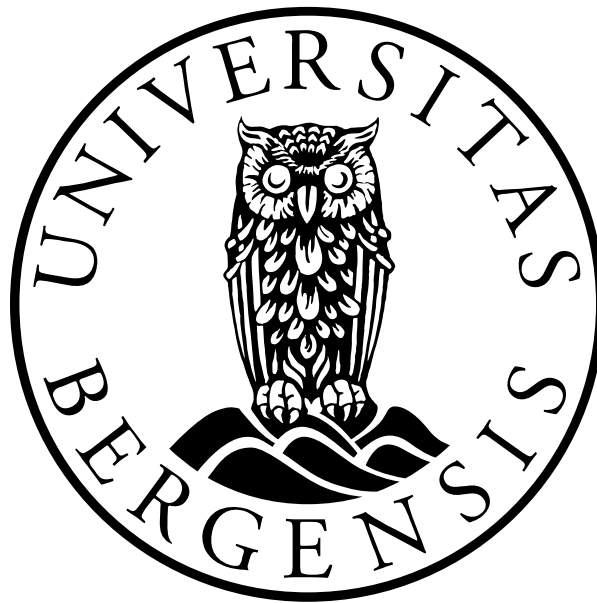


Generative språkmodeller for automatisert tekstforenkling: Tilpasning av nyhetsartikler for lesbarhet hos personer med dysleksi

Thale Knudsen Kirkhorn



Masteroppgave i medie- og interaksjonsdesign
Institutt for informasjons- og medievitenskap
Universitetet i Bergen

1. juni 2023

Sammendrag

I denne masteroppgaven utforskes utviklingen av en prototype for automatisert tekstforenkling med fokus på tilpasning av nyhetsartikler for personer med dysleksi. Prosjektet involverer flere iterasjoner for å finne den mest effektive tilnærmingen for å forenkle tekstene ved hjelp av generative språkmodeller. Prototypen inneholder tre nivåer av forenklet tekst, der nivå null representerer den originale teksten og nivå en til tre gradvis forenkler innholdet. Testing og evaluering av prototypen gjennomføres ved å utføre brukertester med målgruppen og samle tilbakemeldinger fra eksperter innen feltet. Disse tilbakemeldingene er verdifulle i forbedringen av prototypen og tekstforenklingen.

En viktig del av arbeidet er eksperimentering med ulike prompts og instruksjoner for å finne den mest effektive måten å be språkmodellen forenkle artiklene på. Observeringer indikerer at formuleringen av promptene har en innvirkning på resultatene, og at det er utfordringer knyttet til språket i de genererte versjonene. Dette vil kreve manuell redigering av tekstene for å oppnå ønsket lesbarhet og kvalitet. Funnene indikerer at språkmodellen fortsatt har begrensninger når det gjelder å generere grammatisk korrekte setninger og å velge passende ord i konteksten ved omformulering av en eksisterende tekst. Videre identifiseres utfordringer knyttet til maksimumsgrensen for tokens, som begrenser promptlengden, artikkellengden og den forenklede teksten.

Det er viktig å påpeke at denne forskningen blir gjennomført innenfor de nåværende rammene for språkmodeller for å teste reformulering av norske nyhetsartikler. Med fremtidig utvikling av språkmodeller og økende tokenkapasitet, samt inkorporering av retningslinjer for klarspråk direkte i modellene, kan vi forvente ytterligere forbedringer i automatisert tekstforenkling.

Forord

Først ønsker jeg å takke veilederen min Bjørnar Tessem for veiledningen gjennom dette prosjektet fra start til slutt.

Jeg vil også rette en stor takk til TekLab for deres økonomiske støtte, som gjorde det mulig for oss å reise til Oslo for samtaler med fagpersoner, ikke én, men to ganger i løpet av prosjektets gang.

Videre ønsker jeg å uttrykke takknemlighet til NTB, Klar Tale og Dysleksi Norge, som har delt sin ekspertise og gitt tilbakemeldinger.

Takk til Dysleksi Bergen og Dysleksi Ungdom for deres hjelp med å rekruttere deltakere til studien. Jeg vil også takke hver enkelt informant som stilte opp til intervjuer og brukertester, deres bidrag har vært avgjørende for resultatene i oppgaven.

Jeg ønsker også å takke min samarbeidspartner, Svenja Lys Forstner, for det gode samarbeidet og ikke minst de gode pausene.

Til slutt vil jeg takke familie og venner for støtte og oppmuntring gjennom hele denne prosessen. En spesiell takk går til min kjære mamma, pappa og Stine for deres kontinuerlige oppmuntring, verdifulle innspill og uvurderlige korrekturlesing.

Thale Knudsen Kirkhorn

Bergen, juni 2023

INNHALDSFORTEGNELSE

1	INNLEDNING	1
1.1	BAKGRUNN	1
1.2	FORSKNINGSSPØRSMÅL	2
1.3	STRUKTUR FOR OPPGAVEN	2
2	TEORI: DYSLEKSI, AUTOMATISK TEKSTFORENKLING OG KI	3
2.1	HVA ER DYSLEKSI?	3
2.2	KLARSPRÅK.....	4
2.2.1	<i>Retningslinjer</i>	4
2.3	AUTOMATISERT TEKSTFORENKLING	5
2.4	HVA ER KUNSTIG INTELLIGENS?	7
2.5	HVOR LANGT HAR TEKNOLOGIEN KOMMET?	9
2.5.1	<i>Etikk og bias innen kunstig intelligens</i>	10
2.6	EKSISTERENDE TEKSTFORENKLINGSVERTØY.....	12
2.7	AUTOMATISK TEKSTFORENKLING PÅ NORSK.....	14
3	FORSKNINGSMETODE: KVALITATIV METODE	16
3.1	KVALITATIV METODE	17
3.1.1	<i>Hvorfor kvalitativ metode i denne oppgaven?</i>	17
3.2	INTERAKSJONSDESIGN.....	18
3.2.1	<i>Brukersentrert design</i>	18
3.2.2	<i>Prototyping</i>	19
3.3	INNSAMLING AV DATA	20
3.3.1	<i>Litteratur</i>	20
3.3.2	<i>Semistrukturerte intervjuer</i>	20
3.3.3	<i>Use case</i>	21
3.4	EVALUERING AV PROTOTYPE	21
3.4.1	<i>Brukertesting</i>	22
3.4.2	<i>Personas</i>	22
4	DESIGNFORLØPET – DEFINERE ET KONSEPT	23
4.1	FORSKNINGSETIKK.....	23
4.2	MÅLGRUPPE.....	23
4.3	FORSKNINGSDELTAKERE	24
4.3.1	<i>Brukere</i>	24

4.3.2	<i>Ekspertter på klarspråk</i>	24
4.3.3	<i>IT-ekspertter</i>	25
4.4	ETABLERE KRAV FUNKSJONELLE KRAV	25
5	PROTOTYPING	26
5.1	UTVIKLINGSVERKTØY	26
5.1.1	<i>GPT-3</i>	26
5.1.2	<i>Figma</i>	26
5.2	ITERASJONSOVERSIKT	27
5.3	FØRSTE ITERASJON	27
5.3.1	<i>Intervju med brukere</i>	28
5.3.2	<i>Intervjuer med journalister</i>	29
5.3.3	<i>Low-fidelity prototype</i>	29
5.3.4	<i>Teknisk prototype</i>	30
5.3.5	<i>Redefinere kravene til prototypen</i>	31
5.4	ANDRE ITERASJON AV PROTOTYPEN	33
5.4.1	<i>Tekstforenklingsnivåer</i>	35
5.4.2	<i>Brukertesting av andre iterasjon</i>	35
5.5	TREDJE ITERASJON AV PROTOTYPEN	36
5.5.1	<i>Tekstforenklingsnivåer</i>	37
5.5.2	<i>Brukertesting av tredje iterasjon</i>	38
6	RESULTATER	41
6.1	FJERDE ITERASJON – ET FERDIG PRODUKT?	41
6.1.1	<i>Evaluering av prototypen</i>	41
6.1.2	<i>Endelig resultat av prototypen og videre endringer</i>	42
6.1.3	<i>Ikke-funksjonelle krav</i>	43
6.2	SAMMENLIGNE AUTOMATISK TEKSTFORENKLING AV TI NYHETSARTIKLER	44
6.2.1	<i>Problem med sitater</i>	46
7	DISKUSJON	47
7.1	AUTOMATISK TEKSTFORENKLING	47
7.1.1	<i>En sammenligning av tekstforenklingstilnærminger: NTB, Klar Tale og GPT-3</i>	47
7.1.2	<i>To iterasjoner av automatisk tekstforenkling: Fremgangsmåte og refleksjoner</i>	49
7.1.3	<i>Evaluering av prompts</i>	52
7.1.4	<i>Gjennomgående utfordringer med forenklete tekster</i>	52
7.1.5	<i>Lesbarhetsindeks</i>	54
7.1.6	<i>Retningslinjer for klarspråk: den ideelle teksten vs. den genererte teksten</i>	56
7.2	BRUKERSENTRETT DESIGN	57

7.2.1	<i>Datainnsamling</i>	57
7.2.2	<i>Iterativ utvikling</i>	57
7.2.3	<i>Evaluering</i>	57
7.3	HVORDAN KAN GENERATIVE SPRÅKMODELLER TILPASSES FOR Å FORENKLE NYHETSARTIKLER FOR PERSONER MED DYSLEKSI? 58	
7.4	HVORDAN KAN VI UTNYTTE GRADERTE FORENKLINGSNIVÅ FOR TILPASSING AV NYHETSARTIKLER FOR PERSONER MED DYSLEKSI? 59	
8	AVSLUTNING	62
8.1	RESULTATER	62
8.2	STUDIENS BEGRENSNINGER OG VIDERE UTVIKLING.....	63
9	REFERANSER	65
10	VEDLEGG	71
	VEDLEGG A: INTERVJUGUIDE TIL SEMI-STRUKTURERTE INTERVJUER MED MÅLGRUPPEN.....	71
	VEDLEGG B: PYTHON-KODE OG HTML-KODE FRA TEKNISK PROTOTYPE	72
	VEDLEGG C: TEKSTFORENKLINGER	74
	VEDLEGG D: PERSONAS.....	74

1 INNLEDNING

Dysleksi er en nevrologisk betinget lese- og skrivevanske som påvirker en betydelig del av befolkningen. For personer med dysleksi kan tradisjonell tekstpresentasjon være utfordrende, da kompleksiteten og den krevende språklige strukturen kan føre til redusert lesbarhet og forståelse (Statped, 2020). I denne masteroppgaven utforskes metoder innen interaksjonsdesign for å tilpasse nyhetsartikler til personer med dysleksi. Det er utviklet en prototype gjennom flere iterasjoner med formål om å forenkle nyhetsartikler på en måte som forbedrer lesbarheten. Prototypen har tre nivåer av forenklet tekst som har blitt generert av kunstig intelligens. Nivåene representerer ulike grader av forenkling og har som hensikt å imøtekomme ulike behov og preferanser blant brukerne. Gjennom denne masteroppgaven ønsker jeg å utforske effektiviteten og brukeropplevelsen av disse forenklete tekstnivåene. Gjennomføringen av denne masteroppgaven innebærer en kombinasjon av forskning innen interaksjonsdesign, automatisk tekstforenkling og brukertesting. For å samle inn data og vurderinger av prototypen er det utført tester og evalueringer med deltakere som har dysleksi. Arbeidet med prototypen vil forhåpentligvis kunne gi innsikt og bidra til utviklingen av mer tilgjengelige og leservennlige nyhetsartikler for personer med dysleksi. Dette prosjektet har blitt utviklet i samarbeid med Svenja Forstner, som i sin masteroppgave tar for seg den visuelle utformingen av prototypen (Forstner, Under utvikling).

1.1 Bakgrunn

Å ha kjennskap til aktuelle hendelser har vært viktig gjennom historien. Nyhetsbrevet oppsto i romertiden, og aviser slik vi kjenner dem i dag har eksistert siden begynnelsen av 1600-tallet (Britannica, u.å.). Nyheter har aldri vært mer tilgjengelig enn i dag. Med bare et klikk kan du lese om, lytte til eller se en hendelse i sanntid. Det finnes en rekke ulike medier å velge mellom, som papiraviser, nettaviser, podkaster, nyhetsvideoer, nyheter på ulike sosiale medium. Alternativene er mange, men tilgjengelighetsnivået er ikke alltid like godt.

Nyhetsartikler kan være krevende for personer med lese- og skrivevansker når informasjonen oppleves som lang og kompleks. Det er derfor viktig å sørge for at artikler er brukervennlige

og tar hensyn til denne målgruppen. Aviser tilbyr i liten grad tilpasninger for grupper som kunne ha nytte av mer lettleste tekster. Som en løsning på dette, vil denne oppgaven utforske bruken av kunstig intelligens, spesifikt generative språkmodeller, for å forenkle og tilpasse nyhetsartikler slik at de blir mer lettleste og tilgjengelige for lesere med dysleksi. Generative språkmodeller er avanserte algoritmer som, gjennom trening på store mengder data, er i stand til å produsere tekster av høy kvalitet (Chowdhary, 2020, s. 603–605).

1.2 Forskningsspørsmål

Det overordnede forskningsspørsmålet i denne oppgaven er:

Hvordan kan nyhetsartikler tilpasses og gjøres mer brukervennlig for personer med dysleksi ved hjelp av kunstig intelligens (KI)?

Med utgangspunkt i nylige fremskritt innen generative språkmodeller har jeg valgt å rette oppmerksomheten mot bruk av denne kunstig intelligens-teknologien i forskningsspørsmålet mitt. Som en spesifikk tilnærming vil jeg fokusere på følgende:

1. Hvordan kan generative språkmodeller tilpasses for å forenkle nyhetsartikler for personer med dysleksi?
2. Hvordan kan vi utnytte graderte forenklingsnivå for tilpassing av nyhetsartikler for personer med dysleksi?

1.3 Struktur for oppgaven

Kapittel 2: **Teori** går gjennom bakgrunn om dysleksi, automatisk tekstforenkling og kunstig intelligens. Kapittel 3: **Forskningsmetode** beskriver datainnsamlingsmetodene benyttet.

Kapittel 4: **Designforløpet** tar for seg målgruppe, forskningsdeltakere og funksjonelle krav.

Kapittel 5: **Prototyping** presenterer hvordan tre iterasjoner av prototype-utvikling ble utført.

Kapittel 6: **Resultater** gjør rede for den siste iterasjonen av prototypen, ikke-funksjonelle krav og gjør en videre evaluering av tekstforenklingsmetoden. Kapittel 7: **Diskusjon**

analyserer og evaluerer resultatene, reflekterer over fremgangsmåten og drøfter

forskningsspørsmålene. Kapittel 8: **Avslutning** presenterer hovedfunn, begrensninger og foreslår metoder for videre utvikling.

2 TEORI: Dysleksi, automatisk tekstforenkling og KI

Dette kapitlet begynner med en gjennomgang av dysleksi (2.1) og ulike retningslinjer innenfor klarspråk (2.2). Videre presenterer 2.3 ulike tilnærminger til tekstforenkling med spesielt fokus på automatisk tekstforenkling. 2.4 presenterer deretter kunstig intelligens og språkmodeller samt ulike tekstbehandlingsverktøy som bruker KI.

