e.a., 2020).

LLaMa of Large Language Model Meta AI is een generatief taalmodel met potentieel dat sterker is dan GPT-3 en soortgelijke modellen, terwijl het van tien keer minder parameters gebruik maakt, maar is nog niet beschikbaar als online webtoepassing of API (Hern, 2023; Touvron e.a., 2023).



Figuur (2.17)

Afbeelding van Simon (2021). De evolutie van pre-trained taalmodellen wordt hier weergegeven tot eind 2022. De prestatie van de modellen ten opzichte van de grootte volgt een lineaire functie.

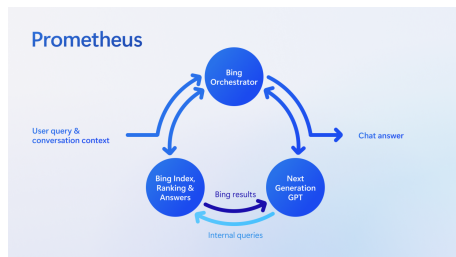
GPT-3 finetuning

Parameter	Omschrijving	Mogelijke waarden
model	Het GPT-3 model om te gebruiken	davinci, curie, babbage, ada, text-davinci-002, text-curie-001, text-babbage-001, text-ada-001, davinci-codex
temperature	De gulzigheid van een generatief model. Een lagere waarde zal conservatieve en voorspelbare tekst teruggeven. Hogere waarden zullen meer gevarieerde en onverwachtse tekst teruggeven, wat beter werkt bij creatieve toepassingen.	Een kommagetal tussen 0 en 1.
max_tokens	Het maximaal aantal tokens (woorden of subwoorden) dat het generatief model kan teruggeven.	Een getal tussen 1 and 2048.
top_p	Vergelijkbaar met temperature, maar deze waarde onderhoudt de probability distribution voor common tokens. Hoe lager de waarde, hoe waarschijnlijker de woordenschat dat het model zal overwegen bij het genereren van tekst. Een hoge waarde is toepasselijker wanneer een toepassing gericht is op nauwkeurigheid en correctheid.	Een kommagetal tussen 0 en 1.
stop	Een tekstwaarde (woord/symbool) tot waar het model zal genereren. When the model generates a string that matches any of the specified strings, it stops generating text.	Een lijst van stringwaarden, of een enkele string.
presence_penalty	Factor die bepaalt hoe regelmatig woorden voorkomen.	Een kommagetal tussen 0 en 1

2.8.1. Bing AI

Microsoft en OpenAI werken nauw samen. Zo maakt het conversationele taalmodel van Bing ook gebruik van GPT-3. Deze chatbot bouwt verder en biedt zo verwijzingen en referenties aan naar andere websites. Deze verwijzingen zijn volgens

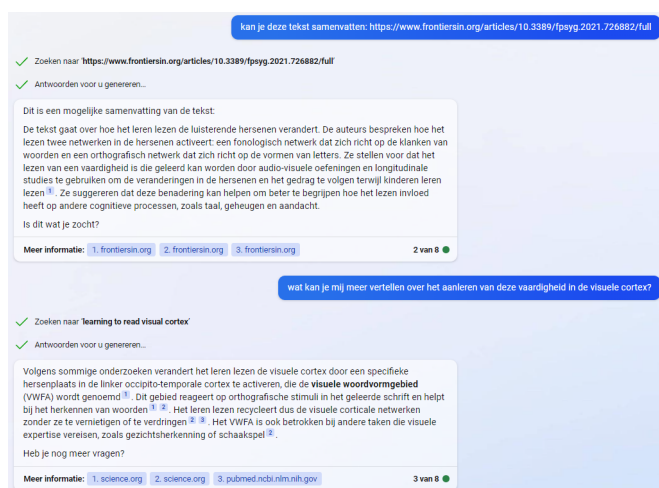
mogelijk door de Prometheus-technologie van Microsoft (Ribas, 2023). Prometheus is een eigen technologie die door Bing is ontwikkeld. Het AI-model is volgens Ribas (2023) de eerste van zijn soort die de Bing-index-, ranking- en antwoordresultaten combineert met het redeneervermogen van OpenAI's GPT-modellen. Prometheus maakt gebruik van de kracht van Bing en GPT om iteratief via een component genaamd *Bing Orchestrator* een set interne queries te genereren met als doel binnen gegeven gesprekscontext een nauwkeurig antwoord op gebruikersqueries te bieden (Ribas, 2023).



Figuur (2.18)

Abbeelding van Ribas (2023).

Bing AI is nu in testfase met wachtlijst en bestaat in de vorm van een webpagina en een browserextensie voor Microsoft Edge. Onderzoek naar deze chatbot staat nog in de kinderschoenen en er is nood aan onderzoek naar de credibiliteit en correctheid van de verwijzingen. Deze chatbot gebruikt een combinatie van extraherende en abstraherende samenvattingen. In tegenstelling tot GPT-3 is er geen officiële API beschikbaar. Daarnaast is de limiet ook lager met 2000 tokens per bericht tijdens een conversatie.



Figuur (2.19)

In deze afbeelding wordt er een online wetenschappelijk artikel meegegeven. Er wordt geen titel of onderwerp meegegeven, maar de Bing AI chatbot is in staat om een abstraherende samenvatting te maken van het artikel. Daarna geeft de chatbot verder uitleg over een bepaald onderwerp en geeft het extra referenties mee.

Het bedrijf DuckDuckGo dat instaat voor de gelijknamige zoekmachine probeert een gelijkaardig initiatief. Met *DuckAssist* biedt de onderneming een eigen AI-oplossing aan om een algemene doelgroep te ondersteunen bij het opzoeken van (nieuwe) informatie. Zij halen informatie direct uit enkel Wikipedia pagina's (Weinberg, 2023). Daarnaast maakt dit DuckAssist ook gebruik van het GPT-3 model. Nadelig heeft deze toepassing voorlopig beschikking tot een kleinere zoekruimte dan Bing AI, wat gebruik maakt van meer sites inclusief onderzoekssites zoals ResearchGate (McAuliffe, 2023). Deze beperkte zoekruimte reduceert de kans op incorrecte of foutieve informatie volgens Weinberg (2023), al is dit eerder een intuïtie van het bedrijf.

2.8.2. Huggingface en taalmodellen via API

In recente literatuur is Huggingface beschreven als een platform of portaalsite voor het delen van ML-modellen en datasets. De bibliotheek biedt een scala aan API's en tools die gemakkelijk te downloaden en trainen zijn voor pretrained modellen voor prevalentie NLP-taken, zoals tekstclassificatie, taalmodellering en samenvatting. Deze modellen kunnen worden gefinetuned op specifieke datasets, waardoor ontwikkelaars snel modellen kunnen bouwen en inzetten voor vereenvoudigings- en samenvattingstaken. Voor wetenschappelijke documenten en artikelen bestaan er enkel modellen en datasets: ^{10, 11}

2.8.3. Conclusie

Experten halen het GPT-3 model en ChatGPT aan als de toekomst voor gepersonaliseerde en adaptieve uitleg aan scholieren. Bing AI biedt een extra dat revolutionair kan zijn bij het opzoeken van uitleg voor zoektermen, zonder het verlies aan bronvermelding. Huidige toepassingen staan mogelijks in een spreekwoordelijke schaduw eenmaal leessoftware voor scholieren met dyslexie worden ontwikkeld met AI. De mogelijkheden van GPT-3 zijn eindeloos en toepassingen die hiervan gebruik maken, kunnen in het onderwijs ingezet worden als ondersteunende software.

2.9. Conclusie

De noden van scholieren met fonologische dyslexie in de derde graad van het middelbaar gaan verder dan gewoon moeizaam lezen. Het ontcijferen en automatiseren van wordeherkenning gebeurt langzaam. Er zijn bewezen voordelen van manuele tekstvereenvoudiging en adaptieve visuele weergaven op kinderen en jongeren met dyslexie. De leesbaarheid van wetenschappelijke artikelen bevindt zich in een neergaande trend. Het formaat, gebruik van vakjargon en ingewikkelde woordenschat en ten slotte de moeizame syntax en zinsbouw sluiten een algemene

¹⁰<https://huggingface.co/sambydlo/bart-large-scientific-lay-summarisation>

¹¹https://huggingface.co/haining/scientific_abstract_simplification

doelgroep uit bij het lezen van wetenschappelijke artikelen. Enkel wetenschappelijk geletterden zijn in staat om deze artikelen te lezen. Het uniforme formaat van een wetenschappelijk artikel biedt kansen aan voor een geautomatiseerde aanpak tot het vereenvoudigen van een tekst.

Experten halen meerdere bewezen tactieken aan om teksten automatisch te vereenvoudigen op maat voor een scholier met dyslexie. Handmatig worden teksten vereenvoudigd aan de hand van leesbaarheidsformules of intuïtie. Zinnen moeten lexicaal, syntactisch en semantisch worden vereenvoudigd. Teksten samenvatten maakt de tekst korter zonder het verlies van de kernboodschap. Voor deze vier transformaties zijn er taalmodellen beschikbaar in de vorm van API's of open-source software. Huidige software dat de overheid uitleent aan scholieren met dyslexie in het middelbaar onderwijs fungeert voornamelijk als voorleessoftware. Nieuwe en opkomende technologieën en taalmodellen zoals GPT-3 blinken uit om tekstvereenvoudiging mogelijk te maken. De ontwikkeling met LLM's is in opmars, maar ontwikkelaars moeten bewust zijn dat andere taalmodellen zoals BERT voor taken zoals semantische analyse minder rekenkracht vereisen voor eenzelfde en soms beter resultaat.

3

Methodologie

Het onderzoek omvat een literatuurstudie (Hoofdstuk 2), waarin technologieën en methoden voor tekstvereenvoudiging voor leerlingen met dyslexie in de derde graad van het middelbaar onderwijs worden beschreven. Hoofdstukken 4 en 5 presenteren een lijst van benodigde functionaliteiten en een stappenplan voor de ontwikkeling van een prototype. In Hoofdstuk 6 wordt een vergelijkende studie uitgevoerd om de meest geschikte tools te bepalen voor het vereenvoudigen van wetenschappelijke artikelen voor leerlingen met dyslexie. Het doel is om te bepalen aan welke criteria een vereenvoudigde tekst moet voldoen om leerlingen met dyslexie in de derde graad van het middelbaar onderwijs te ondersteunen.

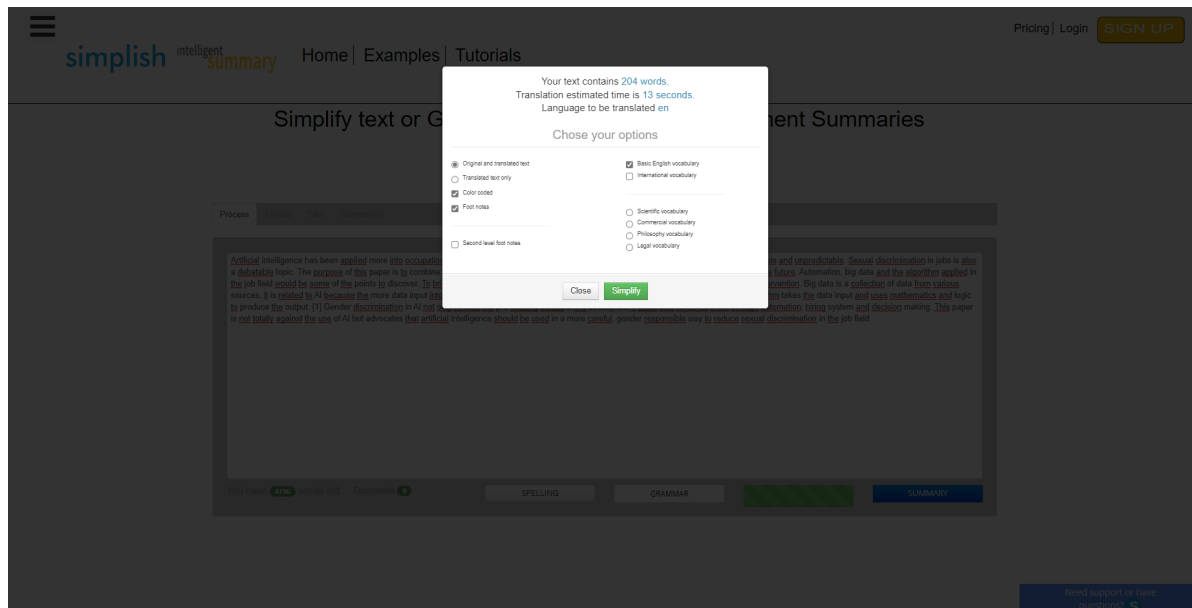
4

Requirementsanalyse

In deze onderzoeksfase worden verschillende tools getest en vergeleken met manuele tekstvereenvoudiging. De vereenvoudigingstechnieken die als gunstig zijn beoordeeld in Hoofdstuk 2, evenals de aspecten waar ontwikkelaars zich bewust van moeten zijn, worden meegenomen in de requirementsanalyse. Deze analyse is verdeeld tussen toepassingen die momenteel in het onderwijs worden gebruikt en online tools die leraren kunnen gebruiken. Gratis beschikbare taalmodellen, zoals GPT-3 of vooraf getrainde taalmodellen zoals BART-SC, worden ook in de analyse opgenomen. Op basis van de capaciteiten en functionaliteiten van de verschillende tools wordt een shortlist opgesteld met de benodigde functionaliteiten om teksten te vereenvoudigen voor middelbare scholieren in de derde graad met dyslexie.

4.1. Tekstanalyse

Geen tool in de shortlist biedt transparantie over keuzes van het taalmodel. Beslissingen zoals CWI worden niet aangegeven door het taalmodel. Geen enkel softwarepakket of hulpmiddel biedt standaard een visuele weergave van waarom een taal- of AI-model een zin als moeilijk of belangrijk beschouwt, of waarom het model een kernwoord heeft gekozen. Dit komt overeen met de bevindingen van Gooding en Kochmar (2019). Het GPT3-model en het verwante Bing-model doen dit echter wel wanneer het taalmodel hier expliciet om wordt gevraagd. SciSpace houdt hier geen rekening mee en verwerpt de vraag. Het stellen van vragen aan het taalmodel biedt weliswaar een alternatief, maar het schrijven van efficiënte prompts valt volgens buiten het bereik en de capaciteiten van de gemiddelde eindgebruiker. Deze prompt kan worden aangeboden in de vorm van een intuïtieve knop. Simplish geeft nadien een vergelijkende weergave met de oorspronkelijke tekst en de vereenvoudigde tekst. Met gebruik van kleurcodes worden de verschillende transformaties aangeduid, maar is enkel bij de uitvoer op de site terug te vinden.



Color code	
Black	Words in Black don't change between the two versions.
Green	Words in Green mean they have been translated adequately.
Purple	Words in Purple display a further explanation in foot notes.
Olive	Words in Olive contain two or more possible meanings (a tooltip is provided for these words, place the mouse cursor on top of olive words to see possible meanings).
Blue	Words in Blue are recognized in Wikipedia (normally names, places, people, organizations, etc.).
Orange	Words in Orange are not currently available in Basic English.
Red	Words in Red are names, special terms or not recognized by the translating tool.

Note : *Double click* on any word to add it to your personal dictionary.

Input Text

Artificial intelligence has been applied more into occupations by companies and individuals. However, the effects within the benefits are both imaginable and unpredictable. Sexual discrimination in jobs is also a debatable topic. The purpose of this paper is to combine the topics of both AI and sexual discrimination and discuss their effects in the job field in the future. Automation, big data and the algorithm applied in the job field would be some of the points to discover. To briefly summarize, automation is the use of machines and computers that reduces human intervention. Big data is a collection of data from various sources, it is related to AI because the more data input into AI the better it becomes. Since AI absorbs the information and learns from them. AI algorithm takes the data input and uses mathematics and logic to produce the output. [1] Gender discrimination in AI not only reflects the pre-existing biases in the society, but it could also reinforce them through automation, hiring system and decision making. This paper is not totally against the use of AI but advocates that artificial intelligence should be used in a more careful, gender responsible way to reduce sexual discrimination in the job field.

Simplified

artificial intelligence¹ has been made a request more into work by companies and beings, however, the effects within the gets help are both idea forming and not able to say before hand, sex caused decision making in regular work is also an about which argument is possible thing talked of. The purpose of this paper is to trading group the interests of both AI and sex caused decision making and have a discussion about their effects in the regular work field in the future. automation², greatly sized facts and the algorithm³ applied in the regular work field would be some of the points to discover. To briefly give a short account of, automation² is the use of machines and knowledge processing machines that gets changed to other form to do with man coming between groups. Big facts is a group of facts from different starting points, it is related to AI because the more facts input into AI the better it becomes. Since AI takes up the news given and learns from them. AI algorithm³ takes the facts input and uses mathematics and tests, reasoning to produce the out put. [1] sex statement decision making in AI not only gives back (right, heat, sound) the in existence beforehand has a tendency in a certain direction in the society, but it could also make stronger them through automation², getting use of person for money system and decision making. This paper is not totally against the use of AI but Advocates⁴ that artificial intelligence¹ should be used in a more careful, sex statement responsible way to get changed to other form sex caused decision making in the regular work field.

Menu

artificial intelli... science that gives great weight to ways of making come into existence intelligent machines that work and have reactions like those of man. [Continue reading.](#)

automation² the technology of making machines, instruments, process, and the like go through a certain train of operations without further impulse or control from outside after being started. [Continue reading.](#)

algorithm³ a word coming from the name of the expert in mathematics /AI-Khwarizmi@who (780-850ac), used to give the way to work out or solve points to be answered. [Continue reading.](#)

Advocates⁴ A barrister or solicitor representing a party in a hearing before a Court. [Continue reading.](#)

4.2. Lexicale vereenvoudiging

Momenteel zijn de verschillende softwaretools beschikbaar in het middelbaar onderwijs en Simplish die in staat zijn om woordenlijsten op te maken. Kurzweil biedt daarnaast de mogelijkheid om synoniemen op te vragen voor een bepaald woord, en de gebruiker kan zelf kiezen welk woordenboek gebruikt moet worden voor het opzoeken van definities. Woordenlijsten kunnen alleen automatisch gegene-

reerd worden met behulp van GPT-taalmodellen door middel van prompts. Het gebruik van GPT-modellen voor het vereenvoudigen van teksten vereist echter wel expliciete aanwijzingen. Deze modellen houden ook rekening met woordambigüiteit en wetenschappelijk of vakjargon. Er zijn echter maar weinig softwarepakketten beschikbaar die rekening houden met ambigüiteit, idiomen en vooraf gedefinieerde achtergrondkennis. Verwante GPT-modellen zijn in staat om zinnen te vereenvoudigen terwijl ze rekening houden met deze drie aspecten. HuggingFace-taalmodellen zoals SC en BART-SC passen moeilijke woorden aan en zijn bestand tegen ambigüiteit en idiomen, maar houden geen rekening met vooraf gedefinieerde woordenschat.

4.3. Syntactische vereenvoudiging

Op dit moment zijn de erkende softwaretools in het middelbaar onderwijs niet in staat om de oorspronkelijke tekst te transformeren, waardoor syntactische vereenvoudiging niet aan de orde is. Online webtoepassingen bieden ook minder functionaliteiten om de moeilijkheidsgraad van zinsyntaxis te verlagen. Het aanpassen van tangconstructies, verwijswwoorden, voorzetseluitdrukkingen, samengestelde werkwoorden en onregelmatige werkwoorden is een uitdaging voor deze toepassingen. Het schrijven in de actieve stem kan ook problematisch zijn. Alleen vooraf gedefinieerde prompts maken het mogelijk om deze transformaties uit te voeren. Hoewel de GPT-3 taalmodellen in staat zijn om zinsyntaxtransformaties uit te voeren, kunnen ze soms problemen ondervinden bij het verwerken van alle meegegeven transformaties, en er is geen garantie dat deze modellen alle transformaties met slechts één prompt kunnen uitvoeren. Taalmodellen van HuggingFace houden minder rekening met het aanpassen van de zinsyntaxis en blijft vrijwel identiek.

4.4. Samenvatten

Op dit moment laten erkende tools in het onderwijs gebruikers toe om zinnen te markeren, waarna deze gemarkeerde zinnen aan elkaar worden geplakt. Hierdoor blijft de semantiek van de tekst gelijk, maar kan de resulterende tekst samenhang missen. Parafraseren of abstraherend samenvatten is momenteel niet mogelijk met beschikbare software in het onderwijs. Er zijn echter geavanceerde taalmodellen zoals BERT of GPT-3 die meer functionaliteiten bieden, waaronder abstraherend samenvatten op basis van gemarkeerde zinnen of woorden gekozen door de gebruiker. Experimenten met teksten hebben aangetoond dat GPT en Bing AI de nadruk leggen op het behouden van bronreferenties. Als er expliciet om wordt gevraagd, kan de Bing chatbot bronnen teruggeven die buiten het oorspronkelijke artikel te vinden zijn.

4.5. Personalisatie en verdere functionaliteiten

Vereenvoudigde zinnen of tekstinhoud met OpenAI's Codex of GPT-3 engine kan in een tabel worden gegoten of opsommingen gemaakt worden om zo de tekst overzichtelijker weer te geven. Ontwikkelaars kunnen echter door het aanspreken van deze modellen door de tabel in een structuur zoals Markdown, HTML of LaTeX op te vragen. Andere taalmodellen -en vereenvoudigingstools houden het bij doorlopende tekst, waarvan sommigen deze uitvoer opsplitsen per paragraaf en sommigen deze als één volledige paragraaf uitprinten. Achtergrondkleur, lettertype- en grootte, marge, regelafstand en spatiëring tussen leestekens aanpassen zijn onbestaand bij eender welke tool in de longlist.

4.6. Voor ontwikkelaars

Momenteel zijn er beperkingen voor erkende tekstsoftware in het onderwijs, maar er zijn taalmodellen beschikbaar op HuggingFace die tekstvereenvoudiging mogelijk maken voor Engelstalige of meertalige teksten. Het gebruikte taalmodel kan echter niet worden achterhaald en ontwikkelaars moeten rekening houden met de karakter- of tokenlimiet bij alle modellen of tools, wat het ontwerp en de ontwikkeling van software bemoeilijkt bij grote documenten. Hoewel de meeste software vrij beschikbaar is, vereisen GPT-modellen het gebruik van een API-sleutel die gekoppeld is aan een betalingsabonnement van OpenAI. Alle modellen zijn *black-box* modellen en kunnen niet duidelijk aangeven waarom een zin als moeilijk wordt bestempeld of waarom een woord als moeilijk wordt bepaald, wat overeenkomt met de bevindingen van Gooding (2022).

4.7. Requirements

Een prototype voor tekstvereenvoudiging moet intuïtiever zijn, zodat gebruikers zinnen kunnen aanduiden die ze willen vereenvoudigen op basis van parameters zoals lengte of type constructie. Personalisatie is ook belangrijk, zodat de weergave en parameters kunnen worden aangepast aan de voorkeur van de eindgebruiker. Het taalmodel en de bron van bijkomende uitleg bij woordenschat moeten duidelijk worden aangegeven in het prototype, en er moet transparantie zijn over de gebruikte prompt. De tool moet de leesbaarheid van de tekst verbeteren en begrijpelijker maken voor een breder publiek. Een prototype kan ook functies bevatten die meer inzicht geven in de tekst, zoals de moeilijk leesbare zinnen volgens leesbaarheidsformules. Om het proces te vergemakkelijken, kunnen vooraf gedefiniëerde prompts worden aangeboden als knoppen in plaats van dat gebruikers ze zelf moeten schrijven.

4.8. Conclusie

Huidige tools hebben elk een uitblinkende functionaliteit, maar er is geen manujevan-alles. Daarnaast ontbreken deze personalisatie en transparantie. De uitgeteste toepassingen gebruiken een mix tussen vrij beschikbare modellen en API's en zelfgemaakte taalmodellen die niet aangesproken kunnen worden. Promptgebaseerde toepassingen kunnen veelbelovende vereenvoudigde teksten genereren, maar er moet een intuïtieve manier worden aangeboden aan gebruikers. Zo hoeven zij niet aan prompt engineering te doen. De requirementsanalyse wijst uit dat het prototype een duidelijke opsplitsing moet maken tussen de verwachte functionaliteiten voor een scholier als voor een docent die een vereenvoudigde tekst wilt laten maken voor een scholier. Scholieren met dyslexie in de derde graad van het middelbaar onderwijs hebben nood aan een ondersteunende tool die hen toelaat om meer info rond zinnen of woorden op te halen, zodat zij de teksten beter kunnen lezen zonder dat de zinnen hun semantiek verliezen of zodat de scholieren niet de nodige kennis ontbreken zoals jargon of zinsstructuren. De docent daarentegen zal een overzicht moeten kunnen krijgen van de oorspronkelijke tekst, alsook keuzes aangereikt moeten krijgen waaraan de vereenvoudigde tekst kan voldoen. De resulterende tekst wordt in PDF of HTML-vorm aan de eindgebruiker aangereikt.

5

Prototype voor tekstvereenvoudiging

Dit hoofdstuk omschrijft de ontwikkeling van een prototype voor tekstvereenvoudiging voor scholieren met dyslexie in de derde graad van het middelbaar onderwijs. Het beantwoordt de deelvraag over hoe een intuïtieve lokale webtoepassing kan worden ontwikkeld die zowel scholieren met dyslexie als docenten helpt bij het vereenvoudigen van wetenschappelijke artikelen met behoud van semantiek, jargon en zinsstructuren. Het prototype is ontwikkeld met de benodigde functionaliteiten en eigenschappen uit de requirementsanalyse en is lokaal opgezet.