2.1 Hva er dysleksi?

Lese- og skrivevansker kan betraktes som en paraplybetegnelse for flere språkvansker. Omtrent fem til ti prosent av befolkningen i Norge har dysleksi, som gjør dette til den vanligste lærevansken (Helland, 2021). Så hva er egentlig dysleksi?

Jeg har valgt å bruke Dysleksi Norge sin definisjon av begrepet. Dysleksi Norge er en interesseorganisasjon som jobber for å bedre hverdagen til personer med lese- og skrivevansker, matematikkvansker og språkvansker (Dysleksi Norge, 2017, s. 4). De har valgt å ta utgangspunkt i International Dyslexia Association, British Dyslexia Association og ROSE-rapporten i utviklingen av definisjonen. De definerer dysleksi som en medfødt disposisjon som varer livet ut. Kjentetegn kan være «[...] omfattende vansker med ordavkodning og staving, i tillegg til vansker med andre språkrelaterte ferdigheter. Mest vanlig er vansker med fonologisk prosessering, hurtig benevning og fonologisk korttidsminne.». Enkelte kan også ha problemer med at de prosesserer det de leser saktere enn vanlig i tillegg til at de sliter med automatisering (Dysleksi Norge, 2017, s. 10). Lese- og skriveutfordringene «avviker fra personens øvrige kognitive ferdigheter» og som regel er vanlig undervisningsmetoder ikke tilstrekkelig. De understreker at god veiledning og teknologiske hjelpemidler kan være effektivt for å bedre lese- og skriveferdighetene (Dysleksi Norge, 2017, s. 10).

Opplevelsen av dysleksi er svært individuell, og graden av dysleksi kan variere betydelig fra person til person. Mens noen opplever en mild grad av dysleksi, kan andre ha langt mer alvorlige utfordringer. Noen vanlige kjennetegn på dysleksi er:

- Vansker med fonembevissthet. Et fonem er den minste enheten i talespråket (SIL International, 2003). I skriftsystemet representeres fonemene som regel av bokstaver, for eksempel *a* og *l*, men også av bokstavgrupper, for eksempel *skj* og *ng*. Fonemisk

bevissthet er “evnen til å identifisere og leke med individuelle lyder i talte ord” (National Center on Improving Literacy, 2018, oversatt fra engelsk). Fonemisk bevissthet er en del av fonologisk bevissthet: evnen til å innse at talte ord er sammensatt av individuelle lydenheter (National Center on Improving Literacy, 2022).

- Utfordringer med fonologisk lesing. Fonologisk lesing handler om å automatisere forbindelsen mellom bokstaver og lyder. Leseren leser nye ord sakte ved å trekke sammen lydene ordet er bygget opp av eller leser i stavelser (Johansen, 2022). Dette krever at leseren har både fonemisk og fonologisk bevissthet.
- Vansker med å raskt gjenkjenne ord (ortografisk lesing). Ved ortografisk lesing er avkodingen av nye ord automatisert, og ordgjenkjenning skjer raskt og presist. Det frigjør leserens kognitive ressurser slik at oppmerksomheten heller kan brukes til meningskonstruksjon og å forstå teksten (Johansen, 2022).
- Vanskeligheter med leseflyt
- Lav lesehastighet som påvirker leseforståelsen.
- Utfordringer med rettskriving.
- Språket i friskriving er enkelt og har begrenset variasjon (Dysleksi Norge, 2017, s. 14).

2.2 Klarspråk

Selv om tanken om klarspråk ikke er ny, har det i det offentlige vært større oppmerksomhet rundt viktigheten av at informasjon skal være lett å forstå for alle. Siden 2009 har Kommunal- og distriktsdepartementet delt ut Klarspråkprisen for å belønne klart og brukervennlig språk (Moderniseringsdepartementet, 2017). Klarspråk er ifølge Språkrådet «korrekt, klar og mottakertilpasset språk.». Kjentegn på klarspråk er for eksempel korte setninger, enkle ord, presist språk og tydelig struktur (Språkrådet, 2013). Språkrådet har utviklet en liste med generelle skriveråd for å skrive klart.

2.2.1 Retningslinjer

Det finnes flere retningslinjer og standarder for hvordan man skriver lettest tekst. Et eksempel på dette er den europeiske standarden rettet mot personer med kognitive funksjonshemninger, *Information for all, European standards for making information easy to read and understand*. Personer med dysleksi er ikke en del av gruppen disse retningslinjene er

utviklet for, men mange av reglene i disse retningslinjene overlapper med de som er utviklet spesielt for dysleksi. Det finnes flere artikler om nett-tilgjengelighet for personer med dysleksi. En av disse er artikkelen *Search, Read and Write: An Inquiry into Web Accessibility for People With Dyslexia* av Berget et al, som sammenligner ulike retningslinjer for lettest tekst i forhold til dysleksi (Berget et al., 2016). Artikkelen nevner blant annet WCAG 2.0 (Web Content Accessibility Guidelines) som er anerkjente retningslinjer utviklet med mål om å gjøre nettinnhold mer tilgjengelig for ulike grupper med ulike nedsatte funksjonsevner (*Web Content Accessibility Guidelines (WCAG) 2.1*, u.å.). British Dyslexia Association er et talerør for personer med dysleksi og arbeider med å fremme et dysleksivennlig samfunn (British Dyslexia Association, u.å.-a). De har også utviklet noen prinsipper som tar spesifikt hensyn til utfordringer personer med dysleksi ofte opplever. Berget et al. argumenterer at tilgjengeligheten forbedres dersom man kombinerer retningslinjer som WCAG 2.0 og retningslinjer spesifikt utviklet for dysleksi (Berget et al., 2016, s. 457). Språkrådet har også utviklet flere retningslinjer for klarspråk. I dette prosjektet er det benyttet en samling av disse retningslinjene.

Tungvint setning:	Forenklet setning:
«Det forekommer tilfeller hvor utførelse av skredforebyggende tiltak initiert av ansvarlige myndigheter medfører endringer i terrengformasjonene, og slike modifikasjoner kan bidra til en reduisering av skredrisikoen i området.»	«Ansvarlige myndigheter kan endre terrengformasjoner som tiltak mot skredfare. Slike endringer kan redusere skredrisikoen i området.»

Tabell 1: Eksempel på en tung setning og en forenklet versjon der det er benyttet tekstforenklingsprinsipper.

2.3 Automatisert tekstforenkling

Hensikten med tekstforenkling (TS) er å redusere kompleksiteten av en tekst, samtidig som man bevarer den originale betydningen og innholdet. Forskning på automatisert tekstforenkling har pågått siden 90-tallet og er fortsatt et aktivt forskningsfelt. Det finnes flere tilnærminger for automatisk tekstforenkling, og det deles ofte opp i leksikalsk, syntaktisk, enspråklig maskinoversetting og hybridteknikker. Leksikalsk tekstforenkling identifiserer og erstatter vanskelige ord med enklere og mer forståelige synonymer uten å forenkles syntaksen til teksten. Det arbeides fra ord til ord og tas ikke hensyn til sammenhengen resten av teksten

(Al-Thanyyan & Azmi, 2022, s. 7). Syntaktisk forenkling er prosessen med å forenkle en tekst samtidig som man beholder informasjonsinnholdet og betydningen. Syntaktisk forenkling refererer til å gjøre en tekst mindre komplisert ved å redusere bruken av grammatiske strukturer eller setningskonstruksjoner. Dette kan være for eksempel å endre konjunksjonen i setninger for å gjøre teksten lettere å lese for en bestemt målgruppe eller lettere å behandle for et dataprogram (Siddharthan, 2006, s. 77). Enspråklig maskinoversettelse handler om å bruke nevralt nettverk til å oversette en tekst til en enklere versjon på samme språk. (Al-Thanyyan & Azmi, 2022, s. 23). Hybridteknikker handler om å kombinere flere av disse metodene for å oppnå bedre resultater. For eksempel kan en hybridteknikk bruke én metode for å lage et sammendrag av en tekst, og deretter bruke en annen type metode for å forenkle ordforrådet og setningsstrukturen i sammendraget (Zaman et al., 2020).

Tekstforenkling blir ofte delt inn i følgende operasjoner:

- Erstatte komplekse ord og uttrykk med enklere synonymer
- Dele opp setninger
- Fjerne ord eller deler av en setning som ikke er essensielle
- Omrokere setningsoppbygging (Niklaus, 2022, s. 21)

En av de viktigste oppgavene til tekstforenkling er å identifisere kompleksiteten i teksten. Dette er vanskelig å definere, og i tillegg varierer det fra én person til en annen. Men, for å kunne avgjøre om teksten bør forenkles eller ikke, er det nødvendig å gjøre en slik vurdering. I tillegg vil det være nyttig når man skal evaluere lesbarheten til teksten som har blitt forenklet (Al-Thanyyan & Azmi, 2022, s. 2).

Automatisk tekstforenkling har mange mulige fordeler, som å øke tilgjengeligheten, klarspråket og leseforståelsen for ulike grupper. Det er kanskje spesielt nyttig for personer med lese- og skrivevansker eller personer som har et annet morsmål. Studier har vist at lange, kompliserte ord påvirker forståelse og lesbarhet negativt for personer med dysleksi. Tekstforenkling, som erstatter vanskelige ord med kortere og mer brukte ord, vil dermed forbedre lesbarheten (Al-Thanyyan & Azmi, 2022, s. 2). I en studie utført av Rello et al. (2013), ble det undersøkt i hvilken grad ordfrekvens og ordlengde påvirker lesbarhet og forståelighet for spansktalende personer med dysleksi. Forskerne erstattet substantiver med synonymer av ulik lengde og frekvens for å se hvordan det påvirket lesbarheten og forståeligheten av teksten. Resultatene viste at økt ordfrekvens bidro til bedre lesbarhet for

personer med dysleksi, altså at de mest vanlige ordene blir gjentatt oftere i teksten. I tillegg viste studien at bruk av kortere ord hadde en positiv effekt på forståeligheten. Videre fant studien at det var mer effektivt å presentere enklere synonymer etter behov i stedet for å erstatte slike synonymer gjennom hele teksten. Dette betyr at når det oppstår behov for å erstatte et vanskelig ord med en enklere synonym, så vil det ha en bedre effekt på forståeligheten å gjøre dette selektivt, i stedet for å gjennomføre generelle utskiftninger gjennom hele teksten (Rello et al., 2013, s. 215–216). Når det gjelder dyslektikere vil hovedformålet med tekstforenkling blant annet være å forbedre lesehastighet og generell forståelse.

Selv om automatisk tekstforenkling har potensiale til å forbedre lesbarhet og forståelighet, finnes det visse begrensninger og utfordringer. En av de sentrale utfordringene er å bevare meningen, tonen og stilen til originalteksten under forenklingen. Dette er spesielt viktig for å sikre at den forenklete teksten beholder den opprinnelige intensjonen og budskapet. Videre må automatiske tekstforenklingsverktøy håndtere flertydighet, tvetydighet, ironi og humor på en hensiktsmessig måte. Slike språklige elementer kan være vanskelige å forenkles uten å miste essensen eller den rette konteksten. En annen utfordring er å tilpasse seg ulike kontekster og formål. Forenklingen av tekst til en spesifikk kontekst eller formål kan føre til tap av informasjon eller endring av betydningen avhengig av sammenhengen. Derfor krever automatisk tekstforenkling en avansert språkteknologi som er i stand til å analysere, generere og evaluere tekster på en intelligent måte slik at forenklingen kan tilpasses sammenhengen.

2.4 Hva er kunstig intelligens?

Definisjonen av kunstig intelligens har endret seg etter hvert som teknologien har blitt utviklet. I denne oppgaven har jeg valgt å ta utgangspunkt i Regjeringens definisjon som er basert på EUs ekspertgruppe på feltet:

Kunstig intelligente systemer utfører handlinger, fysisk eller digitalt, basert på tolkning og behandling av strukturerte eller ustrukturerte data, i den hensikt å oppnå et gitt mål. Enkelte KI-systemer kan også tilpasse seg gjennom å analysere og ta hensyn til hvordan tidligere handlinger har påvirket omgivelsene (Kommunal- og moderniseringsdepartementet, 2020).

Teknologier basert på kunstig intelligens er programmert til å ta til seg informasjon og bearbeide denne mot et formål. Dataprogrammet bruker så målet til å forbedre seg, enten

gjennom direkte tilbakemelding eller ved å analysere og sammenligne resultatet mot sluttmålet.

En annen definisjon er koblet direkte mot intelligens-begrepet: Kunstig intelligens innebærer å simulere intelligent atferd i maskiner, med målet om å skape maskiner som kan etterligne menneskelig intellekt på viktige områder som språkforståelse, resonnering og problemløsning. Selvlæring er en viktig komponent i KI, og gjør det mulig for systemer å skaffe ny informasjon og forbedre sine kunnskapsbaserte vurderinger og konklusjoner gjennom erfaring og data. Maskinlæring og dypplæringsteknikker har blitt standard-tilnærmingen for å utvikle KI, der kunstige nevralt nettverk tillater modeller å lære og resonnerer som mennesker, og utføre stadig mer komplekse oppgaver. KI har allerede hatt betydelig innvirkning på flere bransjer og forventes å spille en avgjørende rolle i å hjelpe virksomheter med å holde seg konkurransedyktige i mange felt. KI omfatter en rekke områder, inkludert maskinlæring, regelbaserte systemer og naturlig språkbehandling (NLP) (Lund et al., 2023, s. 571).

Naturlig språkbehandling er et felt dedikert til utviklingen av metoder og teknikker for automatisk analyse og representasjon av menneskelig språk. Målet med NLP er å lære datamaskiner å forstå, tolke, generere og samhandle med menneskelig språk på en effektiv og betydningsfull måte. NLP omfatter et bredt spekter av oppgaver og teknikker, inkludert tekstanalyse, språkmodellering, automatisk oversettelse og semantisk forståelse. Ved hjelp av ulike algoritmer og maskinlæringsteknikker har NLP-systemer evnen til å bearbeide tekstdata, trekke ut informasjon, identifisere mønstre og tolke betydningen til ord og setninger. Gjennom bruk av NLP-teknikker kan datamaskiner raskt og effektivt tolke og behandle store mengder tekstdata. Dette har potensialet til å forbedre og automatisere oppgaver og gi verdifull innsikt fra store informasjonskilder (Chowdhary, 2020, s. 603 – 605).

Kunstig intelligens er et stort fagområde, og denne oppgaven vil fokusere spesifikt på naturlige språkmodeller. Store språkmodeller (LLM) refererer til en spesifikk type NLP-modell som er trent på store datasett med naturlig språk for å produsere og analysere tekst på en måte som skal speile menneskelig tankegang. De benytter seg av avanserte maskinlæringsteknikker og trening på store datamengder for å oppnå høy tekstkvalitet (Carlini et al., 2021).