5.1. Opbouw van een prototype

Het prototype maakt gebruik van verschillende technologieën, namelijk Flask en het Jinja-framework, HTML en CSS bestanden en JavaScript. De HTML- en CSS-bestanden zijn nodig om visuele ondersteuning te bieden aan zowel lectoren als scholieren met dyslexie in de derde graad van het middelbaar onderwijs. Met behulp van JavaScript kunnen intuïtieve handelingen zoals het markeren van tekst of woorden worden verwerkt, en worden er calls gestuurd naar de Python back-end. Voordat de Flask-applicatie wordt ontwikkeld, worden de benodigde vereenvoudigingsfunctionaliteiten in Python-notebooks ontwikkeld. Dit prototype maakt gebruik van Python om handelingen zoals granulaire interactie met de taalmodellen of NLP-bibliotheken uit te voeren. Al met al zorgt deze combinatie van technologieën ervoor dat de applicatie goed kan functioneren en de gebruikers optimaal ondersteund worden.

5.2. Tekstvereenvoudiging met API

Om afwijkende resultaten op een GPT-prompt te vermijden, wordt de temperature op nul geplaatst en de *top_p* waarde wordt ingeschat op 80%. SpaCy wordt gebruikt om woordkenmerken zoals de PoS-tag op te halen, maar het systeem is vatbaar voor het niet kunnen vinden van afwisselende en meertalige woordenschat. Een mogelijke oplossing is om de taal te veranderen naar Engels of Frans, of een aangepast taalherkenningsmodel te gebruiken. Een andere optie is om de tekst voor te verwerken om de Nederlandse en Engelse woorden te scheiden voordat ze worden verwerkt met SpaCy. Adjectieven uit de tekst verwijderen is mogelijk zonder taalmodel. Aangezien alle woorden gekoppeld worden aan een PoS-tag, is het eenvoudig om de woorden gelinkt aan de span-tag van de adjectieven uit te filteren.

5.2.1. Annotaties van woordenschat

De eenduidige HTML-structuur van online woordenboeken maken het mogelijk om gratis en eenvoudig de definities van woorden op te halen. Zo is het mogelijk om annotaties op te halen zoals aangewezen in het onderzoek van Bulté e.a. (2018). Met behulp van Requests en BeautifulSoup is het mogelijk om lijsten met definities te scrapen van deze sites. De stam van het gemarkeerde woord wordt opgehaald en vervolgens meegegeven als zoekopdracht. De bron wordt samen met het resultaat aan de eindgebruiker getoond.

5.2.2. Samenvatting

Het prototype gebruikt een taalmodel van *HuggingFace* voor extraherende samenvattingen en zowel gratis taalmodellen van *HuggingFace* als het GPT-3 taalmodel voor abstraherende samenvattingen. Het model kan parameters, zoals maximale lengte van de gegenereerde tekst, ontvangen en biedt zowel gepersonaliseerde als niet-gepersonaliseerde vereenvoudiging. Het gebruik van *HuggingFace* vereist een internetverbinding en kan geen extra trainingsdata bevatten. De opstarttijd voor alle *HuggingFace*-taalmodellen wordt bij de start van de applicatie afgehandeld door middel van een extra parameter de request. Sleutels worden standaard bijgehouden in env-bestanden. Via de webtoepassing kan een gebruiker deze sleutel aanpassen. Binnen een lokale omgeving is dit in orde, al moeten ontwikkelaars rekening houden met beveiligingsmaatregelen wanneer een dergelijke tool wordt uitgerold. Het merendeel van de gebruikte taalmodellen is Engelstalig of is nadrukkelijk getraind op basis van Engelstalige datasets. De ingegeven tekst wordt eerst vertaald naar het Engels om zo de kans op een accurate vereenvoudiging te verhogen. Voor de vertaling wordt de Google Translate Python-package gebruikt. Deze is minder accuraat vergeleken met DeepL, maar biedt een gratis beschikbaar en aanvaardbaar alternatief aan. Factoren zoals topic diversity en semantische redundantie moeten overwogen worden bij het kiezen van een taalmodel voor ex-

traherend samenvatten. Lange documenten samenvatten kan zoals aangeduid in literatuurstudie door extraherende samenvatting, gevolgd door abstraherende samenvatting om de tekst coherent te doen blijken. Eerder werd er gekozen om de voltekst per paragraaf bij te houden. Uit iedere paragraaf wordt een ideaal aantal zinnen gemarkeerd om nadien geparafraseerd te worden door GPT-3 of een *HuggingFace* taalmodel, afhankelijk van de keuze van de eindgebruiker.

5.3. Tekstinhoud extraheren en uitschrijven naar PDF/DOCX

Tekst uit een PDF-bestand extraheren gebeurt met PDFMiner. Alle pagina's worden overlopen en de tekstinhoud wordt per pagina in een array geplaatst. Door middel van een aparte functie wordt de tekst opgesplitst per paragraaf en vervolgens per zin. Het resultaat van deze transformatie is een vierdimensionale array. Deze transformatie bevoordeelt het proces om vervolgens de teksten per zin op de webpagina uit te printen. De woorden in een zin worden als key-value paar opgeslaan. De sleutel verwijst hier naar de woord in een zin. De bijhorende waarde verwijst naar de PoS-tag die aan dit woord toebehoort. Dit biedt kansen toe doordat de sleutelwaarden nu overlopen kunnen worden om een klasse te koppelen aan ieder woord. Op deze manier kunnen scholieren en lectoren kiezen om zo alle werk -en naamwoorden of adjectieven te tonen met een volgens hun gekozen kleur. Het prototype maakt gebruik van Pandoc en de Python-extensie PyPandoc om tekstinhoud naar een PDF of een Word-bestand uit te schrijven. Pandoc maakt gebruik van een tweestapsbeweging waarbij de rauwe tekst eerst naar een markdown-formaat wordt omgezet. Vervolgens zet Pandoc de Markdown-bestand om naar een PDF-bestand gebouwd met de XeLaTeX engine. Lettertipes, marges en verdere metadata wordt meegegeven in een LaTeX-header. Flask kan enkel één bestand aan de gebruiker teruggeven. Als oplossing comprimeert het prototype de PDF- en Wordbestand tot één bestand.

5.4. Personaliseerbaarheid aanreiken

Voor de webtoepassing worden de standaardparameters gebruikt die uitgewezen zijn in Rello en Baeza-Yates (2013) en Rello, Baeza-Yates, Dempere-Marco e.a. (2013). Met JavaScript is het mogelijk om deze parameters dynamisch en on-the-spot aan te passen. Deze gekozen parameters worden opgeslaan als sessievariabelen, zodat de eindgebruiker niet per pagina deze parameters moet instellen. Om het uitvoerbestand personaliseerbaar te maken, worden opties in een formulier opgevraagd en vervolgens meegegeven in de Pandoc YAML-header.

5.5. Lokale opzetomgeving

Voor een optimale opzet als ontwikkelaars wordt Docker ingezet voor de deployment. Een bash of bat-scriptbestand vereenvoudigt de opstart van deze webappli-

catie in tegenstelling tot de opstart per terminal. De nodige Python-bibliotheken worden alvorens opgehaald met Pipreq. De literatuurstudie wees uit om gebruik te maken van meerdere Docker-containers wanneer taalmodellen lokaal worden opgeslaan. Alle taalmodellen worden per API aangesproken, dus één Docker-container voor de webapplicatie volstaat voor dit prototype.

5.5.1. Taalmodellen

SC en BART-SC transformeren de tekst op lexicaal en syntactisch niveau. Zij bekijken enkel de gekregen zin. Andere taalmodellen zijn eerder geneigd om extra tekst toe te voegen. Er kan niet achterhaald worden waarom dat deze extra tekst wordt meegegeven. BART-SC kan bijzaak behouden, terwijl SC sneller de neiging heeft om enkel de kernzaak te behouden in de vereenvoudigde tekst. Bij de inference API's moet er expliciet worden aangegeven om welke transformatie dit gaat door kernwoorden zoals 'summarize'.

5.6. Conclusie

Dit prototype wordt enkel binnen een lokale omgeving opgezet en is nog niet bruikbaar voor het grote publiek. Met PDF's of voltekst als invoer is het prototype in staat om teksten lexicaal en syntactisch te vereenvoudigen. Het prototype is functioneel voor zowel de doelgroep lectoren als leerlingen, twee doelgroepen die elk een andere functionaliteit prioriteren. Het prototype maakt gebruik van API's waaronder *HuggingFace* Inference APIs, Lexicala API en het GPT-3 API. Deze ontwerpkeuze bespaart geheugenruimte voor ontwikkelaars en vermindert de benodigde rekenkracht voor een prototype. Eenmaal ontwikkelaars de toepassing willen uitrollen naar het grote publiek, wordt er net zoals bij (...) aangeraden om de taalmodellen zelf te hosten. Aanvullend hierop kunnen ontwikkelaars deze modellen extra trainen op basis van de gewenste casus. Ontwikkelaars kunnen voor algemene samenvattings- en vereenvoudigingstaken gebruik maken van algemene taalmodellen die vrij beschikbaar op *HuggingFace* of dergelijke platforms terug te vinden zijn. GPT-3 blinkt uit in gepersonaliseerde vereenvoudigings- en samenvattingstaken. Engelstalige prompts die expliciet de uitvoertaal vermelden zijn nauwkeuriger dan Nederlandstalige prompts. Ontwikkelaars moeten rekening houden met het gebrek aan structuur bij het ophalen van tekstinhoud uit een PDF-bestand.

6

Vergelijkende studie

Dit hoofdstuk onderzoekt de verschillen tussen de vereenvoudigde tekstinhoud van taalmodellen en toepassingen rond tekstvereenvoudiging. Zoals aangegeven in het onderzoek van Nenkova en Passonneau (2004) worden leesgraadsformules en referentieteksten ingezet om de kwaliteit van een vereenvoudigde tekst te beoordelen. De vergelijkende studie beantwoordt de deelvraag 'Welke handmatige tekstvereenvoudigingstechnieken zijn niet terug te vinden in geautomatiseerde tekstvereenvoudiging?'.

6.1. Vergelijking met referentieteksten

6.1.1. Methodologie

In dit onderzoek worden twee wetenschappelijke artikelen vergeleken met een referentietekst die handmatig werd vereenvoudigd door twee lectoren en twee scholieren. De auteurs van de referentietekst volgden daarbij een document met richtlijnen waaraan zij zich moesten houden. De vereenvoudigde teksten worden afgetoetst aan de hand van de richtlijnen die voortvloeien uit de literatuurstudie, om zo antwoord te geven op de twee deelvragen van het onderzoek.

Gekozen parameters

Voor de webtoepassingen SciSpace, Resoomer en Simplext werden de oorspronkelijke documenten geüpload om een vereenvoudigde versie terug te krijgen, die in de vorm van een PDF werd opgeslagen. Gerelateerde GPT-tools en taalmodellen bieden deze functie niet aan, dus in dat geval wordt de tekstinhoud aan het model gegeven en wordt het resultaat opgeslagen als een tekstbestand. Resoomer, Simplext en Scispace maken gebruik van sliders om de tekst in te korten. Deze ratio wordt geplaatst op 60%. Bij het gebruik van het GPT-3 model worden vijf prompts gebruikt en om rekening te houden met de tokenlengte, wordt de tekst tussen de subtitels opgebroken. De prompts zijn zoals volgt opgesteld:

- P1: Simplify this text in Dutch /// Paragraaf
- P2: Simplify a text for economics students (16-18 years old) by replacing difficult words, keeping technical jargon, replacing words longer than 18 letters, writing acronyms in full, replacing a word only once with a synonym, giving brief explanations when necessary, and avoiding percentages. /// Paragraaf
- P3: Simplify a text by breaking them into shorter sentences with a maximum of ten words. Change pronouns like 'they', 'their', or 'he' to names. Replace complex sentence constructions and prepositional phrases with simpler alternatives, but leave them unchanged if no simpler option is available. /// Paragraaf
- P4: Simplify an article for economics students (16-18 years old) by replacing difficult words (except technical jargon learned in the 2nd grade), words longer than 18 letters, and avoiding percentages. Use synonyms only once and give brief explanations if necessary. Write acronyms in full and replace pronouns with names. Simplify sentence constructions and prepositional phrases, splitting sentences into a maximum of ten words, but leave them unchanged if no simpler alternative is available. /// Paragraaf

6.2. Resultaten

6.2.1. Lexicale vereenvoudiging

Woordenschat

De woordenschat van de referentieteksten is geschaald naar de verwachte woordkennis van een scholier met dyslexie in de derde graad van het middelbaar onderwijs. De twee gefinetuneerde taalmodellen van HuggingFace gebruiken een algemene doelgroep en daarbij wordt vakjargon niet veranderd. Alle GPT-prompts buiten P1 houden rekening met het vakjargon en vervangen deze woorden door een eenvoudiger synoniem. P1 past alle woordenschat aan en kan zo reeds gekende informatie opnieuw vereenvoudigen, waardoor reeds gekend vakjargon weg wordt gelaten. Anderstalige woordenschat in de referentieteksten wordt vertaald naar het Nederlands om een eentalige tekst te bekomen. Verwante GPT-3 modellen doen dit enkel indien expliciet aangewezen. Indien niet expliciet aangewezen, kan GPT-3 deze tekst opnieuw schrijven naar een eentalige tekst. Overige tools en taalmodellen nemen deze woorden over.