Transformer-modeller er en type modell innenfor naturlig språkbehandling som har vist seg å være svært effektiv i ulike språkrelaterte oppgaver (Kalyan et al., 2021, s. 2). Modellen, som ble introdusert i 2017, introduserte en ny tilnærming ved å benytte seg av oppmerksomhetsmekanismer (attention) til å lære av store mengder tekstdata. Ved å konvertere inndataord til vektorrepresentasjoner som sendes gjennom enkodere og dekodere, oppnår transformermodellene en representasjon av konteksten og sammenhengen mellom ordene. Dette gjør dem effektive for ulike NLP-oppgaver som språkmodellering, oversettelse og tekstgenerering (Acheampong et al., 2021, s. 5793). Transformer-modellen har bidratt til betydelige fremskritt innenfor feltet og fortsetter å være et sentralt verktøy for å forbedre språklig analyse og tekstgenerering (Kalyan et al., 2021, s. 2). Blant de mest kjente Transformer-baserte modellene er GPT (Generative Pre-trained Transformer) utviklet av OpenAI og BERT (Bidirectional Encoder Representations from Transformers) utviklet av Google. Disse modellene trenes på store mengder data for å lære rike språkrepresentasjoner og kan deretter finjusteres for å oppnå avansert språkforståelse (Devlin et al., 2019). Transformer-modeller kan bruke alle de nevnte metodene for tekstforenkling (se delkapittel 2.3), avhengig av hvordan de er trent og evaluert.

Selvfølgelig! En språkmodell er en type datamaskinalgoritme som er trent på store mengder tekst for å kunne forstå og generere naturlig språk. Modellen lærer av mønstre og strukturer i språket ved å analysere og bearbeide store mengder tekstdata. Dette gjør det mulig for språkmodellen å forstå og tolke tekst, generere sammenhengende setninger og til og med svare på spørsmål eller utføre språkrelaterte oppgaver. Språkmodeller som GPT-3 og BERT har oppnådd betydelige fremskritt innen naturlig språkbehandling og brukes i en rekke applikasjoner, inkludert oversettelse, chatbots, tekstgenerering og mer.

Figur 1: ChatGPT forklarer hva en språkmodell er.

2.5 Hvor langt har teknologien kommet?

Kunstig intelligens brukes allerede i mange ulike felt som for eksempel bildeanalyse, aksjehandel og spill som sjakk, for å nevne noe. Ifølge Statistisk sentralbyrå (SSB) brukte 11 prosent av norske foretak kunstig intelligens i 2021 (Walther-Zhang, 2021). En undersøkelse gjennomført av Eurostat om bruk av KI i nordiske bedrifter fra 2022 viser likevel at Norge ligger på en tredjeplass bak Danmark og Finland på henholdsvis 24 og 16 prosent (Kvadsheim, 2022).

Siden 2015 har Lånekassen gjennomført landsdekkende bokkontroll. Formålet med denne kontrollen er å sikre at studenter har gitt korrekte opplysninger om sin bosituasjon i forhold til søknad om lån og stipend (Lånekassen, 2016, s. 15–16). I 2018 implementerte Lånekassen kunstig intelligens for utvelgelsen av studenter til bokkontroll. Resultatene fra denne kontrollen viste at maskinlæringsbasert utvelgelse var dobbelt så effektiv som tilfeldige utvalg i identifiseringen av studenter som ikke kunne dokumentere at de bodde separat fra foreldrene. Selv om antallet avvik oppdaget var sammenlignbart med året før, ble antallet kontroller betydelig redusert med nesten 50 prosent (Funnemark & Aagesen, 2019). Dette eksemplet illustrerer potensialet til kunstig intelligens for å forbedre effektiviteten til bokkontroller. Samtidig frigjør det også ressurser som Lånekassen kan bruke til andre interne oppgaver.

Bruksområdene for store språkmodeller er mangfoldige, og det har allerede hatt en betydelig innvirkning på flere områder. Disse modellene har evnen til å generere tekst av høy kvalitet innen ulike sjangre, inkludert fortellinger (Frantz, 2020), dikt (D. Anderson, 2021), og akademiske tekster (Lindland, 2023). De har også vist seg å være nyttig innen automatisk oversettelse ved å lære sammenhenger og språkstrukturer. Videre er språkmodeller i stand til å forstå og trekke ut informasjon fra tekst, inkludert gjenkjennelse av sammenhenger og følelser (Acheampong et al., 2021, s. 5821). Dette åpner opp for dypere innsikt og forståelse ved analyse av store datasett. Språkmodeller kan også benyttes som grammatiske verktøy, for eksempel ved korrekturlesning av tekster (Lancaster, 2023). Ved å tilby autokorreksjon og språkstøtte kan de hjelpe enkeltpersoner med språkutvikling og sikre et høyt språknivå. Dette kan være spesielt nyttig for personer med lese- og skrivevansker eller elever som ikke har tilgang til slik hjelp hjemmefra.

2.5.1 Etikk og bias innen kunstig intelligens

Mens kunstig intelligens blir mer utbredt i samfunnet er det voksende bekymringer om etiske dilemmaer og bias knyttet til utviklingen og bruken av store språkmodeller. Med bias menes forutinntatthet som påvirker oppfatninger og beslutninger. Den økende bruken av teknologier basert på store språkmodeller har ført til spørsmål rundt etikk og fordommer. En sentral bekymring er knyttet til videreføring av eksisterende fordommer og skjevheter i språkmodell-algoritmene. Ettersom disse modellene er trent på store mengder tekst fra internett, er de utsatt for å tilegne seg eventuelle fordommer eller feilinformasjon som finnes i kildedataene. Dette

kan føre til konsekvenser som videreføring og forsterkning av fordommer knyttet til kjønn, rase, religion eller sosiale ulikheter i KI-genererte tekster. Flere studier har undersøkt bias i store språkmodeller og gjort funn som bekrefter dette (Abid et al., 2021, s. 1; Bordia & Bowman, 2019). Videre stilles det også spørsmål spesielt knyttet til personvern og datasikkerhet. Det finnes flere metoder for å innhente sensitiv informasjon fra tekstdata. I mønstergjenoppbyggingsangrep blir mønstre i den opprinnelige teksten identifisert, og segment av den opprinnelige sekvensen som inneholder sensitiv informasjon blir rekonstruert. En annen metode er å bruke «stikkordinferens» for å avgjøre om tekster, for eksempel medisinske opplysninger, inneholder bestemte sensitive stikkord (Pan et al., 2020, s. 1315).

Et eksempel som illustrerer fordommer i datasett er det amerikanske selskapet Amazon, som utviklet et kunstig intelligens-verktøy med hensikt å forenkle utvelgelsen av de best kvalifiserte jobbsøkerne. Imidlertid erkjente Amazon i 2015 at deres nye system ikke vurderte søkere til stillinger som programvareutviklere og andre tekniske stillinger på en kjønnsnøytral måte. Årsaken til dette var at Amazons datamodeller ble trent opp ved å observere mønstre i CV-er som hadde blitt sendt til selskapet over en 10-års periode. Dette resulterte i at systemet favoriserte kandidater som brukte uttrykk som var hyppigere brukt blant mannlige søkere. Siden flertallet av disse CV-ene kom fra menn, ble denne skjevheten reflektert i AI-teknologien. I tillegg oppstod det andre problemer med dataene, noe som førte til at ukvalifiserte søkere ble anbefalt for ulike stillingstyper (Dastin, 2018).

Eksempelet om Lånekassens implementering av kunstig intelligens er et eksempel på bruk av teknologien som gir ønsket resultat. Likevel er det viktig å være bevisst på mulige implikasjoner knyttet til utvalgsriterier og mulig skjevhet i utvalgsgruppen. Selv om kunstig intelligens kan forbedre effektiviteten og ressursutnyttelsen i kontrollprosessen, er det viktig å sikre rettferdighet og lik behandling av alle studenter. Dette innebærer en nøye overveielse av utvalgsmetodikk, opplæring av algoritmer og evaluering av resultatene for å unngå utilsiktede skjevheter og sikre at beslutningene er rettferdige. Ved å opprettholde et kritisk blikk på databehandlingen kan man maksimere fordelene mens man samtidig reduserer mulige negative konsekvenser og opprettholde prinsippene om likeverdig behandling.

2.6 Eksisterende tekstforenklingsverktøy

Det finnes flere verktøy som er utviklet for å oppsummere eller skrive om tekst. Mange av disse bruker AI-teknologi for å analysere eller generere tekst. Teknologier som baseres på AI er et raskt voksende felt og gjennom tiden dette prosjektet har pågått har det både blitt lansert og forsvunnet verktøy.

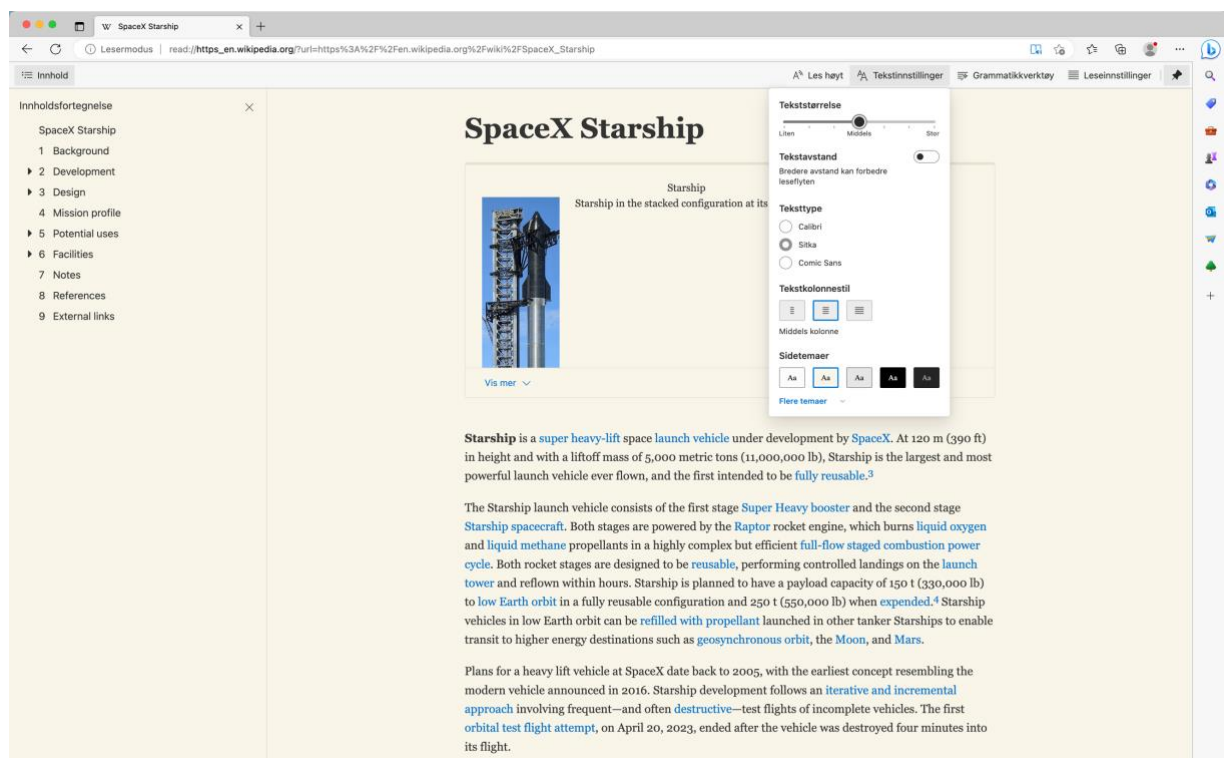
Mange av verktøyene som er tilgjengelig i dag er utvidelser man kan laste ned i nettleseren eller på datamaskinen. Hero er et eksempel på verktøy som dette. Appen lover også å forenkle kompliserte setninger til et enklere språk. Den er for tiden tilgjengelig på engelsk og dansk. Den engelske versjonen er mer avansert og kan forenkle hele setninger. Den danske versjonen erstatter imidlertid bare vanskelige ord med enklere synonymmer (Hero, 2022). De bruker en KI-algoritme som de har utviklet selv (Hero, u.å.). Wordhero er et annet skriveprogram som produserer tekst for brukeren. Her trenger brukeren kun å legge inn noen stikkord for at programmet kan lage en sammenhengende tekst med fullstendige setninger. Wordhero er basert på GPT-3 modellen og virker også på norsk (WordHero, u.å.).

Wordtune er en app som bruker AI til å generere tekst. Programmet kan blant annet parafrasere (gjengi tekst med egne ord uten å endre betydning), oppsummere, fikse skrivefeil, foreslå forbedringer i teksten og segmentere tekster. Programmet er utviklet for å tolke engelsk, men er også mulig å bruke på et lite antall andre språk. Norsk er imidlertid ikke implementert (Wordtune Team, u.å.).

Hemingway editor og Grammarly er to andre eksempler på populære tekstredigerere. De er begge designet for å hjelpe brukeren med å skrive bedre og tydeligere setninger. Hemingway editor gir deg en poengsum basert på lesbarheten til teksten. Ulike farger brukes for å fremheve ord eller setninger og angi hva problemet er. For eksempel, hvis en setning er veldig vanskelig å lese, er den merket med rødt (Hemingway Editor, u.å.). Grammarly hjelper brukeren ved å gi spesifikke forslag til ord som skal fjernes eller erstattes i tillegg til tegnsetting. Det er også en «tonedetektor» som forteller deg om teksten fremstår morsom, formell, osv. Grammarly benytter også AI for å drive applikasjonen (Grammarly, u.å.).

Mange nettlesere, som Firefox og Safari, har implementert en lesemodus som fjerner elementer som blant annet reklame og videoer. Det er også ofte mulig å gjøre egne

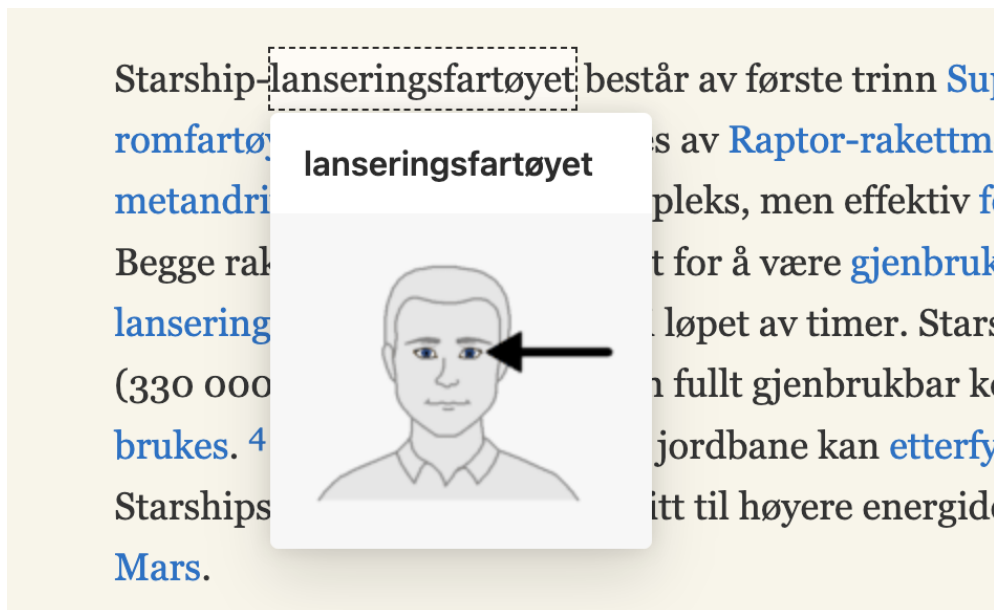
tilpasninger (Firefox, u.å.). Microsoft Edge er en nettleser utviklet av Microsoft. Nettleseren tilbyr en rekke funksjoner som blant annet innebygde tilgjengelighets- og læringsverktøy. «Immersive reader» ble opprinnelig utviklet for personer med dysleksi (lesevansker) og dysgrafi (skrivevansker) (Nicolson & Fawcett, 2011) (se Figur 2). Microsoft beskriver at verktøyet «simplifies web page layouts, removes clutter, and helps you customize your reading experience» (Microsoft, u.å.). Flere funksjoner er implementert i verktøyet. Det er blant annet mulig å endre tekststørrelse og mellomrom, kolonnebredde og forskjellige fargetemaer. Linjefokus er en funksjon som skyggelegger hele siden bortsett fra en regulerbar linje som skal gjøre det lettere å fokusere på noen få linjer av gangen. En bildeordbok forklarer ulike ord med et eller flere bilder. Markøren endres til en tryllestav dersom man plasserer musen over ord som har forklaring.