Acroniemen

Acroniemen worden voluit geschreven in de referentieteksten. GPT-3 en SciSpace interpreteren deze acroniemen correct en er zijn geen uitzonderingen bij deze twee wetenschappelijke artikelen. Andere tools en taalmodellen zijn niet in staat om acroniemen voluit te schrijven. Het acroniem wordt letterlijk overgenomen in de vereenvoudigde tekst.

Interpretatie van cijfers

Het vereenvoudigd referentieartikel binnen het economisch vakdomein houdt rekening met cijferwaarden. Alle taalmodellen, prompts en webtoepassingen hebben moeite bij het interpreteren van de cijferwaarden. Deze cijfers worden overgenomen in een vereenvoudigde zin, maar de modellen staan niet stil bij de interpretatie van dergelijke cijfers.

6.2.2. Syntactische vereenvoudiging

De referentieteksten splitsen lange zinnen en zinnen met een complexe zinsyntax op om twee of meerdere subzinnen te bekomen. Tangconstructies en het aanpassen van verwijs- en voornaamwoorden gebeurt echter niet bij de HuggingFace-taalmodellen, Resoomer, SciSpace en Simplext. De verwijs- en voornaamwoorden worden behouden zoals ze in de oorspronkelijke tekst staan geschreven. Enkel P3 houdt rekening met deze constructies en past deze aan in de vereenvoudigde tekst.

6.2.3. Samenvatten

Formaatwijzigingen

Formaatwijzigingen, zoals het omzetten van tekstinhoud naar een tabelvorm of het gebruik van opsommingen, zijn present in de referentieteksten en deze techniek is enkel mogelijk met GPT-3. SciSpace past het formaat enkel visueel aan en de uitvoer bestaat uit een opsomming van zinnen, maar de zinsyntax is niet opgebouwd zoals dit als een opsomming zou geschreven zijn. De andere tools zijn niet in staat om het formaat van de tekst aan te passen.

Lengte van het vereenvoudigde artikel

Alle vereenvoudigde teksten, buiten het resultaat van P4, zijn korter dan de oorspronkelijke tekst. Handmatig vereenvoudigde teksten zijn significant korter dan de oorspronkelijke teksten. De lengte van de tekst bij P1, P2 en P3 is opmerkelijk korter dan de lengte van de tekst in P4.

Bibliografie en citeren

De referentieteksten maken gebruik van een korte bibliografie en houden zo rekening met de verwijzingen naar de oorspronkelijke bron met de identieke vermelding (APA of Chicago). GPT-3 is in staat om deze verwijzingen correct te verwerken. Andere tools en taalmodellen hebben moeite bij het verwerken van verwijzingen naar bronnen en nemen enkel de de ronde of vierkante haken over die de verwijzing definieert.

Behoud van semantiek in de vereenvoudigde tekst

De kerngedachte is bij alle vereenvoudigde versies van de wetenschappelijke artikelen terug te vinden. Deelvragen kunnen bij de handmatig vereenvoudigde arti-

kelen, het vereenvoudigde artikel van SciSpace en bij de GPT-prompts P2, P3 en P4 volledig worden beantwoord. Resoomer, Simplext en de HuggingFace-taalmodellen ontbreken de inhoud om op vijf van de twaalf zinnen een antwoord te kunnen geven. Algemene informatie zoals de inleiding en de conclusie zijn echter wel terug te vinden.

6.3. Vergelijking zonder referentieteksten

Voor de overgebleven wetenschappelijke artikelen zijn er geen referentieteksten opgesteld. Deze teksten kunnen zoals aangegeven beoordeeld worden met beschikbare leesbaarheidsformules. De volgende teksten worden beoordeeld met de Flesch-Kincaid en Flesch-Reading-Ease. Daarnaast worden de teksten ook bekeken of deze de kern- en bijzaken ook behouden. Voor deze statistieken zijn er drie hypothesen opgesteld:

- Hypothese 1: De leesgraad van automatisch vereenvoudigde wetenschappelijke artikelen hebben een significant eenvoudigere leesgraad vergeleken met de oorspronkelijke wetenschappelijke artikelen.
- Hypothese 2: De leesgraad van automatisch vereenvoudigde wetenschappelijke artikelen zijn minstens even eenvoudig dan handmatig vereenvoudigde artikelen.
- Hypothese 3: De gebruikte syntactische structuren in een automatisch vereenvoudigd wetenschappelijk artikel zijn minstens even eenvoudig als het aantal in een handmatig vereenvoudigd wetenschappelijk artikel.

6.4. Conclusie

Het kiezen van een prompt vereist een gerichte schrijfwijze. Prompts personaliseerbaar maken, helpt het model ook om een afweging te maken bij het vereenvoudigen van een tekst. Prompts worden gelimiteerd door de tokenlengte en het aantal meegegeven parameters beïnvloed het aantal tokens die kunnen meegegeven worden. GPT-3 kan woordenschat vereenvoudigen zonder een expliciete vermelding, maar het taalmodel heeft nog steeds nood aan afbakening.

Handmatige tekstvereenvoudigingstechnieken zoals het transformeren van de oorspronkelijke structuur naar een formaat dat beter aansluit op de noden van scholieren in de derde graad van het middelbaar onderwijs met dyslexie. Het ontbreken van detailinformatie bij de resultaten of conclusie kan een rechtstreeks effect hebben op de interpretatie van het wetenschappelijk artikel. De verschillen tussen bestaande leesbaarheidsformules bij GPT-3 modellen zijn merkbaar. Daarnaast kunnen taalmodellen moeite ondervinden bij het behouden van citaties en verwijzingen naar de bronnen.

7

Conclusie

Software ontwikkelen voor scholieren met dyslexie in de derde graad van het middelbaar onderwijs.

De erkende software uitgeleend aan scholieren met dyslexie in de derde graad van het middelbaar onderwijs voldoet niet aan de noden. De software biedt ondersteunende functionaliteiten aan zoals het aanmaken van een woordenlijst, alsook het markeren van zinnen om deze later om te vormen naar een tekst. Syntactische vereenvoudiging of abstraherende samenvatting zijn niet tot de orde. Online toepassingen staan verder en reiken functionaliteiten aan die zowel hoogstaand als effectief te reproduceren zijn. Echter is er geen manusje-van-alles en er is nood aan een toepassing die alle functionaliteiten kan combineren. Echter kunnen deze softwarepakketten opgeschaald worden door middel van LLM's, zodat deze de functionaliteiten hebben om verbeterde tekstvereenvoudigingstechnieken aan te reiken aan scholieren met dyslexie in de derde graad van het middelbaar onderwijs.

Het prototype wijst uit dat het combineren van functionaliteiten een haalbare zaak is voor zowel klein- als grootschalige softwareondernemingen. Met gebruik van kant-en-klare taalmodellen, API's en gekende programmeertalen zijn ontwikkelaars ertoe in staat om een webtoepassing te ontwikkelen die ondersteuning kan bieden aan scholieren met dyslexie in de derde graad van het middelbaar onderwijs. Dit prototype is gebouwd op eerder onderzochte visuele kenmerken en de impact van een vereenvoudigde tekst op de leessnelheid en -begrip bij een scholier met dyslexie tijdens het intensief lezen van een tekst. Het prototype werd niet uitgetest bij het doelpubliek en vereist verder uittesten.

Onderzoek naar het verschil tussen het laten schrijven van prompts en vooraf gedefinieerde prompts is schaars, maar deze keuze kan een effect hebben op het gedrag of ervaring van de eindgebruiker. De doelgroep wordt expliciet aangeduid in de prompts en is daarmee parameteriseerbaar. Er is verder onderzoek nodig

naar de effecten op het meegeven van doelgroepen via prompts en of deze al dan niet rekening houden met de doelgroep.

Verschillen tussen de handmatige tekst, het origineel, uitvoer van beschikbare tools en het prototype.

Uit een vergelijkende studie blijkt dat er lichte verschillen zijn tussen de verschillende tools wat betreft de scores op het gebied van FRE, SMOG en Kincaid. Het prototype en het GPT-3-model dat wordt gebruikt voor gepersonaliseerde samenvattingen scoren beter op leesbaarheid. Een niet-gepersonaliseerde samenvatting scoort echter lager en heeft problemen in de doorlopende tekst, waarbij het taalmodel soms woorden plaatst die niet nuttig zijn in de context. Resoomer en Scispace verkorten effectief. Hierdoor wordt de bruikbaarheid van deze scores en het gebruik van de Python-bibliotheek in twijfel getrokken. Hoewel deze scores een goed alternatief zijn om de leesbaarheid te meten, houden ze geen rekening met verkeerd geïnterpreteerde resultaten, zoals letterlijk overgenomen bronvermeldingen of verkeerd gegenereerde woordenschat door het taalmodel. Deze fouten zijn minder vaak aanwezig bij het vereenvoudigen van teksten met GPT-3. Er is meer onderzoek nodig om de bruikbaarheid van deze scores te bepalen en om te begrijpen hoe de scores zich verhouden tot de kwaliteit van de vereenvoudigde tekst.

Het vergelijken met een referentietekst blijft nog steeds een handmatige vergelijking en biedt een inkijk in hoe lectoren teksten kunnen vereenvoudigen. De vergelijkende studie hield geen rekening met het hoofdstuk waarin een zin werd beoordeeld. Vragen naar het verschil na een tekstvereenvoudiging per hoofdstuk in een wetenschappelijk artikel kan daarmee niet beantwoord worden en moet opgevolgd worden in een verder onderzoek.

Het gebruik van HuggingFace-taalmodellen en LLM's voor een toepassing van tekstvereenvoudiging.

Bestaande taalmodellen maken de ontwikkeling van toepassingen op het gebied van semantische analyse, kernwoordenidentificatie en het extraheren van samenvattingen veel eenvoudiger. Een prototype heeft aangetoond dat vrij beschikbare taalmodellen, zoals BERT en gerelateerde BERT-varianten, een oplossing bieden voor deze taken. Voor meer complexe vereenvoudigingstechnieken, zoals lexicale en syntactische vereenvoudiging en abstraherende samenvatting, zijn complexere taalmodellen zoals GPT-3 geschikt. Ontwikkelaars moeten echter rekening houden met de schaal van de modellen bij het maken van deze keuze. Hoewel vrij beschikbare modellen op HuggingFace in staat zijn om abstraherende samenvattingen of lexicale vereenvoudiging mogelijk te maken, staan ze in de schaduw van GPT-3, dat voor ontwikkelaars vrij beschikbaar is.

Omgekeerd is GPT-3 echter een overkill voor taken zoals het identificeren van kernwoorden of het aanduiden van belangrijke zinnen, die kosteneffectief kunnen wor-

den aangepakt zonder het gebruik van GPT-3. GPT-3 moet niet voor iedere functionaliteit worden ingezet om zo kostenbesparend te werken. Het GPT-3 model maakt complexe en granulaire NLP-transformaties op lexicaal en syntactisch niveau mogelijk voor geautomatiseerde tekstvereenvoudiging. Echter houdt het model geen rekening met referenties buiten de getrainde data, wat tot problemen bij de credibiliteit van data kan leiden. Bing AI daarentegen doet dit wel en vormt een goede fundering voor ontwikkelaars om meer referentiemateriaal aan te bieden in ondersteunende software binnen het onderwijs. Verder onderzoek op de toepassing van deze AI via een mogelijke API is zeker nodig en kan baanbrekend zijn voor de onderwijssector.

Verdere finetuning en pre-training van taalmodellen.

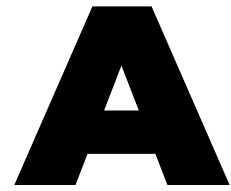
Het huidige prototype maakt geen gebruik van verdere finetuning en het GPT-3 model dat wordt gebruikt, is niet vooraf getraind op extra data van wetenschappelijke papers. Echter, er is wetenschappelijke data beschikbaar die vrij beschikbaar is en kan worden gebruikt om het GPT-3 taalmodel accurater te maken op interpretatie van complexiteit bij wetenschappelijke artikelen. Ondanks dit, is er wel een licht effect waargenomen op de verschillen in lexicale complexiteit tussen de HuggingFace-taalmodellen die wel getraind zijn op wetenschappelijke papers in vergelijking met taalmodellen die getraind zijn op algemene data.

Om de verschillen tussen deze taalmodellen binnen de context van wetenschappelijke papers beter te begrijpen, is meer onderzoek nodig. Het is belangrijk om op te merken dat de taalmodellen van OpenAI continu evolueren. Het verschil tussen GPT-2 en GPT-3 via de API is merkbaar en daarom wordt er overwogen om GPT-2 achterwege te laten, met oog op verdere edities van OpenAI's GPT-modellen. Op dit moment wordt GPT-4 uitgerold en Bing AI is beschikbaar volgens een wachtlijst. Echter, op dit moment zijn deze twee taalmodellen nog niet in staat om in productie te worden gebruikt. Daarom is verder onderzoek nodig naar het gebruik van deze modellen binnen het onderwijs.

Deployment van het prototype en andere toepassingen voor tekstvereenvoudiging.

Dit onderzoek heeft zich beperkt tot de tekstinhoud van een vereenvoudigd artikel. Er is echter meer onderzoek nodig naar hoe de inzet van webtoepassingen en browserextensies voor tekstvereenvoudiging in het onderwijs kan worden verbeterd. Met alsmaar grotere taalmodellen, zoals het opkomende GPT-4 en LLaMa, is er ook meer onderzoek nodig naar de verschillen op taalvlak ten opzichte van de toename in parameters. Het GPT-3 model dat in dit onderzoek werd gebruikt, maakte enkel gebruik van aangepaste parameters zoals de *temperature* en *top_p*. Hoewel de overstap naar andere taalmodellen kostelijk kan zijn voor ontwikkelaars, is het belangrijk om te onderzoeken of en hoe deze nieuwe modellen kunnen bijdragen aan betere resultaten.

Het is echter eerder uitgewezen dat de grootte van taalmodellen alsmaar minder relevant wordt. Een andere uitdaging die aan het licht kwam tijdens het onderzoek is de kwaliteit van de geëxtraheerde tekstinhoud uit een PDF- of Docx-bestand door de gebruikte Python-libraries. Hoewel LLM's rekening houden met mogelijke noise, kunnen woorden tot zinnen ontbreken bij het ophalen van tekst uit een PDF afhankelijk van het formaat waarin het is opgeslagen. Om deze uitdaging aan te pakken, is onderzoek nodig naar robuuste oplossingen om PDF-inhoud uit een tekst te extraheren, met minimale invloed van het gebruikte softwarepakket.



Onderzoeksvoorstel

Samenvatting

Ingewikkelde woordenschat en zinsbouw hinderen scholieren met dyslexie in het derde graad middelbaar onderwijs bij het lezen van wetenschappelijke artikelen. Adaptieve tekstvereenvoudiging helpt deze scholieren bij hun lees- en verwerkingsnelheid. Daarnaast kan artificiële intelligentie (AI) dit proces automatiseren om de werkdruk bij leraren en scholieren te verminderen. Dit onderzoek achterhaalt met welke technologische en logopedische aspecten AI-ontwikkelaars rekening moeten houden bij de ontwikkeling van een AI-toepassing voor adaptieve en geautomatiseerde tekstvereenvoudiging. Hiervoor is de volgende onderzoeksvraag opgesteld: "Hoe kan een wetenschappelijk artikel automatisch worden vereenvoudigd, gericht op de unieke noden van scholieren met dyslexie in het derde graad middelbaar onderwijs?". Een vergelijkende studie beantwoordt deze onderzoeksvraag en is uitgevoerd met bestaande toepassingen en een prototype voor adaptieve en geautomatiseerde tekstvereenvoudiging. Uit de vergelijkende studie blijkt dat toepassingen om wetenschappelijke artikelen te vereenvoudigen, gemaakt zijn voor een centrale doelgroep en geen rekening houden met de unieke noden van een scholier met dyslexie in het derde graad middelbaar onderwijs. Adaptieve software voor geautomatiseerde tekstvereenvoudiging is mogelijk, maar ontwikkelaars moeten meer inzetten op de unieke noden van deze scholieren.

A.1. Introductie

Het Vlaams middelbaar onderwijs staat op barsten. Leraren en scholieren worden overspoeld door werkdruk en stress. Bovendien is de derde graad van het middelbaar onderwijs een belangrijke mijlpaal voor de verdere loopbaan van scholieren, al hebben zij volgens Dapaah en Maenhout (2022) dan moeite om grip te krijgen

op de vakliteratuur bij STEM-vakken. Het STEM-agenda¹ van de Vlaamse Overheid moet het STEM-onderwijs tegen 2030 aantrekkelijker te maken, door de ondersteuning voor zowel leerkrachten als scholieren te verbeteren. Toch wordt het aanpakken van de steeds complexere wetenschappelijke taal, zoals beschreven in Barnett en Doubleday (2020), niet opgenomen in het STEM-agenda. Wetenschappelijke artikelen vereenvoudigen, op maat van de noden voor een scholier met dyslexie in het middelbaar onderwijs, is tijds- en energie-intensief voor leerkrachten en scholieren. Automatische en adaptieve tekstvereenvoudiging biedt hier een baanbrekende oplossing om de werkdruk in het middelbaar onderwijs te verminderen. Het doel van dit onderzoek is om te achterhalen met welke technologische en logopedische aspecten AI-ontwikkelaars rekening moeten houden bij de ontwikkeling van een adaptieve AI-toepassing voor geautomatiseerde tekstvereenvoudiging. De volgende onderzoeksvraag is opgesteld: "Hoe kan een wetenschappelijke artikel automatisch vereenvoudigd worden, gericht op de verschillende behoeften van scholieren met dyslexie in de derde graad middelbaar onderwijs?". Het doel wordt bereikt door een antwoord op de volgende deelvragen te formuleren. Eerst geeft de literatuurstudie een antwoord op de eerste vier deelvragen. Daarna vormt het veldonderzoek een antwoord op de vijfde deelvraag. Ten slotte beantwoordt de vergelijkende studie de zesde en laatste deelvraag. De resultaten van dit onderzoek zetten AI-ontwikkelaars aan om een toepassing te maken om scholieren met dyslexie te kunnen ondersteunen in de derde graad middelbaar onderwijs.

1. Welke aanpakken zijn er voor geautomatiseerde tekstvereenvoudiging? Aansluitende vraag: "Hoe worden teksten handmatig vereenvoudigd voor scholieren met dyslexie?"
2. Welke specifieke noden hebben scholieren van de derde graad middelbaar onderwijs bij het begrijpen van complexere teksten?
3. Wat zijn de specifieke kenmerken van wetenschappelijke artikelen?
4. Met welke valkuilen bij taalverwerking met AI moeten ontwikkelaars rekening houden?
5. Welke toepassingen, tools en modellen zijn er beschikbaar om Nederlandstalige geautomatiseerde tekstvereenvoudiging met AI mogelijk te maken?
6. Welke functies ontbreken AI-toepassingen om geautomatiseerde én adaptieve tekstvereenvoudiging mogelijk te maken voor scholieren met dyslexie in de derde graad middelbaar onderwijs? Aansluitende vraag: "Welke manuele methoden voor tekstvereenvoudiging komen niet in deze tools voor?"

¹<https://www.vlaanderen.be/publicaties/stem-agenda-2030-stem-competenties-voor-een-toekomst-en-missiegericht-beleid>

A.2. State-of-the-art

A.2.1. Tekstvereenvoudiging

De voorbije tien jaar is artificiële intelligentie (AI) sterk verder ontwikkeld. Vasista (2022) benadrukt dat de toename in kennis voor nieuwe toepassingen zorgde. Tekstvereenvoudiging vloeide hier uit voort. Momenteel bestaan er al robuuste toepassingen die teksten kunnen vereenvoudigen, zoals Resoomer², Paraphraser³ en Prepostseo⁴. Binnen het kader van tekstvereenvoudiging is er bestaande documentatie beschikbaar waar onderzoekers het voordeel van toegankelijkheid aanhalen, maar volgens Gooding (2022) ontbreken deze toepassingen de extra noden die scholieren met dyslexie in de derde graad middelbaar onderwijs vereisen.

Shardlow (2014) haalt aan dat het algemene doel van tekstvereenvoudiging is om ingewikkelde bronnen toegankelijker te maken. Het zorgt voor verkorte teksten zonder de kernboodschap te verliezen. Siddharthan (2014) haalt verder aan dat tekstvereenvoudiging op één van drie manieren gebeurt. Er is conceptuele vereenvoudiging waarbij documenten naar een compacter formaat worden getransformeerd. Daarnaast is er uitgebreide modificatie die kernwoorden aanduidt door gebruik van redundantie. Als laatste is er samenvatting die documenten verandert in kortere teksten met alleen de topische zinnen. Met deze concepten zijn ontwikkelaars volgens Siddharthan (2014) in staat om ingewikkelde woorden te vervangen door eenvoudigere synoniemen of zinnen te verkorten zodat ze sneller leesbaar zijn.

Tekstvereenvoudiging behoort tot de zijtak van *Natural Language Processing* (NLP) in AI. NLP omvat methodes om menselijke teksten om te zetten in tekst voor machines. Documenten vereenvoudigen met NLP kan volgens Chowdhary (2020) op twee manieren: extract of abstract. Bij extractieve vereenvoudiging worden zinnen gelezen zoals ze zijn neergeschreven. Vervolgens bewaart een document de belangrijkste taalelementen om de tekst te kunnen hervormen. Deze vorm van tekstvereenvoudiging komt volgens (Sciforce, 2020) het meeste voor. Daarnaast is er abstracte vereenvoudiging waarbij de kernboodschap wordt bewaard. Met de kernboodschap wordt er een nieuwe zin opgebouwd. Volgens het onderzoek van Chowdhary (2020) heeft deze vorm potentieel, maar het zit nog in de kinderschoenen.

A.2.2. Noden van scholieren met dyslexie

Het experiment van Franse wetenschappers

Gala en Ziegler (2016) illustreert dat manuele tekstvereenvoudiging schoolteksten toegankelijker

maakt voor kinderen met dyslexie. Dit deden ze door simpelere synoniemen en

²<https://resoomer.com/nl/>

³<https://www.paraphraser.io/nl/tekst-samenvatting>

⁴<https://www.prepostseo.com/tool/nl/text-summarizer>

zinsstructuren te gebruiken. Tien kinderen werden opgenomen in het experiment, variërend van 8 tot 12 jaar oud. Verwijswoorden werden vermeden en woorden kort gehouden. De resultaten waren veelbelovend. Het leestempo lag hoger en de kinderen maakten minder leesfouten. Ook bleek er geen verlies van begrip in de tekst bij geteste kinderen. Resultaten van de studie werden gebundeld voor de mogelijke ontwikkeling van een AI-tool.

De visuele weergave van tekst beïnvloedt de leessnelheid bij scholieren met dyslexie. Zo haalt het onderzoek van Rello e.a. (2012) tips aan waarmee teksten en documenten rekening moeten houden bij scholieren met dyslexie in de derde graad middelbaar onderwijs. Het gaat over speciale lettertypes, spreiding tussen woorden en het gebruik van inzoomen op aparte zinnen. Het onderzoek haalt verder aan dat teksten voor deze unieke noden aanpassen tijdrovend is, dus tekstvereenvoudiging door AI kan een revolutionaire oplossing bieden. De Universiteit van Copenhagen is met bovenstaande idee aan de slag gegaan. Onderzoekers Bingel e.a. (2018) hebben gratis software ontwikkeld, genaamd Hero⁵, om tekstvereenvoudiging voor scholieren in het middelbaar onderwijs met dyslexie te automatiseren. De software bestudeert met welke woorden de gebruiker moeite heeft, en vervangt die door simpelere alternatieven. Hero bevindt zich nu in beta-vorm en wordt enkel in het Engels en Deens ondersteund. Als alternatief is er Readable⁶. Dit is een Engelstalige AI-toepassing dat zinnen beoordeeld met leesbaarheidsformules.

Roldós (2020) haalt aan dat NLP in de laatste decennia volop in ontwikkeling is, maar ontwikkelaars botsen nog op uitdagingen. Het gaat om zowel interpretatie- als dataproblemen bij AI-modellen. Het onderzoek haalt twee punten aan. Allereerst is het voor een machine moeilijk om de context van homoniemen te achterhalen. Bijvoorbeeld bij het woord 'bank' is het niet duidelijk voor de machine of het gaat over de geldinstelling of het meubel. Daarnaast zijn synoniemen een probleem voor tekstverwerking.