Figur 2: Skjermbilde av en wikipedia-nettside om SpaceX Starship i Microsoft Edges immersive reader.

En utfordring med ordforklaringene er at sammensatte ord ikke blir tilstrekkelig forklart. Et eksempel på dette er sammensetningen «hydrogen-oksygen», der kun det andre ordet, «oksygen», blir forklart. Videre blir ord med dobbel betydning enten forklart ved å vise begge betydningene eller ved å gi feil betydning som ikke passer til sammenhengen. Det kan føre til forvirring og misforståelser for leserne. Det er også mulig å oversette tekster til mange ulike språk. Oversetter man fra engelsk til norsk, virker heller ikke ordforklaringene like godt. Et

eksempel er Starship-lanseringsfartøyet, der «lanseringsfartøyet» blir forklart med et bilde av en mann med pil mot øyet, se **Error! Reference source not found.** Ordforklaringene er gjennomgående for enkle ord, som for eksempel «flytrafikk», i stedet for lange og vanskelige ord som ofte skaper problemer for personer med dysleksi. Tekst til tale-verktøyet fungerer svært godt på engelsk med en naturlig lesestemme. Den norske stemmen er imidlertid mer syntetisk.



Figur 3: Bildeordforklaring av «lanseringsfartøyet» i Microsoft Edge.

Rewordify.com er en nettside utviklet av en lærer som bruker en egen NLP-system til å forenkle tekster på engelsk og er utviklet med spesielt fokus på elevers leseutvikling. Brukeren limer inn en tekst eller en lenke til en nettside, og så vil rewordify.com generere en forenklet versjon av teksten med enklere ord og uttrykk (Rewordify, u.å.). Amazons Word Wise er en funksjon som hjelper deg å lese raskere og lettere på Kindle-enheter og apper. Når du slår på Word Wise, vil du se enkle definisjoner og synonymer over vanskelige ord. Du kan justere antall hint du ser med en skyveknapp. Word Wise er tilgjengelig for mange engelskspråklige bøker (Amazon, u.å.).

2.7 Automatisk tekstforenkling på norsk

Forskning på automatisk tekstforenkling er relativt omfattende for språk som engelsk, spansk, portugisisk, fransk, italiensk og andre språk (Ebling et al., 2022). Det er imidlertid gjort lite forskning gjort rundt kvaliteten av automatisk tekstforenkling på norsk. NorBERT og

NorELMo er store språkmodeller utviklet av Language Technology Group fra Universitetet i Oslo (UiO). Begge modellene har blitt trent på rundt to milliarder tokens av norsk tekst. Det er også planlagt å utvikle separate modeller for bokmål og nynorsk i fremtiden. Mange språk med begrensede ressurser har ikke egne enspråklige, storskala språkmodeller og må i stedet ty til bruk av en flerspråklig modell, som for eksempel Googles flerspråklige BERT-modell - mBERT - som ble trent på data som også inkluderte norsk. Frem til lanseringen av modellene utviklet av UiO var mBERT den eneste BERT-instansen som kunne brukes for norsk. (Kutuzov et al., 2021).

3 FORSKNINGSMETODE: Kvalitativ metode

Dette kapitlet gir en oversikt over metodene brukt i oppgaven. Først gis det en skildring av interpretivistisk forskning. 3.1 presenterer kvalitativ metode. 3.2 presenterer interaksjonsdesign og brukersentrert design. 3.3 presenterer innsamling av data. Til slutt presenterer 3.4 evalueringsmetoder av prototypen.

Williamson og Johnson skiller mellom to hovedforskningstilnæringer: positivisme og interpretivisme. Positivistisk forskning bruker «deduktivt resonnement», noe som betyr at det benyttes generelle premisser for å utvikle objektive og logiske argumenter eller konklusjoner. Det brukes vanligvis kvantitative forskningsmetoder med mange deltakere og legges vekt på målbare aspekter. Målet er å finne ut om en teori kan brukes på lignende tilfeller og oppnå en viss grad av generalisering. Interpretivistisk forskning bruker «induktivt resonnement» for å utvikle teorier. Induktivt resonnement tar spesifikke situasjoner og bruker dem til å danne generelle utsagn ved å finne mønstre. Interpretivistisk forskning anerkjenner at enkeltpersoner opplever situasjonene på forskjellige måter, og at perspektivene er avhengige av hverandre. Både forskere og deltakere har sin egen forståelse og tolkning av forskningstemaet som kan påvirke forskningsresultatene. Ved å følge en interpretivistisk tilnærming må man ta hensyn til dette under forskningen. Interpretivistisk forskning bruker vanligvis kvalitative metoder og fokuserer på individuelle perspektiver. Gjennom forskningsinteraksjoner blir likheter og forskjeller i perspektiver identifisert og kartlagt. (Williamson & Johanson, 2018, s. 5–7)

For å sikre vitenskapelig kvalitet er det utformet fire kriterier: credibility, transferability, dependability og confirmability. Kredibilitet (credibility), sikrer at funn og tolkninger gjenspeiler deltakernes ulike virkeligheter. Overførbarhet (transferability) vurderer i hvilken grad funn kan generaliseres til andre kontekster. Pålitelighet (dependability) fokuserer på å spore og forklare variasjonen i resultatene. Etterprøvbarehet (confirmability) handler om å etablere at resultatene ikke påvirkes av forskerens egne skjevheter og perspektiver. Gjennom å være bevisst på disse aspektene sikrer interpretativ forskning pålitelige og gyldige funn (Williamson & Johanson, 2018, s. 15).

3.1 Kvalitativ metode

Kvalitativ forskning handler om å forstå brukeren og gå i dybden på hva problemet er og hvordan det påvirker brukerne. Kvalitative forskningsmetoder kan omfatte ulike teknikker som intervjuer, brukertester og observasjoner. Disse metodene gir innsikt i brukeropplevelser og tillater forskerne å stille åpne spørsmål for å forstå brukernes perspektiver (Denny & Weckesser, 2022). Kvalitativ forskning er viktig fordi det gir en forståelse av målgruppens synspunkter og atferd som kan brukes til å forbedre produkter og tjenester. Dette kan også bidra til å identifisere problemområder tidlig i utviklingsprosessen, og gjøre at man sparer tid og ressurser på å rette opp feil senere. Det kan også hjelpe utviklere forstå hvorfor designbeslutninger ikke gir ønskede resultater og teste mulige løsninger (Sharp et al., 2019, s. 307–309). Et eksempel på dette kan være en nettavis som opplever nedgang i antall lesere. Uten kvalitativ forskning kan det bli vanskelig å få innblikk i årsaken til at færre brukere kommer tilbake til nettsiden. Kvalitativ forskning lar også utviklerne høre forskjellige brukerperspektiver og undersøke komplekse tema som kan være vanskeligere å utforske i for eksempel en spørreundersøkelse.

Mens forskere kan få mye informasjon fra et lite antall forskningsdeltakere, kan det være vanskelig å se størrelsen, omfanget eller problemet uten å også bruke kvantitativ forskning. En kvalitativ studie kan fange opp ekstreme svar som kan gi skjeve funn. Selv om kvalitative metoder kan gi innsikt, er det ikke statistisk representative. Bredt spekter og variasjon i mulige svar på kvalitative forskningsspørsmål og metoder gjør det utfordrende å oppdage trender og mønstre blant svarene. Alle deltakerne i en studie kan gi svært forskjellige svar på et spørsmål, noe som gjør det vanskelig å kategorisere og komme fram til klare konklusjoner (Sharp et al., 2019, s. 307–309).

3.1.1 Hvorfor kvalitativ metode i denne oppgaven?

Formålet med oppgaven er å finne ut hvordan digitale nyhetsartikler kan gjøres mer tilgjengelig for dyslektikere. Dysleksi er en medfødt lærevanske som i ulik grad gir utfordringer med nøyaktig og/eller flytende lesing av ord og tekst (Statped, 2020) (se delkapittel 2.1). Siden dysleksi oppleves ulikt fra person til person var det derfor ikke hensiktsmessig å benytte kvantitativ metode der deltakerne for eksempel vurderer prototypen med Likert-skala eller lignende. Dette fordi det vil være stort spenn i hvordan brukerne vil oppleve de ulike funksjonene i prototypen. En person med sterk grad av dysleksi, vil ha andre

behov enn en person med mindre grad. I tillegg vil deltakerne ha ulike strategier og erfaringer som vil påvirke hvordan de vurderer prototypen.

I denne oppgaven er det spesielt fokus på språkmodeller og hvordan de kan benyttes for å tilpasse tekst til en målgruppe. Vi har også valgt å benytte retningslinjer for klarspråk som er utviklet for å gjøre tekst tydelig og lett å forstå (se delkapittel 2.2.1). For å undersøke hvor godt automatisk tekstforenkling er, blir de i denne oppgaven vurdert opp mot kravene i retningslinjene.

3.2 Interaksjonsdesign

Interaksjonsdesign tar for seg utformingen og brukeropplevelsen av interaktive systemer. Dette involverer fire grunnleggende aktiviteter. Oppdagelse av krav, design av alternativer, prototyping og evaluering. Oppdagelse av krav vil si at det fokuseres på å forstå sluttbrukerne og deres behov. Dette innebærer datainnsamling og analyse for å danne grunnlaget for produktets krav. Deretter kommer design av alternativer, der ideer for å møte kravene utforskes. Dette inkluderer både konseptuell design, som definerer hvordan produktet kan brukes, og konkret design, som tar hensyn til detaljer som farger og ikoner. Prototyping skaper en interaktiv versjon av designet for å la brukere teste og evaluere produktet. Til slutt kommer evaluering, der brukervennlighet og brukbarhet til produktet vurderes gjennom diverse kriterier. Disse aktivitetene er sammenflettet og bidrar til å skape et optimalt og brukervennlig interaktivt produkt (Sharp et al., 2019, s. 50).

3.2.1 Brukersentrert design

Brukersentrert design er en tilnærming som legger vekt på å forstå og imøtekomme behovene, preferansene og ønskene til brukerne gjennom hele designprosessen. Sentralt i brukersentrert design er å involvere brukerne aktivt i alle faser av designprosessen, fra problemanalyse og idégenerering til prototyping og evaluering. Ved å samle inn og analysere brukernes tilbakemeldinger, innsikter og bruksmønstre kan man utvikle produkter og tjenester som er intuitive, nyttige og tilpasset brukernes kontekst og opplevelse. Brukersentrert design legger også vekt på iterativitet, der løsninger kontinuerlig forbedres basert på tilbakemeldinger og testing. Dette bidrar til å sikre at det endelige produktet er i tråd med brukernes behov og

forventninger. Brukersentrert design innebærer derfor en kritisk tilnærming for å skape meningsfulle og effektive løsninger som skaper positive brukeropplevelser.

John Gould og Cayton Lewis la i 1985 frem tre prinsipper som nå anses som fundamentet for en brukersentrert tilnærming. Det første prinsippet, tidlig fokus på brukere, understreker viktigheten av å forstå brukernes behov på et tidlig stadium av designprosessen. Ved å plassere brukerne i sentrum og analysere deres behov, preferanser og kontekstuelle utfordringer, kan designere utvikle løsninger som er skreddersydde for brukernes behov og gir en best mulig brukeropplevelse. Det andre prinsippet, empiriske målinger, legger vekt på betydningen av å samle inn og analysere faktiske data og observasjoner fra brukere. Ved å gjennomføre empiriske målinger, for eksempel brukertesting og observasjon, kan designere få innsikt i brukernes atferd, preferanser og utfordringer. Dette gir et solid grunnlag for å evaluere og forbedre designløsningen basert på objektive og pålitelige data. Det tredje prinsippet, iterativt design, innebærer å kontinuerlig teste og evaluere designløsninger gjennom hele prosessen. Ved å samle inn tilbakemeldinger og innsikt fra brukerne underveis, kan designere identifisere og håndtere problemområder og forbedre løsningen. Denne iterative tilnærmingen sikrer at produktet stadig blir mer optimalisert og tilpasset brukernes behov (Sharp et al., 2019, s. 47–49).

3.2.2 Prototyping

Prototyping er en prosess der man utvikler noe fra en idé til noe mer konkret. Det er en viktig aktivitet i designprosessen og innebærer å skape en tidlig versjon eller modell av et produkt eller system. Formålet er å visualisere og teste konseptet eller designet for å evaluere og forbedre det før den endelige implementeringen. Dette kan være alt fra en enkel tegning (low fidelity), til en detaljert konstruksjon (high fidelity). Ved å lage prototyper kan designere og utviklere samle tilbakemeldinger fra brukere, identifisere mulige problemer og utforske forskjellige designalternativer for å sikre at sluttresultatet oppfyller brukernes behov og forventninger. Prototyping bidrar til å redusere risikoen for kostbare feil og misforståelser, samtidig som det gir mulighet for eksperimentering og innovasjon. Det er en dynamisk og iterativ prosess som bidrar til å informere og veilede designbeslutninger underveis i utviklingsprosessen (Sharp et al., 2019, s. 421–424). I dette tilfellet handler det om å teste ut hvordan man kan bruke kunstig intelligens til å forenkle tekster for en spesifikk målgruppe.

3.3 Innsamling av data

Denne seksjonen tar for seg datainnsamlingsmetodene som ble brukt i dette prosjektet.

3.3.1 Litteratur

I arbeidet med oppgaven har vi brukt ulike artikler og verk som omhandler dysleksi, tekstforenkling og kunstig intelligens som grunnlag for utviklingen av prototypen. Det har også blitt brukt en samling av ulike retningslinjer for klarspråk som utgangspunkt for å vurdere de genererte versjonene av tekstene.