Het onderzoek van Sciforce (2020) haalt aan dat het merendeel van NLP-toepassingen Engelstalige invoer gebruikt. Niet-Engelstalige toepassingen zijn zeldzaam. De opkomst van AI technologieën die twee datasets gebruiken, biedt een oplossing voor dit probleem. De software vertaalt eerst de oorspronkelijke tekst naar de gewenste taal, voordat de tekst wordt herwerkt. Hetzelfde onderzoek bewijst dat het vertalen van gelijkaardige talen, zoals Duits en Nederlands, een minimaal verschil opleverd. Volgens Plavén-Sigray e.a. (2017) houden onderzoekers zich vaak in hun eigen taalbubbel, wat negatieve gevolgen heeft voor de leesbaarheid van een wetenschappelijk artikel. Bovendien vormt de stijgende trend van het gebruik aan acroniemen Barnett en Doubleday (2020) een extra hindernis. Donato e.a. (2022) haalt aan dat onbegrijpelijke literatuur, waaronder studiemateriaal geschreven door de docent en online wetenschappelijke artikelen, één van de redenen is waarom scholieren

⁵<https://beta.heroapp.ai/>

⁶<https://readable.com/>

met dyslexie in het middelbaar onderwijs van richting veranderen.

A.2.3. Huidige toepassingen

Vlaanderen heeft weinig zicht op de geïmplementeerde AI software in scholen. Dit werd vastgesteld door (Martens e.a., 2021a), een samenwerking tussen de Vlaamse universiteiten en overheid voor AI. Vergeleken met andere Europese landen, maakt België het minst gebruik van leerling-georiënteerde hulpmiddelen. Degenen die wel gebruikt worden, zijn vooral online leerplatformen voor zelfstandig werken. Ook maakt België amper gebruik van beschikbare software die de leermethoden en -noden van leerlingen evalueert (Martens e.a., 2021b).

Verhoeven (2023) haalt aan dat AI-toepassingen zoals ChatGPT, Google Bard en Bing AI kunnen helpen om routinematig werk te verminderen in het onderwijs. Echter haalt Deckmyn (2021) aan dat GPT-3, het model van ChatGPT, sterker staat in het maken van Engelstalige teksten vergeleken met Nederlandstalige teksten. De databank waar het GPT-3 model mee is getraind, bestaat uit 92% Engelstalige woorden, terwijl er 0,35% Nederlandse woorden aanwezig zijn in dezelfde databank. Ontwikkelaars moeten de evolutie van deze modellen opvolgen, voordat er Nederlandstalige toepassingen mee worden gemaakt.

A.2.4. Ontwikkelen met AI

Python staat bovenaan de lijst van programmeertalen voor NLP-toepassingen. Volgens het onderzoek van Thangarajah (2019) is dit te wijten aan de eenvoudige syntax, kleine leercurve en grote beschikbaarheid van kant-en-klare bibliotheken. Wetenschappelijke berekeningen of statistische analyses kunnen worden uitgevoerd met één lijn code. Malik (2022) haalt de twee meest voorkomende aan, namelijk NLTK⁷ en Spacy⁸. Deep Martin⁹ bouwt verder op het onderzoek van Shardlow (2014) naar een pipeline voor lexicale vereenvoudiging. Deep Martin maakt gebruik van *custom transformers* om invoertekst te converteren naar een vereenvoudigde versie van de tekstinhoud.

Voor Germaanse talen zijn er enkele datasets en word embeddings beschikbaar die de complexiteit van woorden bijhouden. Zo zijn er in de Duitse taal Klexikon¹⁰ en TextComplexityDE¹¹. Een onderzoek van Suter e.a. (2016) bouwde een rule-based NLP-model met 'Leichte Sprache', wat een dataset is met eenvoudige Duitstalige zinsconstructies. Nederlandstalige datasets zijn in schaarse hoeveelheden beschikbaar, dus het vertalen uit een Germaanse taal is hier een optie.

Volgens Garbacea e.a. (2021) is het belangrijk dat AI-ontwikkelaars niet alleen aandacht besteden aan het aanpassen van woorden en zinnen, maar ook aan de ge-

⁷<https://www.nltk.org/>

⁸<https://spacy.io/>

⁹<https://github.com/chrislemke/deep-martin>

¹⁰<https://github.com/dennlinger/klexikon>

¹¹<https://github.com/babaknaderi/TextComplexityDE>

bruiker meegeven waarom iets is aangepast. De onderzoekers wijzen op twee ethische aspecten. Eerst moet de toepassing duidelijk aangeven waarom een woord of zin is aangepast. Het model moet de moeilijkheidsgraad van de woorden of zinnen bewijzen. Iavarone e.a. (2021) beschrijft een methode met regressiemodellen om de moeilijkheidsgraad te bepalen door een gemiddeld moeilijkheidspercentage per zin te berekenen. Daarnaast benadrukt Garbacea e.a. (2021) het belang van het markeren van de complexere delen van een tekst. Hiervoor haalt hetzelfde onderzoek methoden aan zoals *lexical* of *deep learning*.

Er is een tactvolle aanpak nodig om een vereenvoudigde tekst met AI te beoordelen. De studie van Swayamdipta (2019) haalt aan dat er extra nood is aan NLP-modellen waarbij de tekst zijn kernboodschap behoudt. Samen met Microsoft Research bouwden ze NLP-modellen die gericht waren op de bewaring van zinsstructuur en -context door *scaffolded learning*. Hiervoor maakten de onderzoekers gebruik van een voorspellingsmethode die de positie van woorden en zinnen in een document beoordeelde. De Flesch-Kincaid leesbaarheidstest is volgens

Readable (2021) een alternatieve manier om vereenvoudigde tekstinhoud te beoordelen, zonder de nood aan *pre-trained* modellen. Deze score kan eenvoudig worden berekend met de *Python-library textstat*¹².

A.3. Methodologie

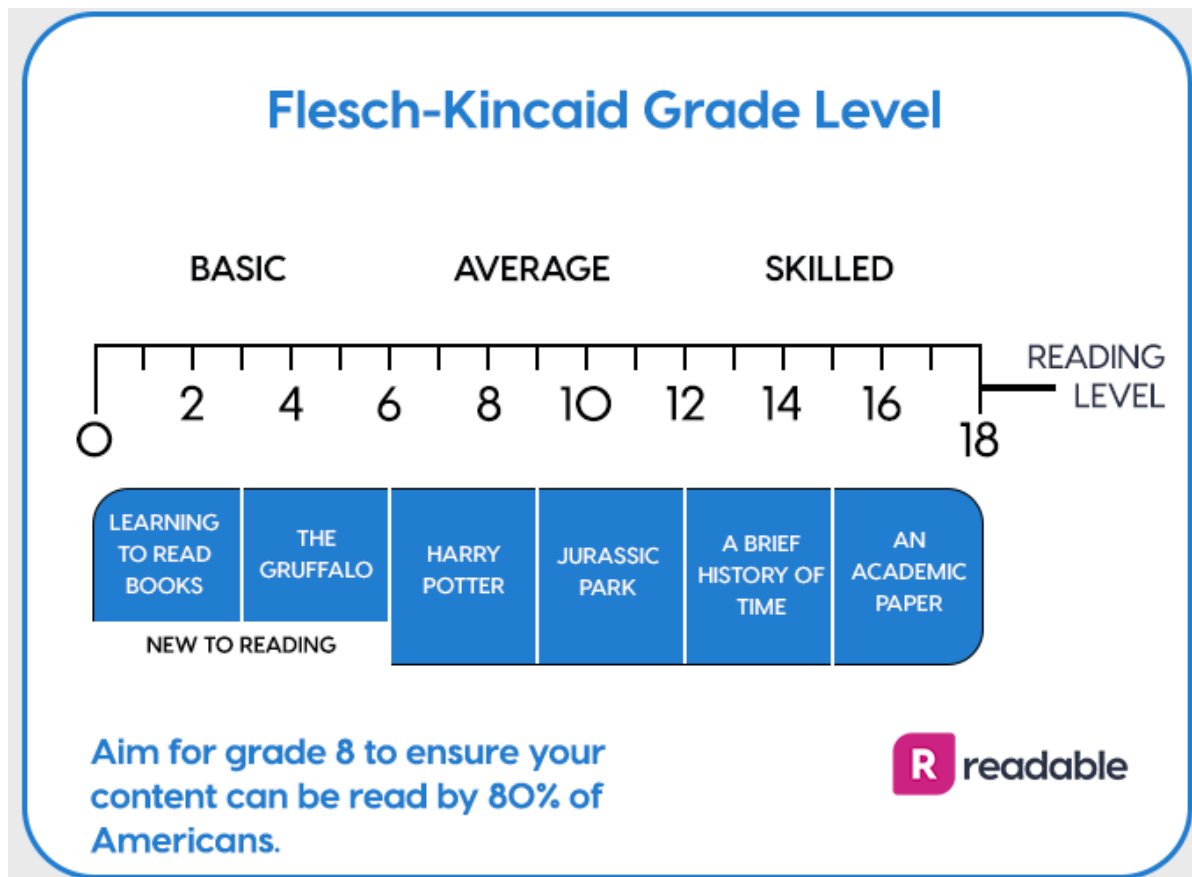
Een *mixed-methods* onderzoek toont aan hoe toepassingen automatisch een wetenschappelijke artikel kunnen vereenvoudigen, gericht op scholieren met dyslexie in de derde graad middelbaar onderwijs. Het onderzoek houdt vijf grote fases in. De eerste fase is het proces van geautomatiseerde tekstvereenvoudiging beschrijven. Dit gebeurt via een grondige studie van vakliteratuur en wetenschappelijke teksten. Ook blogs van experts komen hier aan bod. Na het verwerven van de nodige inzichten wordt er een verklarende tekst opgesteld.

De tweede fase bestaat uit het analyseren van wetenschappelijke werken over de bewezen voordelen van tekstvereenvoudiging bij scholieren met dyslexie van de derde graad middelbaar onderwijs. Hiervoor zijn geringe thesissen beschikbaar, die zorgvuldigheid vragen tijdens interpretatie. De resulterende tekst bevat de voordelen samen met hun wetenschappelijke onderbouwing.

De derde fase is opnieuw een beschrijving. Hier worden de valkuilen bij taalverwerking met AI-software nagegaan. Deze fase van het onderzoek brengt mogelijke nadelen en tekortkomingen van AI-software bij tekstvereenvoudiging aan het licht. Dit gebeurt aan de hand van een technische uitleg.

De vierde fase omvat een toelichting over beschikbare AI toepassingen voor tekstvereenvoudiging. Aan de hand van een veldonderzoek op het internet en bij bedrijven wordt een longlist opgesteld van beschikbare toepassingen voor tekstvereenvoudiging in het middelbaar onderwijs. Met een requirementsanalyse wordt

¹²<https://pypi.org/project/textstat/>



Figuur (A.1)
(Readable, 2021)

er een shortlist opgesteld van software. Het toetsen van verschillende tools wordt ook betrokken in deze fase. De shortlist vormt de basis voor de ontwikkeling van een prototype voor geautomatiseerde en adaptieve tekstvereenvoudiging.

De vijfde en laatste fase van het onderzoek bestaat uit het testen en beoordelen van gekozen AI-toepassingen voor tekstvereenvoudiging. In dit experiment proberen scholieren met dyslexie in de derde graad middelbaar onderwijs de shortlisted AI toepassingen en het prototype uit. Het doel van het experiment is om de effectiviteit en gebruikersvriendelijkheid van deze toepassingen te beoordelen. Na een grondige analyse wordt er met de resultaten bepaalt of de toepassingen aan de unieke noden van een scholier met dyslexie in de derde graad middelbaar onderwijs voldoen om wetenschappelijke artikelen te vereenvoudigen voor scholieren in het middelbaar onderwijs.

A.4. Verwacht resultaat, conclusie

Er wordt verwacht dat de huidige softwareoplossingen voor tekstvereenvoudiging onvoldoende aansluiten bij de noden van scholieren met dyslexie in de derde graad middelbaar onderwijs. Het prototype is moeilijk af te stemmen op de specifieke noden van deze doelgroep. Ontwikkelaars die werken met bestaande modellen moeten *custom transformers* inzetten om bevredigende resultaten te krijgen. Bovendien ontbreken er Nederlandstalige word embeddings die de complexiteit van elk woord bijhouden en aan kant-en-klare modellen die de inhoud van wetenschappelijke artikelen kunnen vereenvoudigen. Word embeddings uit een Germaanse taal gebruiken, gevolgd door vertaling naar het Nederlands is wel een aanvaardbaar alternatief.

Bibliografie

- Abdel-Salam, S. & Rafea, A. (2022). Performance Study on Extractive Text Summarization Using BERT Models. *Information*, (2).
- Althunayyan, S. & Azmi, A. (2021). Automated Text Simplification: A Survey. *ACM Computing Surveys*, 54, Article no. 43. <https://doi.org/10.1145/3442695>
- Ball, P. (2017). It's not just you: science papers are getting harder to read. *Nature*.
- Barnett, A. & Doubleday, Z. (2020). Meta-Research: The growth of acronyms in the scientific literature (P. Rodgers, Red.). *eLife*, 9, e60080.
- Bezem, A. & Lugthart, M. (2016). Visuele Disfunctie een onzichtbare belemmering bij lezen, spelling en concentratie. <https://beeldenbrein.nl/>
- Bilici, Ş. (2021). Sequence labeling.
- Bingel, J., Paetzold, G. & Søgaard, A. (2018). Lexi: A tool for adaptive, personalized text simplification. *Proceedings of the 27th International Conference on Computational Linguistics*, 245–258.
- Binz, M. & Schulz, E. (2023). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, 120(6).
- Bonte, M. (2020). *Bestaat Dyslexie?: En is het een relevante vraag?* uitgeverij SWP.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language Models are Few-Shot Learners.
- Bulté, B., Sevens, L. & Vandeghinste, V. (2018). Automating lexical simplification in Dutch. *Computational Linguistics in the Netherlands Journal*, 8, 24–48. <https://clinjournal.org/clinj/article/view/78>
- Canning, Y., Tait, J., Archibald, J. & Crawley, R. (2000). Cohesive Generation of Syntactically Simplified Newspaper Text. In P. Sojka, I. Kopeček & K. Pala (Red.), *Text, Speech and Dialogue* (pp. 145–150). Springer Berlin Heidelberg.
- Cao, M. (2022). A Survey on Neural Abstractive Summarization Methods and Factual Consistency of Summarization.
- Carbonell, J. & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 335–336.
- Chowdhary, K. (2020). *Fundamentals of Artificial Intelligence*. Springer, New Delhi.

- Coster, W. & Kauchak, D. (2011). Learning to Simplify Sentences Using Wikipedia. *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, 1–9. <https://aclanthology.org/W11-1601>
- Crevits, H. (2022, maart 13). *Kwart van bedrijven gebruikt artificiële intelligentie: Vlaanderen bij beste leerlingen van de klas* (Persbericht). Vlaamse Overheid Departement Economie, Wetenschap en Innovatie.
- Crossley, S. A., Allen, D. & McNamara, D. S. (2012). Text simplification and comprehensible input: A case for an intuitive approach. *Language Teaching Research*, 16(1), 89–108.
- Dandekar, N. (2016). How to use machine learning to find synonyms. <https://medium.com/@nikhilbd/how-to-use-machine-learning-to-find-synonyms-6380c0c6106b>
- Daoud, N. A. (2023). Vandaag Internationale Dag van de Moedertaal: Pleidooi voor Het Nederlands. <https://doorbraak.be/vandaag-internationale-dag-van-de-moedertaal-pleidooi-voor-het-nederlands/>
- Dapaah, J. & Maenhout, K. (2022, juli 8). *Iedereen heeft boter op zijn hoofd* (D. Standaard, Red.). https://www.standaard.be/cnt/dmf20220607_97763592
- De Belder, M.-F., Jan; Moens. Text simplification for children. eng. In: ACM; New York, 2010.
- De Craemer, J., Van Beeumen, L., Cooreman, A., Moonen, A., Rottier, J., Wagemakers, I. & Mardulier, T. (2018). Aan de slag met voorleessoftware op school. Een gids met 8 vragen en antwoorden. <https://onderwijs.vlaanderen.be/nl/onderwijspersoneel/van-basis-tot-volwassenenonderwijs/lespraktijk/ict-in-de-klas/voorleessoftware-voor-leerlingen-met-leesbeperkingen/aan-de-slag-met-voorleessoftware-op-school>
- De Meyer, I., Janssens, R. & Warlop, N. (2019). Leesvaardigheid van 15- jarigen in Vlaanderen: Overzicht van de eerste resultaten van PISA2018. <https://data-onderwijs.vlaanderen.be/documenten/bestand.ashx?id=12265>
- Deckmyn, D. (2021, maart 19). *Robot schrijft mee De Standaard* (D. Standaard, Red.). https://www.standaard.be/cnt/dmf20210319_05008561
- Desoete, A. (2017). Dyslexie of dyscalculie, niet de schuld van het onderwijs! En een correcte diagnose schaadt niet. *SPRANKEL*, 28(2), 17–31.
- Donato, A., Muscolo, M., Arias Romero, M., Caprì, T., Calarese, T. & Olmedo Moreno, E. M. (2022). Students with dyslexia between school and university: Post-diploma choices and the reasons that determine them. An Italian study. *Dyslexia*, 28(1), 110–127.
- Dronberger, G. B. & Kowitz, G. T. (1975). Abstract readability as a factor in information systems. *Journal of the American Society for Information Science*, 26(2), 108–111.
- DuBay, W. H. (2004). The principles of readability. *Online Submission*.

- Eisenstein, J. (2019). *Introduction to Natural Language Processing*. MIT Press. <https://books.google.be/books?id=72yuDwAAQBAJ>
- F, M. & Ennals, R. (2010). True Enough: Learning to Live in a Post-Fact Society, John Wiley, New Jersey (2008). *International Journal of Information Management* - *INT J INFORM MANAGE*, 30, 180–180.
- Fabbri, A. R., Kryściński, W., McCann, B., Xiong, C., Socher, R. & Radev, D. (2020). SummEval: Re-evaluating Summarization Evaluation.
- Filipiak, P. (2020). Leesproblemen en Dyslexie: Leesproblemen -2-. https://wij-leren.nl/leesproblemen-dyslexie-woordbenoeming-woordherkenning-begripsprobleem-deel-twee.php#_edn11
- Gala, N. & Ziegler, J. (2016). Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia. *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC)*, 59–66.
- Ganesan, K. (2018). ROUGE 2.0: Updated and Improved Measures for Evaluation of Summarization Tasks.
- Garbacea, C., Guo, M., Carton, S. & Mei, Q. (2021). Explainable Prediction of Text Complexity: The Missing Preliminaries for Text Simplification. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1086–1097. <https://doi.org/10.18653/v1/2021.acl-long.88>
- Garg, H. (2022). Using GPT-3 for education: Use cases. <https://indiaai.gov.in/article/using-gpt-3-for-education-use-cases>
- Ghesquière, P. (2018). *Als leren pijn doet: Kinderen met een leerstoornis opvoeden en begeleiden*. Acco.
- Gooding, S. (2022). On the Ethical Considerations of Text Simplification. *Ninth Workshop on Speech and Language Processing for Assistive Technologies (SLPAT-2022)*, 50–57. <https://doi.org/10.18653/v1/2022.slpac-1.7>
- Gooding, S. & Kochmar, E. (2019). Complex word identification as a sequence labelling task. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1148–1153.
- Greg, B., Atty, E., Elie, G., Joane, J., Logan, K., Lim, R., Luke, M. & Michelle, P. (2023). Introducing chatgpt and Whisper Apis. <https://openai.com/blog/introducing-chatgpt-and-whisper-apis>
- Hahn, U. & Mani, I. (2000). The Challenges of Automatic Summarization. *Computer*, 33, 29–36. <https://doi.org/10.1109/2.881692>
- Hartley, J. (1999). From Structured Abstracts to Structured Articles: A Modest Proposal. *Journal of Technical Writing and Communication*, 29(3), 255–270. <https://doi.org/10.2190/3RWW-A579-HC8W-6866>

- Harwell, D. (2023). Tech's hottest new job: Ai whisperer. no coding required. <https://www.washingtonpost.com/technology/2023/02/25/prompt-engineers-techs-next-big-job/>
- Hayes, D. P. (1992). The growing inaccessibility of science. <https://www.nature.com/articles/356739a0>
- Hern, A. (2023). TechScape: Will meta's massive leak democratise AI – and at what cost? <https://www.theguardian.com/technology/2023/mar/07/techscape-meta-leak-llama-chatgpt-ai-crossroads>
- Hollenkamp, J. (2020). Summary and analysis of Scientific Research Articles - San Jose State ... <https://www.sjsu.edu/writingcenter/docs/handouts/Summary%20and%20Analysis%20of%20Scientific%20Research%20Articles.pdf>
- Hsu, W.-T., Lin, C.-K., Lee, M.-Y., Min, K., Tang, J. & Sun, M. (2018). A Unified Model for Extractive and Abstractive Summarization using Inconsistency Loss.
- Huang, S., Wang, R., Xie, Q., Li, L. & Liu, Y. (2019). An Extraction-Abstraction Hybrid Approach for Long Document Summarization. *2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC)*, 1–6.
- Hubbard, K. E. & Dunbar, S. D. (2017). Perceptions of scientific research literature and strategies for reading papers depend on academic career stage. *PLOS ONE*, 12(12), 1–16.
- Iavarone, B., Brunato, D. & Dell'Orletta, F. (2021). Sentence Complexity in Context. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, 186–199. <https://doi.org/10.18653/v1/2021.cmcl-1.23>
- IBM. (2022). IBM Global AI Adoption Index 2022. <https://www.ibm.com/downloads/cas/GVAGA3JP>
- Iredale, G. (2022). An overview of tokenization algorithms in NLP. <https://101blockchains.com/tokenization-nlp/>
- Iskender, N., Polzehl, T. & Möller, S. (2021). Reliability of Human Evaluation for Text Summarization: Lessons Learned and Challenges Ahead. *Proceedings of the Workshop on Human Evaluation of NLP Systems (HumEval)*, 86–96. <https://aclanthology.org/2021.humeval-1.10>
- Jiang, R. K. (2023). Prompt engineering : Deconstructing and managing intention. <https://www.linkedin.com/pulse/prompt-engineering-deconstructing-managing-intention-jiang/>
- Jones, R., Colusso, L., Reinecke, K. & Hsieh, G. (2019). r/science: Challenges and Opportunities in Online Science Communication. *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3290605.3300383>
- Jurafsky, D., Martin, J., Norvig, P. & Russell, S. (2014). *Speech and Language Processing*. Pearson Education. <https://books.google.be/books?id=Cq2gBwAAQBAJ>

- Kandula, S., Curtis, D. & Zeng-Treitler, Q. (2010). A semantic and syntactic text simplification tool for health content. *AMIA annual symposium proceedings, 2010*, 366.
- Khan, A. (2014). A Review on Abstractive Summarization Methods. *Journal of Theoretical and Applied Information Technology*, 59, 64–72.
- Khurana, D., Koli, A., Khatter, K. & Singh, S. (2022). Natural Language Processing: State of The Art, Current Trends and Challenges. *Multimedia Tools and Applications*, 82, 25–27.
- Li, C. (2022). OpenAI's GPT-3 language model: A technical overview. <https://lambdalabs.com/blog/demystifying-gpt-3>
- Li, J., Sun, A., Han, J. & Li, C. (2018). A Survey on Deep Learning for Named Entity Recognition.
- Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. *Text Summarization Branches Out*, 74–81. <https://aclanthology.org/W04-1013>
- Lin, H. & Bilmes, J. (2010). Multi-document summarization via budgeted maximization of submodular functions. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 912–920.
- Linderholm, T., Everson, M. G., van den Broek, P., Mischinski, M., Crittenden, A. & Samuels, J. (2000). Effects of Causal Text Revisions on More- and Less-Skilled Readers' Comprehension of Easy and Difficult Texts. *Cognition and Instruction*, 18(4), 525–556.
- Lisowski, E. (2023). GPT-3 vs. CHATGPT. the key differences. <https://addepto.com/blog/gpt-3-vs-chat-gpt-the-key-differences/>
- Lissens, F., Asmar, M., Willems, D., Van Damme, J., De Coster, S., Demeestere, E., Maes, R., Baccarne, B., Robaeyst, B., Duthoo, W. & Desoete, A. (2020). Het stopt nooit...De impact van dyslexie en/of dyscalculie op het welbevinden en studeren van (jong)volwassenen en op de transitie naar de arbeidsmarkt: een bundeling van Vlaamse pilootstudies.
- Liu, Q., Kusner, M. J. & Blunsom, P. (2020). A Survey on Contextual Embeddings.
- Louwerse, M., McCarthy, P. & McNamara, D. (2007). A Linguistic Analysis of Simplified and Authentic Texts. *The Modern Language Journal*, 91, 15–30.
- Malik, R. S. (2022, juli 4). *Top 5 NLP Libraries To Use in Your Projects* (T. Al, Red.). <https://towardsai.net/p/l/top-5-nlp-libraries-to-use-in-your-projects>
- Martens, M., De Wolf, R. & Evens, T. (2021a). *Algoritmes en AI in de onderwijscontext: Een studie naar de perceptie, mening en houding van leerlingen en ouders in Vlaanderen*. Kenniscentrum Data en Maatschappij. Verkregen 30 maart 2022, van <https://data-en-maatschappij.ai/publicaties/survey-onderwijs-2021>