3.3.2 Semistrukturerte intervjuer

Semistrukturerte intervjuer er samtaler mellom forsker og informant der spørsmålene er bestemt på forhånd, men der informanter gjerne får ulike oppfølgingsspørsmål. Dette gjør at det er lett å styre samtalen, men at samtalen blir mer uformell slik at det er større rom for å få ut informasjon som kan være relevant. I tillegg gir det muligheter for å følge opp eventuelle tema som informanten tar opp og som kan være interessant i arbeid med prototypen. Dette var spesielt nyttig i startfasen av prosjektet der vi prøvde å kartlegge hvilke utfordringer målgruppen hadde og hvordan dette eventuelt kunne møtes i et verktøy.

I startfasen intervjuet vi journalister fra NTB, Klar Tale og Aftenposten Junior, med erfaring rundt letteste nyhetsartikler for å lære mer om hvilke metoder de bruker, og hva som legges vekt på i denne typen tekster. Disse intervjuene ga oss viktig informasjon om hvordan journalistene arbeidet. Vi fant blant annet ut at det er mye mer utfordrende å forenkle en eksisterende artikkel enn å skrive en artikkel fra scratch. Journalistene uttrykte at de savnet et verktøy som kunne brukes til å forenkle tekst, slik at det kunne frigjøre tid til å skrive flere dybdeartikler for målgruppen. I de fleste intervjuene som ble gjennomført, ble Liksskår (lesbarhetsindeks) nevnt. Dette er en måling av hvor lett en tekst er å lese. Liksskår beregnes ved å kombinere antall ord og setninger i teksten på en spesiell måte. Liks er prosent av lange ord pluss gjennomsnittlig antall ord per setning (J. Anderson, 1983, s. 491). For å beregne Liksskåren til en tekst legges teksten inn i et program som analyserer teksten og regner ut scoren for deg. Ifølge Klar Tale bør Liks-scoren være under 30 for at en tekst skal regnes som lett å lese, men de prøver å holde seg rundt 24 siden de er en lettlest avis. Det ble også nevnt at Liks kan brukes for å gi en pekepinn, men at det er vanskelig å høre om en tekst er lett å lese med mindre man leser det høyt.

Vi hadde også intervjuer med personer fra NTB med kunnskap om bruk av kunstig intelligens i journalistikkfeltet. Her stilte vi spørsmål om hvordan de jobber, hvilke verktøy/program de bruker og utfordringer knyttet til KI. Informasjon om arbeidet med utviklingen av Nynorskroboten var spesielt interessant med tanke på prosjektet. Nynorskroboten er et oversettingsverktøy utviklet av NTB Arkitektst som oversetter tekster fra bokmål til nynorsk (NTB, u.å.).

Videre intervjuet vi målgruppen til prosjektet for å få et innblikk i deres forhold til nettbaserte nyhetsartikler. Her stilte vi blant annet spørsmål om selve leseprosessen, om det var utfordringer som gikk igjen og hvorfor de eventuelt stoppet å lese en artikkel før de hadde lest den ferdig. Disse intervjuene var med på å forme kravene til prototypen vi utviklet (se VEDLEGG A: INTERVJUGUIDE TIL SEMI-STRUKTURERTE INTERVJUER MED MÅLGRUPPEN for intervjuguide).

3.3.3 Use case

Use case eller brukstilfelle, er en beskrivelse av en typisk eller forventet interaksjon mellom en bruker og et system. Det er en metode som brukes for å forstå og dokumentere hvordan et system skal fungere i ulike scenarier eller situasjoner. Et brukstilfelle identifiserer hvilke handlinger en bruker kan utføre og hvordan systemet responderer på disse handlingene. Dette kan være nyttig for å definere krav, validere designvalg, og sikre at systemet er tilpasset brukernes behov (Sharp et al., 2019, s. 415).

3.4 Evaluering av prototype

I dette delkapittelet presenteres evalueringsmetodene av prototypen som er utviklet gjennom flere iterasjoner. Gjennom hele prosessen har vi tatt hensyn til tilbakemeldinger fra brukertester for å gjøre nødvendige justeringer. Disse testene har bidratt til å identifisere potensielle utfordringer og muligheter for forbedring. I tillegg til brukertester har vi også gjennomført en siste evaluering av prototypen der vi fikk verdifulle tilbakemeldinger fra eksperter innen relevante fagfelt. Denne kombinasjonen av brukerbasert tilbakemelding og ekspertvurdering er ment å gi et grunnlag for å vurdere prototypens kvalitet og brukervennlighet.

3.4.1 Brukertesting

For å gjøre det mer attraktivt for personer å delta valgte vi å få muntlige tilbakemeldinger i både intervjusituasjon og under testing av prototypen. Vi tok opp lyd og skrev notater underveis. Det ble gjennomført to runder med brukertester. Totalt endte vi på 14 brukertester, med ti forskjellige deltakere. Fire av deltakerne var med på to brukertester og seks personer deltok bare på andre runde av brukertestene. Videre ble den fjerde iterasjonen av prototypen vurdert gjennom tilbakemeldinger fra eksperter innen relevante fagfelt.

3.4.2 Personas

Persona er detaljerte beskrivelser av typiske brukere av et produkt og skal gi et helhetlig bilde. De skal representere målgruppen og hvilke behov og ønsker de har. Det brukes mye av designere for å minne dem på personene som kommer til å bruke det endelige produktet og dermed gjøre gode designvalg slik at det blir en god brukeropplevelse (Sharp et al., 2019). I dette prosjektet ble det utformet to ganske forskjellige personaer. Den ene representerer gjennomsnittet av brukerne vi intervjuet og den andre representerer en annen type bruker som kan være vanskeligere å rekruttere. Se Figur 11 og Figur 12. Disse personaene er laget i samarbeid med prosjektpartner for prosjektet og er derfor på engelsk. Personaene ble brukt spesielt til evalueringen av første iterasjon av prototypen (Se VEDLEGG D: PERSONAS for personas), (Forstner, Under utvikling).

4 DESIGNFORLØPET – Definere et konsept

I dette kapitlet presenteres først forskningsetikk i 4.1. 4.2 presenterer målgruppen. 4.3 presenterer forskningsdeltakere. 4.4 presenterer funksjonelle krav for prototypen. Til slutt presenteres produktvisjonen.

4.1 Forskningsetikk

I forbindelse med prosjektet har det blitt gjennomført datainnsamling med forskningsdeltakere. Deltakerne mottok et samtykkeskjema som inneholdt informasjon om forskningsomfanget. Det ble understreket at det var frivillig å underskrive, og det var avgjørende for om brukeren ønsket å delta på intervju eller brukertest. Det ble også klart kommunisert at anonymiteten deres ville bli ivaretatt gjennom hele prosessen, og at de hadde rett til å trekke seg fra forskningen når som helst. Dersom noen bestemte seg for å trekke seg, ville all data som var samlet inn bli slettet. Alle deltakerne ble tilsendt samtykkeskjema på forhånd, men det ble også gjennomgått muntlig før de signerte. På denne måten ble det sikret at alle var fullstendig informert om forskningens formål og omfang.

4.2 Målgruppe

Målgruppen for prosjektet er voksne over 18 år som har dysleksi og leser nyheter. Mye forskning rundt dysleksi er fokusert på barn i skolealder. Ved å rette fokus mot en voksen målgruppe, kan det bidra til å sette søkelys og øke kunnskapen om hvordan dysleksi påvirker voksne og hvilke behov de har i forbindelse med informasjonsinnhenting. Lese og skriveverktøy er også ofte rettet mot barn. Det at voksne ikke bruker verktøy i like stor grad, kan ha flere årsaker. En årsak kan være at verktøyene er tilrettelagt for barns læring. En annen årsak kan være at de ikke er klar over hva som eksisterer. Ved å velge en voksen målgruppe kan det også gjøre at det er lettere å få tilgang til denne gruppen og rekruttere deltakere til studien. Det gjør også prosessen rundt samtykke enklere, da de kan signere samtykkeskjema selv. Dette kan også bidra til å øke antall deltakere og gjøre det lettere å gjennomføre en grundig og pålitelig undersøkelse. Å sikre at alle innbyggere har tilgang til og forstår nyhetsinnhold er en viktig del av en demokratisk samfunnsstruktur. Det å identifisere

barrierer og utfordringer voksne med dysleksi møter når de leser nyheter, kan bidra til å skape en mer inkluderende mediebransje og bidra til å styrke demokratiet.

4.3 Forskningsdeltakere

4.3.1 Brukere

Rekrutteringen av informanter til prosjektet viste seg å være en utfordrende oppgave, da det var nødvendig å benytte ulike metoder for å engasjere deltakere. Noen av deltakerne ble kontaktet gjennom personlige bekjentskaper, mens en tok initiativ etter å ha sett plakater som var opphengt rundt i Bergen. Et innlegg om prosjektet ble også publisert på Dysleksi Bergens Facebook-side, som resulterte i at to personer meldte sin interesse. Dysleksi Ungdom bidro også til å rekruttere fem personer som deltok i brukertesting. Det er viktig å erkjenne at dysleksi er en svært individuell opplevelse. Under rekrutteringen av deltakere til forskningsprosjektet var det enklere å tiltrekke seg studenter, og dermed kan denne gruppen være overrepresentert i studien. Det er viktig å ta hensyn til at studenter kanskje i større grad har en mildere form for dysleksi eller er mer erfarne lesere sammenlignet med en gjennomsnittlig bruker. Deres akademiske bakgrunn og hyppige eksponering for tekstbaserte oppgaver kan ha påvirket deres leseferdigheter og tilnærming til tekster. Dette faktum kan ha implikasjoner for resultatene og generaliserbarheten av funnene. På den annen side var det også en interessant observasjon at flere av deltakerne hadde familiemedlemmer som også hadde dysleksi. Dette perspektivet ga dem en unik innsikt i hvordan andre i deres nære krets opplever tilstanden, og det bidro til å berike diskusjonen og forståelsen av dysleksi i prosjektet.

4.3.2 Ekspertter på klarspråk

Journalister fra NTB, Aftenposten Junior og avisen Klar Tale med kunnskap og erfaring med lettest tekst bidro til datainnsamling om dette temaet samt tilbakemelding på siste iterasjon av prototypen. Tre personer fra Dysleksi Norge ga også verdifulle kommentarer på prototypen. En person fra Dysleksi Norge ga oss også innsikt i opplevelsen av dysleksi og ulike utfordringer knyttet til dette i starten av prosjektet.

4.3.3 IT-eksperter

Totalt seks personer med utviklerbakgrunn bidro til prosjektet. Tre personer ble intervjuet under feltarbeidet og fire personer bidro med tilbakemelding på tredje iterasjon av prototypen. En ekspert ga også tilbakemelding på prosjektet og prototypen midtveis i prosjektet.

4.4 Etablere krav funksjonelle krav

Designkrav beskriver nøyaktig hvilke egenskaper man forventer av et produkt og hvordan det skal utføres (Sharp et al., 2019, s. 384). Det er viktig å vite hvem målgruppen er og forstå evnene og behovene deres. Tradisjonelt sett har to forskjellige typer krav blitt identifisert, funksjonelle krav og ikke-funksjonelle. Funksjonelle designkrav beskriver hva et produkt skal gjøre (Sharp et al., 2019, s. 390). Basert på informasjonen samlet under feltarbeidet og litteratur ble følgende funksjonelle designkrav etablert for prototypen:

- Forenkle tekst
- Tre nivåer av forenklet tekst
- Bytte ut vanskelige ord
- Omformulere vanskelige setninger
- Vis nyhetsartikkel i en «leseversjon» uten reklame og andre forstyrrende elementer
- Forklare vanskelige ord i teksten med «pop-up»
- Meny med ulike verktøy
- Knapp til å bytte skrifttype
- Knapper til å justere skriftstørrelse

Produktvisjon

Etter intervjuer og datasamling ble det også utviklet en produktvisjon for prototypen.

Produktvisjon refererer til en klar og overordnet beskrivelse av formålet til et produkt eller en tjeneste. Det innebærer å identifisere og kommunisere produktets formål, målgruppe, verdier og ønsket fremtidig tilstand (Benassi et al., 2011, s. 165). Følgende produktvisjon ble utformet for prototypen som vi har gitt navnet «enkl.»:

«For lesere med dysleksi som ønsker å lese tilgjengelige nyheter, er «enkl.» et leseverktøy som transformerer nettbaserte nyhetsartikler til en regulerbar lettlest versjon. I motsetning til andre lignende forenklingsverktøy er produktet utviklet for norske nyhetsartikler og spesielt designet for lesere med dysleksi.»

5 PROTOTYPING

Dette kapittelet gir en oversikt over utviklingen av prototypen. 5.1 presenterer utviklingsverktøy. 5.2 presenterer en oversikt over iterasjonene. 5.3. presenterer første iterasjon av prototypen. 5.4 presenterer andre iterasjon av prototypen. 5.5 presenterer tredje iterasjon av prototypen.

5.1 Utviklingsverktøy

Denne delen presenterer hvilke utviklingsverktøy som har blitt brukt til å designe og utvikle de ulike iterasjonene av prototypen.

5.1.1 GPT-3

GPT (Generative Pre-Trained Transformer) er en dyp læringsmodell utviklet av OpenAI for å generere tekst og er trent på store datasett med naturlig språk (Brown et al., 2020, s. 5). GPT-3 er den tredje versjonen av denne modellen og brukes til dette prosjektet. Davinci er en av de mest anerkjente GPT-3 modellene og er basert på treningsdata fram til 2019 (OpenAI, u.å.-b). GPT-språkmodellen bruker tokens for å representere ord og uttrykk. Tokens i GPT er vanligvis innhentet ved hjelp av en prosess som kalles tokenisering, som innebærer å dele en tekst opp i mindre enheter som ord, fraser eller tegn. Disse enhetene representeres som numeriske verdier, som kan mates inn i modellen og brukes til å generere tekst. Det er imidlertid en maksimumsgrense for hvor mye tekst man kan legge inn i språkmodellen på grunn av tokens. En token tilsvarer omtrent fire symboler (OpenAI, u.å.-c). GPT-modellen bruker også en teknikk som kalles kontekstuell embedding, som vil si at modellen kan forstå betydningen av ord, og dermed generere sammenhengende tekst som følger naturlig språk (OpenAI, u.å.-a).

5.1.2 Figma

Figma er et gratis designverktøy som kan brukes til å designe prototyper til mobilapplikasjoner og webapplikasjoner. En av fordelene med Figma er at man kan arbeide

sammen med andre, som var svært nyttig siden prototypene er laget med en medstudent (Staiano, 2022).

5.2 Iterasjonsoversikt

Oversikt over iterasjonene som viser hvilke metoder brukt i hver iterasjon, de største endringene mellom hver iterasjon og antall deltakere (Forstner, Under utvikling).

Iterasjon	Metoder	Evaluering og signifikante funn	Antall deltakere
1	Intervjuer Personas	Definere brukerstudie Nettleserutvidelse Tre forenklingsnivåer	6 personer med dysleksi 7 felteksperter
2	Brukertest	Forenkle forenklingsnivåer Flytte meny	4 personer med dysleksi
3	Brukertest	Tre fargetemaer Lagreknapp	10 personer med dysleksi
4	Tilbakemelding fra eksperter	Forenkle ordforklaringer Tekst-til-tale-knapp	11 felteksperter i fire grupper

Tabell 2: Oversikt over de fire iterasjonene.

5.3 Første iterasjon

I starten av prosjektet ble det gjort intervjuer med journalister som har erfaring med klartekst for å få et innblikk i ulike aspekter som er viktig å ta hensyn til når man skriver lettlest tekst. Intervjuer med utviklere som har kunnskap om KI-journalistikk og kunstig intelligens ble også gjennomført. For å kartlegge opplevelsen av nyhetsartikler for personer i målgruppen og finne ut mer om ulike behov eller utfordringer, ble det også gjennomført intervjuer med personer med dysleksi. Det ble også gjort en gjennomgang av forskning på dysleksi og ulike eksisterende hjelpemidler. Dette ga en bedre forståelse av ulike problemer dyslektikere møter i forbindelse med nyhetsartikler og hvilken retning det brukervennlige verktøyet skulle ta. Informasjonen som ble samlet i denne fasen la grunnlaget for det konseptuelle designet, funksjonelle og ikke-funksjonelle krav og første iterasjon av prototypen. Den første prototypen ble deretter utgangspunkt for neste iterasjon som ble brukertestet med personer i målgruppen.

Valg av plattform for utviklingen av prototypen var også viktig. I startfasen av prosjektet var det usikkert hvordan KI-en skulle brukes og hvordan teksten skulle presenteres. Etter nøye vurdering ble det bestemt å utvikle en nettside for en datamaskin. Dette skyldes at slike enheter vanligvis har større skjermer sammenlignet med mobile enheter som mobiltelefoner eller nettbrett. Ved å velge denne tilnærmingen kunne potensielle utfordringer knyttet til skjermstørrelse unngås, som for eksempel plassering av menylinjen eller synlighet av funksjoner ved bruk av berøringsskjermer. Denne tilnærmingen gjorde det mulig å legge større vekt på selve interaksjonen, spesielt visning og formatering av tekst.

5.3.1 Intervju med brukere

Totalt seks personer fra målgruppen ble intervjuet i forkant av prototyp utviklingen. Det ble valgt å utføre semistrukturerte intervjuer for å samle informasjon. Deltakerne ble spurt om opplevelser med nettaviser og ulike utfordringer de hadde erfart på dette området. Et tema som ble nevnt av flere var at de ofte merket at samme innhold ble nevnt flere ganger i den samme artikkelen. Dette gjorde at de ble usikre på om de leste den samme delen av teksten på nytt. En utfordring mange med dysleksi opplever er å miste plasseringen i teksten når de leser, og at de derfor leser samme linje om igjen (Gregor et al., 2003, s. 343). En annen erfaring flere delte var at dersom en tekst er delt opp enten i mindre paragrafer eller med bilder virker teksten mindre lang, og man får avbrekk som hjelper på konsentrasjonen. Flere nevner også at de ikke setter seg ned for å lese en artikkel så ofte, spesielt dersom den er lang, både fordi det er tidkrevende og energikrevende. Det må i så fall omhandle noe de er veldig interessert i. Dersom man støter på vanskelige ord, skal det mye til for å søke det opp, og det blir dermed gjort sjeldent.

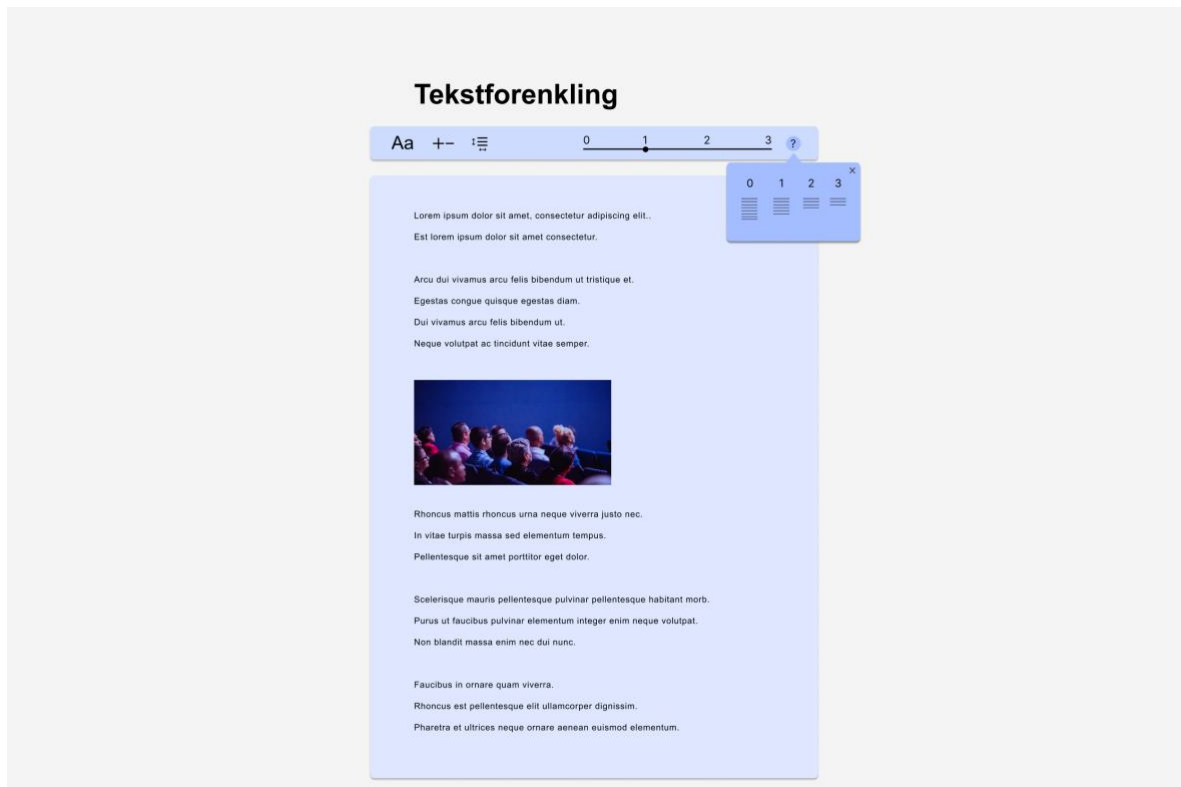
Det ble også stilt spørsmål angående en funksjon som noen nettaviser har implementert de siste årene, at i begynnelsen av en artikkel står det estimert lesetid. Noen svarte at dette ikke var noe de hadde lagt merke til, eller at de ikke pleide å se på det. Mens noen svarte at dette var et tips som de brukte mye. Dersom de synes lesetiden var for lang, gjorde det at de ikke leste artikkelen.

5.3.2 Intervjuer med journalister

Gjennom intervjuene fikk vi innsikt i ulike utfordringer ved manuell forenkling av nyhetsartikler. Klar Tale delte spesielt mye av hvordan de arbeider med forenkling. Prosessen innebærer å lese gjennom teksten for å finne hovedinnholdet og det mest viktige, og deretter enten gå gjennom setning for setning eller skrive om teksten fra scratch. Enkelte teknikker som brukes inkluderer konvertering av tall til brøk fordi det er mindre abstrakt og ofte lettere å forstå. Ord med dobbel betydning blir byttet ut, men endring av et enkelte ord kan ofte kreve at hele setningen må skrives på nytt. Det er også fokus på å endre setninger fra passiv til aktiv form. Direkte sitater kan ofte være for lange og blir gjort om til kortere indirekte sitat. Setninger med mye informasjon blir også delt opp. Det er ofte nødvendig å bryte skriveregler med å for eksempel dele opp lange ord med bindestrek og starte setninger med ord som «men» eller «og». Det kan også være utfordrende å forenkle nyhetsartikler dersom teksten mangler vesentlig bakgrunnsinformasjon som er nødvendig for å forstå hva saken handler om. Uten denne bakgrunnskunnskapen kan det være vanskelig å trekke ut de mest relevante punktene fra originalteksten og forenkle dem på en forståelig måte. I tillegg kan det hende at spørsmål som må besvares for å skrive en lettfattelig versjon av artikkelen, ikke er enkle nok i originalteksten. Å løse disse utfordringene krever derfor at journalistene noen ganger må kontakte kilden fra artikkelen for å fylle inn eventuelle mangler i informasjonen. Klar Tale delte også at de har skrevet lettleste partiprogrammer i samarbeid med politiske partier. De fikk tilbakemeldinger fra personer som sa at de stemte ved valget på grunnlag av informasjonen de tilegnet seg gjennom disse tekstene.

5.3.3 Low-fidelity prototype

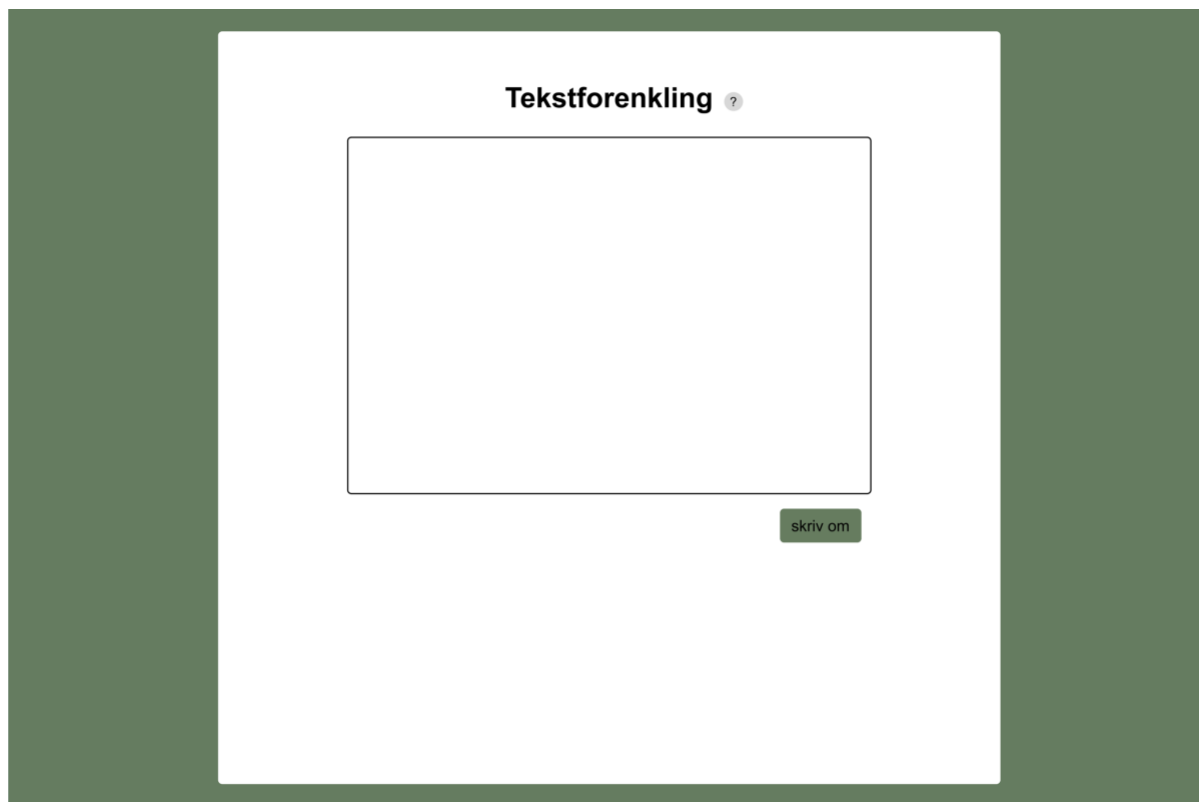
Den første versjonen av prototypen er laget i Figma og er kun illustrasjoner. Startsidene inneholder en tekstboks, en meny med diverse verktøy og en knapp som forenkler teksten. På denne siden kan brukeren lime inn tekst. Neste side viser teksten i en forenklet versjon med den samme verktøylinjen øverst. Man har mulighet til å endre skrifttype, linjeavstand og ordavstand i tillegg til tre nivåer av forenkling. Alle funksjonene har en tilhørende pop-up boks med flere funksjoner. Ved siden av forenklingsnivåene er det en informasjonsknapp med en enkel illustrasjon som gir en indikasjon på tekstlengden til hvert nivå.



Figur 4: Figma-skisse fra første iterasjon av prototypen.

5.3.4 Teknisk prototype

Den tekniske prototypen er utviklet ved hjelp av Python, HTML, CSS og GPT-3. Koden i Python tar teksten fra HTML-dokumentet og kjører den gjennom GPT-3 sammen med en prompt som instruerer språkmodellen til å forenkle teksten. Prompt: «Rewrite the following text in an easy-to-understand language for people with dyslexia». Deretter blir man omdirigert til en annen side som presenterer den nye teksten. Den nye versjonen av teksten er kortet ned til noen få setninger. Grunnet tidsbruk ble denne tekniske prototypen satt til side og videre iterasjoner ble utformet i Figma.



Figur 5: Første iterasjon av prototypen, teknisk prototype.

5.3.5 Redefinere kravene til prototypen

Den tekniske prototypen og low-fidelity prototypen har omtrent samme design. Etter å ha gjennomført en brukerstudie for å forstå hvordan typiske brukere leser nyhetsartikler, ble utformingen av verktøyet revurdert, og nye krav ble definert. Personaene ble brukt til å evaluere den første iterasjonen av prototypen, samt utvikle brukerstudien. I stedet for å måtte legge inn tekst selv, ville det være en bedre løsning for brukeren å ha tilgang til verktøyet direkte fra en åpen nyhetsartikkel. Dette vil både redusere antall nødvendige klikk og spare tid. Derfor ble det besluttet å bytte til en nettleserutvidelse som prototypens format for de påfølgende iterasjonene. Videre ble prototypen konseptualisert som en utvidelse som kan lastes ned i nettleseren og dermed brukes på alle nettaviser. Ikonet for justering av mellomrom mellom ord ble også redesignet for å være tydeligere. Verktøylinjen ble erstattet med en mindre meny, og det ble lagt til en funksjon for å endre tekstfarge og bakgrunn. Ordforklaringer for vanskelige ord og uttrykk ble også implementert. Kjernen av ideen er å tilby et hjelpemiddel som er brukervennlig og forenkler hverdagen, uten behov for å kopiere og lime inn tekst i et separat verktøy. Verktøyet er tilpasset alle nettaviser, men det kan også

utvides til å inkludere forskningsartikler og lignende. Vi har hovedsakelig fokusert på nettaviser, og derfor har kravene våre gjenspeilet dette området.

Videre ble det besluttet å fokusere på designaspektet i de påfølgende utviklingsiterasjonene i stedet for å videreutvikle den tekniske siden. Dette valget ble hovedsakelig tatt på grunn av tidsbegrensninger. Prosjektet er også utviklet med hensyn til brukeropplevelse, og fokuset bør derfor ligge på dette perspektivet. Derfor er de påfølgende utviklingsiterasjonene i hovedsak rettet mot brukeropplevelsen og i mindre grad mot den tekniske realiseringen.