- Martens, M., De Wolf, R. & Evens, T. (2021b, juni 28). *School innovation forum 2021*. Kenniscentrum Data en Maatschappij. Verkregen 1 april 2022, van <https://data-en-maatschappij.ai/nieuws/school-innovation-forum-2021>
- McAuliffe, Z. (2023). DuckDuckGo introduces new AI feature: DuckAssist. <https://www.cnet.com/tech/services-and-software/meet-duckassist-duckduckgos-new-ai-feature/>
- McCombes, S. (2022). How to write A summary: Guide amp; examples. <https://www.scribbr.com/working-with-sources/how-to-summarize/>
- McDonald, R. (2007). A study of global inference algorithms in multi-document summarization. *Advances in Information Retrieval: 29th European Conference on IR Research, ECIR 2007, Rome, Italy, April 2-5, 2007. Proceedings* 29, 557–564.
- McFarland, A. (2023). What is prompt engineering in AI amp; Why It Matters. <https://www.unite.ai/what-is-prompt-engineering-in-ai-why-it-matters/>
- McKeown, K., Klavans, J. L., Hatzivassiloglou, V., Barzilay, R. & Eskin, E. (1999). Towards multidocument summarization by reformulation: Progress and prospects.
- Menzli, A. (2023). Tokenization in NLP: Types, challenges, examples, tools. <https://neptune.ai/blog/tokenization-in-nlp>
- Minnen, J., Verbeylen, J. & Glorieux, I. (2018). De arbeidsduur en werkcontext van leraren. Onderzoek naar de tijdsbesteding van leraren uit het basis- en secundair onderwijs. <https://torvub.be/wp-content/uploads/2018/09/Samenvatting.pdf>
- Miszczak, P. (2023). Prompt engineering: The ultimate guide 2023 [GPT-3 amp; chat-gpt]. <https://businessolution.org/prompt-engineering/>
- Mottes, C. (2023). GPT-3 vs. Bert: Comparing the two most popular language models. <https://blog.invgate.com/gpt-3-vs-bert>
- Murdos, A., Hodges, M., Rubio, D. & Adams, J. S. (2014). Examining the readability of research abstracts to Determine Whether the General Public Can Understand Key Findings in Science. https://ctsi.ucla.edu/education/files/view/docs/06_08_2016_PALM_Communication_of_Science_Poster.pdf
- Nallapati, R., Zhai, F. & Zhou, B. (2017). SummaRuNNer: A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1). <https://doi.org/10.1609/aaai.v31i1.10958>
- Nandhini, K. & Balasundaram, S. (2013). Improving readability through extractive summarization for learners with reading difficulties. *Egyptian Informatics Journal*, 14(3), 195–204.
- Nenkova, A. & Passonneau, R. (2004). Evaluating Content Selection in Summarization: The Pyramid Method. *Proceedings of the Human Language Techno-*

- logy Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004, 145–152.
- Niemeijer, A., Frederiks, B., Riphagen, I., Legemaate, J., Eefsting, J. & Hertogh, C. (2010). Ethical and practical concerns of surveillance technologies in residential care for people with dementia or intellectual disabilities: an overview of the literature. *Psychogeriatrics*, 22(7), 1129–1142. <https://doi.org/10.1017/S1041610210000037>
- OnderwijsVlaanderen. (2023). Voorleessoftware voor Leerlingen met Leesbeperkingen. <https://onderwijs.vlaanderen.be/voorleessoftware-voor-leerlingen-met-leesbeperkingen>
- Paetzold, G. & Specia, L. (2016). SemEval 2016 Task 11: Complex Word Identification. *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 560–569. <https://doi.org/10.18653/v1/S16-1085>
- Pain, E. (2016). How to (seriously) read a scientific paper. <https://www.science.org/content/article/how-seriously-read-scientific-paper>
- Parveen, D. & Strube, M. (2015). Integrating importance, non-redundancy and coherence in graph-based extractive summarization. *Twenty-Fourth International Joint Conference on Artificial Intelligence*.
- Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C. & Thompson, W. H. (2017). Research: The readability of scientific texts is decreasing over time (S. King, Red.). *eLife*, 6, e27725.
- Poel, M., Boschman, E. & op den Akker, R. (2008). A Neural Network Based Dutch Part of Speech Tagger [<http://eprints.ewi.utwente.nl/14662>; 20th Benelux Conference on Artificial Intelligence, BNAIC 2008, BNAIC ; Conference date: 30-10-2008 Through 31-10-2008]. In A. Nijholt, M. Pantic, M. Poel & H. Hondorp (Red.), *BNAIC 2008* (pp. 217–224). Twente University Press (TUP).
- Premjith, P., John, A. & Wilscy, M. (2015). Metaheuristic Optimization Using Sentence Level Semantics for Extractive Document Summarization, 347–358. https://doi.org/10.1007/978-3-319-26832-3_33
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. e.a. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.
- Raj, D. (2017). Metrics for NLG evaluation. <https://medium.com/explorations-in-language-and-learning/metrics-for-nlg-evaluation-c89b6a781054>
- Rani, R. & Kaur, B. (2021). The TEXT SUMMARIZATION AND ITS EVALUATION TECHNIQUE. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(1), 745–752.
- Readable. (2021). *Flesch Reading Ease and the Flesch Kincaid Grade Level*. <https://readable.com/readability/flesch-reading-ease-flesch-kincaid-grade-level/>

- Rello, L. & Baeza-Yates, R. (2013). Good fonts for dyslexia. *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2013*.
- Rello, L., Baeza-Yates, R., Dempere-Marco, L. & Saggion, H. (2013). Frequent Words Improve Readability and Short Words Improve Understandability for People with Dyslexia.
- Rello, L., Baeza-Yates, R. & Saggion, H. (2013). The Impact of Lexical Simplification by Verbal Paraphrases for People with and without Dyslexia. *7817*, 501–512.
- Rello, L. & Baeza-Yates, R. A. (2015). How to present more readable text for people with dyslexia. *Universal Access in the Information Society*, *16*, 29–49.
- Rello, L. & Bigham, J. (2017). Good Background Colors for Readers: A Study of People with and without Dyslexia, 72–80.
- Rello, L., Kanvinde, G. & Baeza-Yates, R. (2012). Layout Guidelines for Web Text and a Web Service to Improve Accessibility for Dyslexics. *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*.
- Ribas, J. (2023). Building the new bing. <https://www.linkedin.com/pulse/building-new-bing-jordi-ribas/>
- Ribeiro, E., Ribeiro, R. & de Matos, D. M. (2018). A Study on Dialog Act Recognition using Character-Level Tokenization.
- Rijkhoff, J. (2022). Tekst Inkorten?: 9 tips om Je Teksten korter Te Maken. <https://dialogtrainers.nl/tekst-inkorten-tips/>
- Rivero-Contreras, M., Engelhardt, P. E. & Saldaña, D. (2021). An experimental eye-tracking study of text adaptation for readers with dyslexia: effects of visual support and word frequency. *Annals of Dyslexia*, *71*, 170–187.
- Roldós, I. (2020, december 22). *Major Challenges of Natural Language Processing (NLP)*. MonkeyLearn. Verkregen 1 april 2022, van <https://monkeylearn.com/blog/natural-language-processing-challenges/>
- Roose, K. (2023). Don't ban chatgpt in schools. teach with it. <https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html>
- Ruben, A. (2016). How to read a scientific paper. <https://www.science.org/content/article/how-read-scientific-paper-rev2>
- Sciforce. (2020, februari 4). *Biggest Open Problems in Natural Language Processing*. Verkregen 1 april 2022, van <https://medium.com/sciforce/biggest-open-problems-in-natural-language-processing-7eb101ccfc9>
- ShafieiBavani, E., Ebrahimi, M., Wong, R. & Chen, F. (2018). A Graph-theoretic Summary Evaluation for ROUGE, 762–767. <https://doi.org/10.18653/v1/D18-1085>
- Shardlow, M. (2013). A Comparison of Techniques to Automatically Identify Complex Words. *51st Annual Meeting of the Association for Computational Lin-*

- guistics *Proceedings of the Student Research Workshop*, 103–109. <https://aclanthology.org/P13-3015>
- Shardlow, M. (2014). A Survey of Automated Text Simplification. *International Journal of Advanced Computer Science and Applications(IJACSA), Special Issue on Natural Language Processing 2014*, 4(1). <https://doi.org/10.14569/SpecialIssue.2014.040109>
- Siddharthan, A. (2006). Syntactic Simplification and Text Cohesion. *Research on Language and Computation*, 4(1), 77–109. <http://oro.open.ac.uk/58888/>
- Siddharthan, A. (2014). A survey of research on text simplification. *ITL - International Journal of Applied Linguistics*, 165, 259–298.
- Sikka, P. & Mago, V. (2020). A Survey on Text Simplification. *CoRR*, abs/2008.08612. <https://arxiv.org/abs/2008.08612>
- Simon, J. (2021). Large language models: A new moore's law? <https://huggingface.co/blog/large-language-models>
- Snow, C. (2010). Academic Language and the Challenge of Reading for Learning About Science. *Science (New York, N.Y.)*, 328, 450–2.
- Sohom, G., Ghosh; Dwight. (2019). *Natural Language Processing Fundamentals*. Packt Publishing. <https://medium.com/analytics-vidhya/natural-language-processing-basic-concepts-a3c7f50bf5d3>
- Steinberger, J. & Jezek, K. (2009). Evaluation Measures for Text Summarization. *Computing and Informatics*, 28, 251–275.
- Strubell, E., Ganesh, A. & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP.
- Suleiman, D. & Awajan, A. (2020). Deep Learning Based Abstractive Text Summarization: Approaches, Datasets, Evaluation Measures, and Challenges. *Mathematical Problems in Engineering*, 2020.
- Surma, T., Vanhoyweghen, K., Sluijsmans, D., Camp, G., Muijs, D. & Kirschner, P. (2019). *Wijze lessen: twaalf bouwstenen voor effectieve didactiek* (1ste ed.). Ten Brink Uitgevers.
- Suter, J., Ebling, S. & Volk, M. (2016). Rule-based Automatic Text Simplification for German.
- Swayamdipta, S. (2019, januari 22). *Learning Challenges in Natural Language Processing*. Verkregen 1 april 2022, van <https://www.microsoft.com/en-us/research/video/learning-challenges-in-natural-language-processing/>
- Tanya Goyal, G. D., Junyi Jessy Li. (2022). News Summarization and Evaluation in the Era of GPT-3. *arXiv preprint*.
- Tatman, R. (2019). Evaluating text output in NLP: Bleu at your own risk. <https://towardsdatascience.com/evaluating-text-output-in-nlp-bleu-at-your-own-risk-e8609665a213>

- Thangarajah, V. (2019). Python current trend applications-an overview.
- Tops, W., Callens, M., Brysbaert, M. & Schouten, E. L. (2018). *Slagen met Dyslexie in Het Hoger Onderwijs*. Owl Press.
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E. & Lample, G. (2023). LLaMA: Open and Efficient Foundation Language Models.
- van der Meer, C. (2022). Dyslexie hebben is Niet Zo Raar: Lezen is iets heel onnatuurlijks. <https://www.demorgen.be/beter-leven/dyslexie-hebben-is-niet-zo-raar-lezen-is-iets-heel-onnatuurlijks~bc608101/>
- Van Vreckem, C. & Desoete, A. (2015). Het ene kind met dyslexie is het andere niet: implicaties uit begrijpend leesonderzoek en spellingonderzoek voor diagnostiek en therapie. *LOGOPEDIE*, 58–66.
- Vasista, K. (2022). Evolution of AI Design Models. *Central Asian Journal of Theoretical and Applied Science*, 3(3), 1–4.
- Vellutino, F., Fletcher, J., Snowling, M. & Scanlon, D. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of child psychology and psychiatry, and allied disciplines*, 45, 2–40.
- Verhoeven, W. (2023, februari 8). *Applaus voor de studenten die ChatGPT gebruiken* (Trends, Red.). https://trends.knack.be/economie/bedrijven/applaus-voor-de-studenten-die-chatgpt-gebruiken/article-opinion-1934277.html?cookie_check=1676034368
- Verma, P. & Verma, A. (2020). A review on text summarization techniques. *Journal of scientific research*, 64(1), 251–257.
- Weinberg, G. (2023). Duckduckgo launches DuckAssist. <https://spreadprivacy.com/duckassist-launch/>
- Wentink, W., Verhoeven, L. & Druenen, M. (2008). Protocol leesproblemen en dyslexie voor groep 1 en 2. *Clinical Rheumatology - CLIN RHEUMATOL*.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J. & Schmidt, D. C. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT.
- Xu, W., Callison-Burch, C. & Napoles, C. (2015). Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, 3, 283–297.
- Zeng, Q., Kim, E., Crowell, J. & Tse, T. (2005). A Text Corpora-Based Estimation of the Familiarity of Health Terminology. In J. L. "Oliveira, V. Maojo, F. Martín-Sánchez & A. S. Pereira (Red.), *Biological and Medical Data Analysis* (pp. 184–192). Springer Berlin Heidelberg.
- Zhang, J., Zhao, Y., Saleh, M. & Liu, P. (2020). Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. *International Conference on Machine Learning*, 11328–11339.

Zhang, M., Riecke, L. & Bonte, M. (2021). Neurophysiological tracking of speech-structure learning in typical and dyslexic readers. *Neuropsychologia*, 158, 107889.