Etter evalueringen av den første iterasjonen ble det også utviklet en detaljert «use case» som beskriver interaksjonen mellom en gjennomsnittlig bruker og leseverktøyet. Hensikten med å utarbeide en «use case» var å gi en bedre forståelse og konseptualisering av hvordan en typisk bruker kan benytte seg av et leseverktøy i sitt daglige bruk. Ved å identifisere og beskrive handlingene til brukeren i ulike steg, kunne man få innsikt i hvordan verktøyet kunne støtte brukeren på en effektiv måte. Denne «use casen» illustrerer en typisk brukers anvendelse av verktøyet og fremhever de spesifikke trinnene som er nødvendige for å oppnå ønsket resultat. Ved å konstruere en gjennomgang av brukerens handlinger og interaksjoner kan man identifisere potensielle utfordringer og forbedringsområder i designet og funksjonaliteten. Dette bidro til videre iterasjoner og utviklingen av leseverktøyet for å sikre en bedre brukeropplevelse. Handlingene til brukeren blir beskrevet i følgende steg:

1. Bruker åpner en nettavis
2. Ser en overskrift som virker interessant
3. Åpner nyhetsartikkelen
4. Ser at artikkelen er veldig lang
5. Ser at utvidelsesikonet er synlig
6. Åpner utvidelse
7. Tilpasser teksten med ulike funksjoner i menyen
8. Leser den tilpassede versjonen av artikkelen
9. Lukker utvidelsen

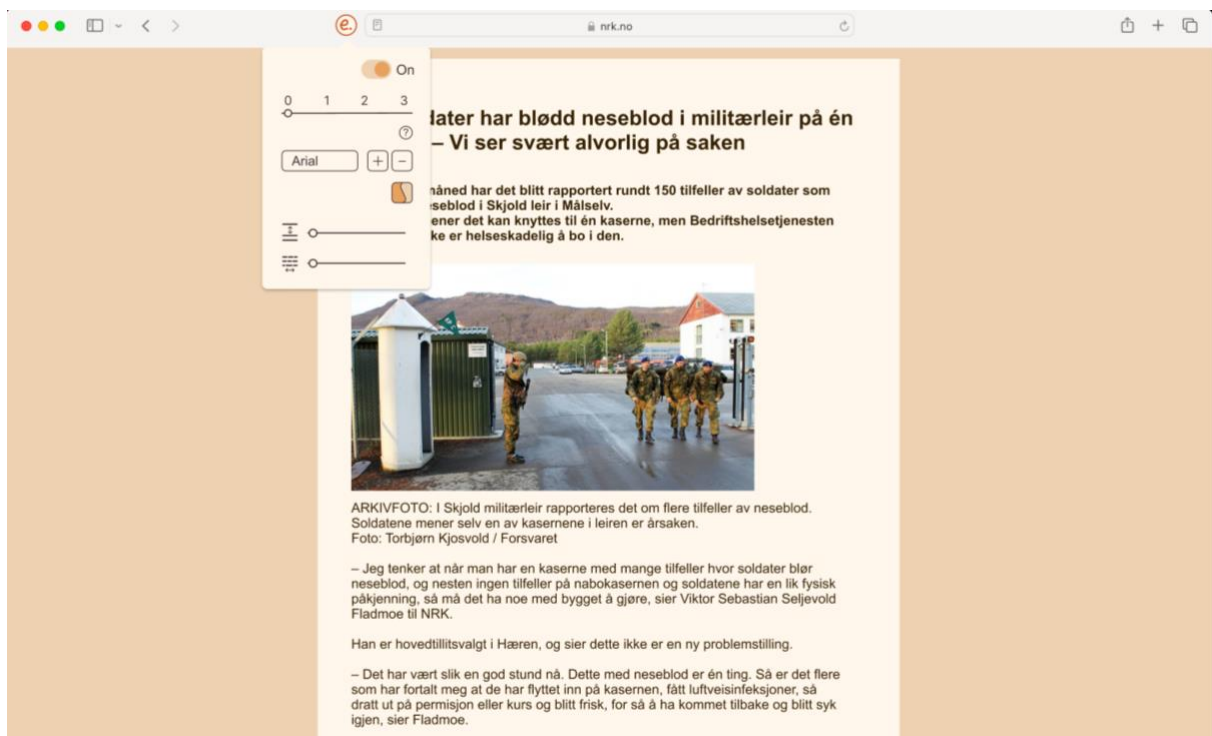
I tillegg ble det også utarbeidet flere funksjonelle krav som ble inkludert i de eksisterende kravene for prototypen:

- Knapp til å slå av/på nettleserutvidelse

- Knapp til å endre farge på tekst og bakgrunn
- Skyveknapp til å justere mellomrom mellom linjer
- Skyveknapp til å justere mellomrom mellom ord

5.4 Andre iterasjon av prototypen

Andre iterasjon av prototypen ble også utviklet i Figma og det er mulig å interagere med flere av de ulike funksjonene. Når man åpner prototypen, ser man en nyhetsartikkel åpen i en nettleser på en datamaskin. Øverst til venstre på siden er det en knapp for å åpne nettleserutvidelsen. Når man trykker på den, kommer det opp en av/på knapp som tar brukeren til en «leseversjon» av nyhetsartikkelen. Innholdet i artikkelen er lik, men reklamer og andre mulige forstyrrende elementer er fjernet. Dette er for å gi leseren valgmuligheten til å lese hele artikkelen dersom man vil, men i en mer minimalistisk versjon. Hvis man trykker på utvidelsesikonet igjen, vises en liten meny med ulike verktøy.



Figur 6: Første forenklingsnivå av andre iterasjon av prototypen.

Menyen har den samme av/på knappen som lar brukeren slå av utvidelsen. Det første verktøyet i menyen er en skyveknapp med fire nivåer for å justere grad av forenkling. I nærheten av denne er det en informasjonsknapp med et spørsmålstegn som gir en enkel

forklaring av de ulike forenklingsnivåene. Under er knappene for å endre skrifttype og størrelse på skriften. Et ikon med forskjellige farger viser til at man kan endre bakgrunnsfarge og skriftfarge. Helt nederst i menyen er det to skyveknapper for å justere avstand mellom linjer og ord. Det er enkle figurer ved siden av for å indikere hva de gjør. Retningslinjer utarbeidet spesielt for personer med dysleksi viser til at avstand mellom ord og linjer kan gjøre tekst mer lettlest for dyslektikere, og dette ble implementert i prototypen (British Dyslexia Association, u.å.-b). Man kan åpne og lukke menyen ved å trykke på utvidelsesikonet øverst i nettleseren.

For å minimere forstyrrelser er «leseversjonen» designet med et minimalistisk grensesnitt og dempede farger. En studie om lesbarhet for personer med dysleksi viser at bakgrunnsfarger kan ha en positiv effekt på lesbarheten, både for personer med og uten dysleksi. Det er funnet at mørk tekst på en varm og lys bakgrunnsfarge, som for eksempel fersken eller gul, er lettere å lese enn svart tekst på hvit bakgrunn (Rello & Bigham, 2017, s. 78). Derfor ble det valgt en lys varm bakgrunnsfarge og en mørk brun skriftfarge. I den opprinnelige teksten er det flere lange fremmedord, som for eksempel «bedriftshelsetjeneste», «oberstløytnant» og «hovedtillitsvalgt». Disse ordene ble derfor valgt som eksempler for å teste forklaring av ord i verktøyet.

I prototypen blir det brukt en artikkel fra NRK for å teste forenklingsnivåene og andre verktøy. Den valgte artikkelen omhandler en kaserne i en militærleir der 150 soldater har blødd neseblod i løpet av en måned og hvilke tiltak som blir gjort. Den valgte nyhetsartikkelen besto en rekke krav som ble utformet for å best mulig teste ulike funksjoner. Innholdet burde ikke være kontroversielt, slik at det ikke ble for mye fokus på temaet. I tillegg var det viktig at det ikke var en veldig stor sak som hadde mye mediedekning og derfor mindre sjanse for at deltakerne hadde lest artikkelen før brukertesten. For å teste ordforklaringsfunksjonen var det også et poeng å finne en artikkel som inneholdt fremmedord eller lange vanskelige ord. Det siste kravet var at artikkelen ikke skulle være for lang, slik at det var lettere for deltakerne å se gjennom hele artikkelen under brukertesten. Dette gjorde det også lettere å gjennomføre forenklingen, som blir diskutert nærmere i neste delkapittel.

menyen betydde før de fikk mulighet til å utforske prototypen gradvis. Det ble stilt spørsmål underveis i tillegg til en spørsmålsdel på slutten av testen der det ble stilt oppfølgingsspørsmål og generelle spørsmål om selve verktøyet. Resultatene fra brukertestene viste at skyveknappen med forenklingsnivå var vanskelig å forstå da ingen visste hva dette kunne være uten å prøve den først. Plasseringen av menyen var også forstyrrende da den skjuler deler av teksten når den er åpen. Det ble nevnt at menyen burde være på siden slik at man kan gjøre endringer på teksten underveis mens man leser.

Tilbakemeldingene på forenklingsnivåene av teksten viste at det var for liten endring mellom nivå én og to. Det ble påpekt at det var vanskelig å se forskjell mellom disse to nivåene. Det siste nivået av forenkling, som hadde den korteste versjonen av teksten, ble også oppfattet som for lang. Det ble bemerket at spesielt personer med kraftig dysleksi vil ha nytte av en enda kortere tekst. En annen tilbakemelding som gikk igjen på det siste nivået var at man savnet overskriftene og at det ikke var like lett å lese siste halvdel av artikkelen fordi det ikke var delt opp i avsnitt. Ved å ha lik avstand mellom alle linjene mistet man blikkfanget som man får ved å dele opp i avsnitt. Ordforklaringen ble fremhevet av samtlige deltakere som svært nyttig.

Etter å ha observert deltakernes interaksjon med prototypen, bestemte vi oss for å gjøre en forbedring av på-knappen til nettleserutvidelsen. Tidligere var det slik at brukeren måtte trykke på utvidelsesikonet og deretter på en separat «på»-knapp for å bli sendt til leseversjonen av nyhetsartikkelen. Etter grundig vurdering bestemte vi oss for å endre dette, slik at når brukeren trykker på utvidelsesikonet, blir de umiddelbart sendt til leserversjonen uten behov for et ekstra klikk. Dette eliminerer ett trinn i prosessen, gjør verktøyet mer tilgjengelig og gir en opplevelse som oppleves som raskere for brukeren.

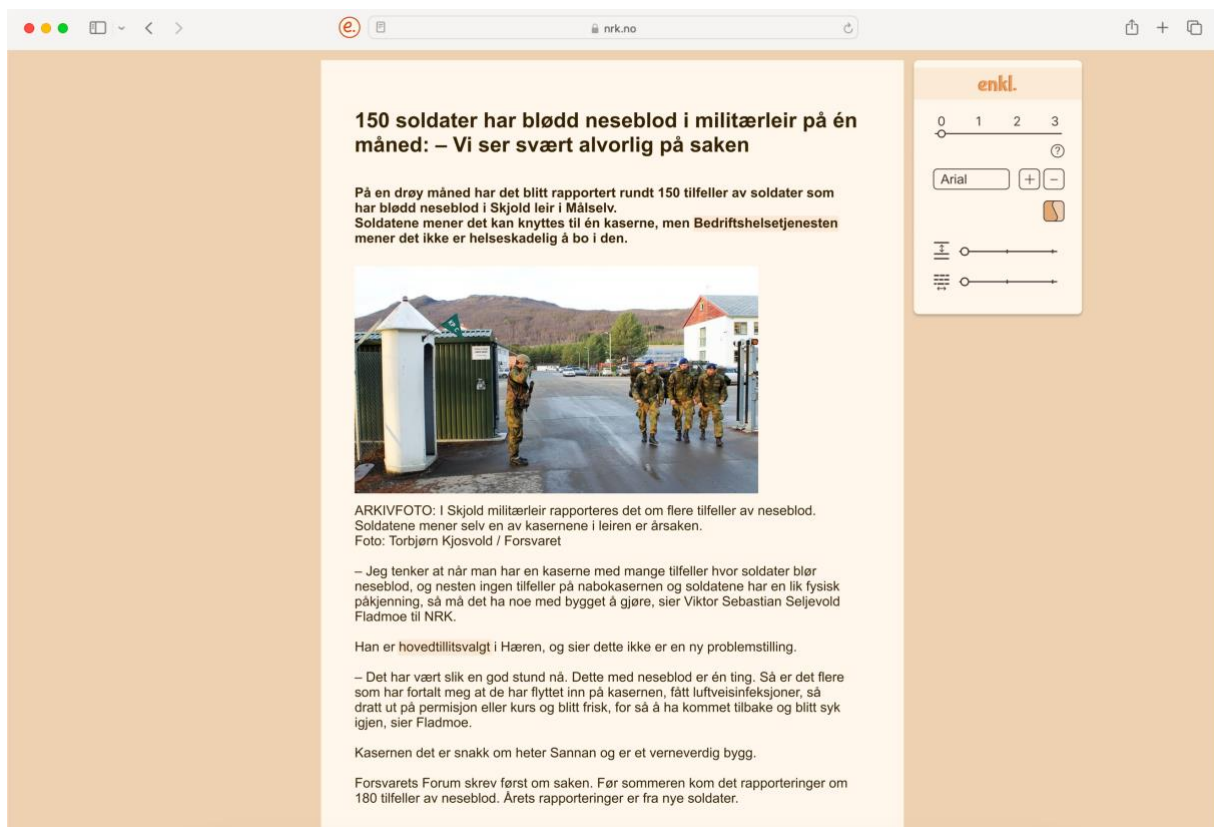
5.5 Tredje iterasjon av prototypen

Etter å ha mottatt tilbakemeldinger fra brukertestene, ble det gjort flere store justeringer for å forbedre teksten. Det første forenklingsnivået ble erstattet av nivå to slik at det var kortere setninger. Videre ble både nivå to og tre ytterligere forkortet. I det siste nivået ble det lagt til deloverskrifter og teksten ble delt opp i korte avsnitt for å gi økt lesbarhet, noe som var en direkte respons på tilbakemeldinger om at det var enklere å lese når det var visuelle punkter å feste blikket på. Det ble også tatt hensyn til at justering av mellomrom mellom linjene kan

gjøres ved behov, heller enn å ha en standard lik avstand gjennom hele teksten. Ingressen ble også forenklet på nivå to og i nivå tre ble setningene punktvis listet opp for å gjøre det enklere å skille mellom setningene. Ordavstand og linjeavstand ble nedjustert, da tilbakemeldingene viste at det var for stort mellomrom i første iterasjon. Flere farger i prototypen ble også implementert for å teste om det utgjorde forskjell i lesbarheten.

Følgende funksjonelle krav ble inkludert i de eksisterende kravene for prototypen:

- Ingressen i forenklingsnivå tre blir skrevet punktvis
- Setningslengden for nivå tre skal ikke overstige en linje
- Meny plasseres på høyre side av artikkelen
- Ordforklaringene kommer til syne når man trykker på ordet



Figur 7: Første forenklingsnivå i tredje iterasjon av prototypen.

5.5.1 Tekstforenklingsnivåer

I den tredje iterasjonen av prototypen valgte vi å bruke den samme artikkelen som utgangspunkt. Den forenklete teksten fra nivå to i den andre iterasjonen ble brukt som nivå én i denne versjonen. Det ble ikke gjort videre endringer i artikkelen.

I nivå to ga vi denne instruksen til språkmodellen: «Rewrite the following text in an easy-to-understand language for people with dyslexia and remove long, difficult words». Teksten ble redusert fra 543 ord til 178 ord og fra 37 setninger til 13. Siden vi vurderte at teksten ikke fulgte retningslinjer for klarspråk, se delkapittel 2.2, ble følgende grep gjort. En setning ble delt i to setninger. Vanskelige ord som bedriftshelsetjeneste ble fjernet i tillegg til deler av setninger som vi vurderte som mindre viktig i forhold til hovedbudskapet i artikkelen. I tillegg endret vi for eksempel «rengjøringsmidler» til det enklere «vaskemidler». I ett tilfelle ble verbtid endret fra «har vist» til «viser». Preposisjon «for» ble endret til «i». Pronomenet «du» ble endret til «de» to steder fordi det i den forenklete teksten viste til feil substantiv. På slutten av artikkelen endret vi rekkefølgen på to setninger for å gjøre innholdet mer tilgjengelig.

Det tredje nivået ble videre forenklet ved å be språkmodellen om å: «Rewrite the following text in a short, easy-to-understand language for people with dyslexia and remove, long difficult words». Også her anså vi det som nødvendig å gjøre manuelle justeringer i teksten. Som tidligere ble to setninger delt og deler av setninger fjernet. Vi valgte å fjerne fornavn og to mellomnavn og la etternavnet så igjen. Språkmodellen endret tittelen hovedtillitsvalgt til hovedansvarlig, noe som ikke har samme betydning. Dette ble derfor endret tilbake. Også her endret vi i et tilfelle verbtid fra fortid til nåtid (sa til sier). Fem av de seks siste setningene i den automatisk forenklete teksten startet med pronomenet «vi». For å skape bedre flyt i teksten endret vi for eksempel setningene «Vi skal undersøke området på våren» til «Området skal undersøkes på våren» og «Vi skal se om vi kan ha færre personer per rom» til «De skal se om de kan ha færre personer per rom».

5.5.2 Brukertesting av tredje iterasjon

Det ble gjennomført totalt ti brukertester hvorav fire deltakere hadde deltatt i den første runden, mens seks deltakere var nye og ikke hadde sett prototypen tidligere. Testene ble gjennomført med den samme grunnleggende strukturen som i den første brukertesten. Deltakerne startet med å åpne NRK-artikkelen for å få en generell oversikt. Deretter ble de bedt om å aktivere nettleserutvidelsen og gradvis utforske de ulike funksjonene i menyen. Deltakerne som ikke hadde deltatt i den første brukertesten, ble også spurt om deres førsteinntrykk av verktøyet og hva de antok de forskjellige alternativene i menyen gjorde.

Deltakerne som deltok for andre gang, ble også spurt om deres tanker rundt endringene som var gjort siden forrige iterasjon.

Det var betydelige variasjoner i tilbakemeldingene når det gjaldt de ulike nivåene av tekstforenkling. Noen deltakere var spesielt positive til det tredje forenklingsnivået, blant annet på grunn av korte setninger og avsnitt, samt evnen til å gi en oppsummering av den opprinnelige artikkelen. Det var flere som uttrykte at de sannsynligvis ville velge forenklingsnivå tre, da de følte at de fikk tilstrekkelig informasjon til å få et overblikk over innholdet i artikkelen. En av deltakerne uttrykte ønske om ytterligere forenkling av det tredje nivået av tekstforenkling. Vedkommende foreslo at teksten kunne bli kortere og presentert i punktform, lignende ingressen i artikkelen. Det er verdt å merke seg at denne deltakeren også deltok i den første brukertesten og hadde derfor allerede sett artikkelen én gang tidligere. Denne tidligere erfaringen kan ha påvirket vedkommendes forståelse av artikkelen og bidratt til ønsket om ytterligere forenkling.

To deltakere uttrykte bekymring for å gå glipp av informasjon når de leste en forkortet versjon av artikkelen. Dette var spesielt på det tredje forenklingsnivået at de uttrykte at det følte som om det manglet informasjon. En bruker, spesielt, påpekte at de vanligvis husker tekst visuelt, så de merket umiddelbart hvis det manglet innhold i teksten i de ulike forenklingsnivåene. Det ble likevel påpekt at dersom man savnet informasjon på et høyere forenklingsnivå, kunne man gå ned et nivå for å få flere detaljer. Enkelte sa at de ville foretrekke å holde seg til den originale teksten og gjøre mindre endringer, som justering av linjeavstand eller fargetema.

Som nevnt uttrykte en av deltakerne ønske om ytterligere forenkling. Dette synspunktet illustrerer at det er ulike preferanser og behov blant deltakerne når det gjelder tekstforenkling. Mens noen deltakere kan oppleve det tredje nivået som tilstrekkelig, kan andre ha behov for en enda mer strukturert og kortfattet presentasjon av informasjonen. Det er også viktig å ta hensyn til deltakernes tidligere erfaringer og kunnskap, da dette kan påvirke deres oppfatning av tekstens forståelighet. Ved å ta i betraktning individuelle preferanser og tidligere erfaringer kan man utvikle tekstforenklingsverktøy som bedre imøtekommer brukernes behov og bidrar til en mer tilgjengelig leseopplevelse.

Selv om de fleste brukerne hadde en preferanse for et bestemt forenklingsnivå, ble det påpekt av flere at deres valg av forenklingsnivå kunne variere avhengig av innholdet og lengden på en artikkel. For eksempel, hvis artikkelen omhandlet et tema som de hadde spesiell interesse for, ville de være mer tilbøyelige til å velge å lese hele artikkelen for å få med seg alle detaljene. På den andre siden, når det gjaldt generelle emner eller artikler som var vanskelig å forstå, ville de heller velge en mer forenklet versjon. Den individuelle tilnærmingen til tekstforenkling varierte også basert på brukernes personlige behov og preferanser. Noen deltakere uttrykte at de foretrakk å få en grundig forståelse av et emne, og derfor ville de velge å lese den originale teksten selv om den kunne være mer kompleks. De var villige til å investere tid og innsats for å fordype seg i materialet. Andre deltakere mente imidlertid at det var mer hensiktsmessig å velge en mer forenklet versjon for å få en rask oversikt eller et overblikk over artikkelen uten å bli overveldet av detaljer. Deres mål var å oppnå en grunnleggende forståelse uten å gå for dypt inn i faglige eller tekniske detaljer.

Deltakerne ga tydelig uttrykk for at underoverskriftene i artikkelen var spesielt nyttige og bidro til å skape orientering og en jevn overgang mellom ulike avsnitt mens de leste. De fungerte som navigasjonspunkter som hjalp brukerne med å forstå strukturen i teksten og holde seg fokusert på relevante deler av artikkelen. Ordforklaringsfunksjonen viste seg å være ekstremt støttende for alle deltakerne, og den ble ofte nevnt som en av favorittfunksjonene i prototypen. Dette skyldes at deltakerne vanligvis måtte søke etter forklaringer på vanskelige ord på egenhånd, noe som kunne være tidkrevende og distraherende fra hovedteksten. Med ordforklaringsfunksjonen fikk de umiddelbart tilgang til definisjoner av vanskelige ord ved å bare klikke på dem. Deltakerne satte også pris på at de gitte forklaringene i eksempelartikkelen var korte og presise. Dette gjorde det enklere for dem å få en rask forståelse av ordene uten å bli overveldet av omfattende eller komplisert forklaring. Noen deltakere påpekte også at denne funksjonen kunne hjelpe dem med å utvide vokabularet sitt og lære nye ord i prosessen.

Etter tilbakemeldingene fra brukertestene ble det tillagt noen flere funksjonelle krav for prototypen:

- Knapp som lar brukeren lagre nåværende innstillinger til neste gang man åpner verktøyet.
- Linjeavstand og ordavstand på ordforklaringene korresponderer med innstillingene valgt for teksten.

6 RESULTATER

Dette kapittelet presenterer fjerde iterasjon av prototypen og tilbakemeldinger. 6.1 presenterer fjerde iterasjon av prototypen, endelig resultat og ikke-funksjonelle krav. 6.2 presenterer en sammenligning av automatisk tekstforenkling av ti ulike nyhetsartikler.

6.1 Fjerde iterasjon – et ferdig produkt?

I den fjerde iterasjonen brukte vi samme artikkel som i tredje iterasjon, og de samme forenklingsnivåene for teksten. Vi gjorde kun små, visuelle endringer basert på evalueringen i den andre runden med brukertester. Det ble implementert en lagringsfunksjon som tillater brukerne å lagre sine preferanser, inkludert forenklingsnivå, linjeavstand og ordavstand. Når dette valget blir aktivert, vil de samme innstillingene være aktive ved neste bruk. Hensikten med denne nye funksjonen er å unngå ekstra trinn med å angi de samme innstillingene hver gang en ny artikkel blir valgt. Samtidig har brukerne som ikke ønsker automatisk innstilling fortsatt mulighet til å starte fra standardinnstillingene hver gang, hvis de ikke aktiverer denne valgmuligheten. Videre ble det også utført endringer i ordforklaringene slik at linjeavstand og ordavstand var i samsvar med de valgte innstillingene for teksten.

6.1.1 Evaluering av prototypen

Beslutningen om å presentere prototypen for eksperter og relevante organisasjoner i stedet for å gjennomføre en ny runde med testing blant målgruppen var en strategisk avgjørelse. Målet var å få innspill fra personer med ekspertise og potensielle brukere som kunne gi verdifulle perspektiver. Ved å involvere Dysleksi Norge, NTB/Arkitekt og Klar Tale, ønsket vi å samle innspill fra fagpersoner og aktører med kunnskap og erfaring innen feltet. Responsen på prototypen var overveiende positiv, noe som indikerte at de implementerte endringene og funksjonalitetene ble godt mottatt.

En gjennomgående tilbakemelding var imidlertid ønsket om å inkludere en tekst-til-tale-funksjon i prototypen. Det ble blant annet nevnt at for mange med dysleksi er det lettere å lese tekst når den blir lest opp samtidig. Selv om denne ideen tidligere hadde blitt vurdert, ble den ikke implementert på grunn av prosjektets primære fokus på tekstforenkling og tekniske

begrensninger innenfor Figma-utviklingsmiljøet. Det var ikke mulig å teste en slik funksjon i den nåværende utviklingsfasen. Disse begrensningene og prioriteringene var avgjørende for beslutningen om å ikke inkludere tekst-til-tale-funksjonen i prototypen, til tross for forespørslene om dette.

Tilbakemeldingene vi mottok fra Klar Tale angående prototypen var overveiende positive. De viste stor interesse for verktøyet og var ivrige etter å teste forskjellen mellom den automatiske forenklete versjonen av teksten og teksten som de forenkler manuelt. En spesifikk tilbakemelding dreide seg om ordforklaringene. Det ble påpekt viktigheten av å unngå ord eller uttrykk som har flere betydninger. Det ble også foreslått å forenkle ordforklaringene ytterligere og kanskje tilby flere nivåer av ordforklaringer for å imøtekomme ulike behov og forståelsesnivåer.

6.1.2 Endelig resultat av prototypen og videre endringer

Basert på tilbakemeldingene av den fjerde iterasjon av prototypen, ble det gjort noen små justeringer. Ettersom det ble påpekt flere ganger at en lytt-til-tekst-funksjon ville være svært nyttig, ble det lagt til et lytteikon i menyen som indikerer at hele artikkelen blir opplest.

Ordforklaringene ble nøye gjennomgått og videre forenklet for å sikre en enda bedre brukeropplevelse. Som en del av denne prosessen gjennomførte vi tester der vi eksperimenterte med ulike instruksjoner for å forklare begrepene «bedriftshelsetjeneste», «hovedtillitsvalgt» og «forsvarsbygg», på samme måte som vi gjorde med forenklingsnivåene. Målet vårt var å finne den mest effektive tilnærmingen for å presentere ordforklaringene, slik at de var lette å forstå og ikke inneholdt flertydige begrep. Etter en grundig evaluering kom vi frem til at instruksjonen «Forklar _ i en kort setning med enkle ord» ga de beste resultatene for to av ordforklaringene. Denne instruksjonen klarte å formidle betydningen av ordene på en klar og enkel måte. Imidlertid fant vi ut at den samme instruksjonen ikke fungerte like godt for alle forklaringene. For «hovedtillitsvalgt» ga instruksjonen: «Forklar i en kort setning med korte og enkle ord hva _ betyr» en litt kortere og enklere forklaring. Likevel var vi bevisste på behovet for en enhetlig tilnærming og ønsket å unngå å ha forskjellige instruksjoner for ulike ordforklaringer. Selv om ordforklaringen for «hovedtillitsvalgt» ikke var dårlig, var den litt lengre enn vi hadde ønsket. Derfor besluttet vi at det var mer hensiktsmessig å ha kun én instruksjon for alle ordforklaringene. Totalt sett

valgte vi derfor å bruke instruksjonen som ga de beste resultatene for to av de tre ordforklaringene, slik at vi kunne sikre en enhetlig tilnærming og optimal brukeropplevelse.

Selv om vi har gjennomført forenkling av ordforklaringene, erkjenner vi viktigheten av å få en ekstern vurdering fra eksperter på klarspråk eller personer i målgruppen. Ved å la erfarne språkspecialister eller personer med dysleksi evaluere de nye ordforklaringene, kan vi sikre at de oppfyller de nødvendige kravene til tydelighet, tilgjengelighet og effektiv kommunikasjon. Deres innsikt og tilbakemeldinger vil være viktig for å validere kvaliteten og gjøre eventuelle justeringer for å ytterligere forbedre ordforklaringene og sikre at de er optimalisert for brukernes behov.

6.1.3 Ikke-funksjonelle krav

- Sharp et al. beskriver ikke-funksjonelle krav som karakteristikkene til produktet (Sharp et al., 2019, s. 390). Følgende ikke-funksjonelle krav ble utformet for prototypen:
- Minimalistisk design med et begrenset antall funksjoner som er nøye utvalgt for å møte brukerens behov og unngå overflødige eller forvirrende elementer.
- Maks 1 sekund ventetid når man endrer forenklingsnivå
- Respons på endringer i forenklingsnivå skal være umiddelbar, slik at brukeren raskt kan se resultatene av endringene de gjør på teksten.
- Tilgjengelig for nedlastning til alle store nettlesere, inkludert Firefox, Chrome, Safari, Microsoft Edge, osv.
- Minimere gjennomsnittslesetid
- Verktøyet bør være tilgjengelig som en nettleserutvidelse for PC og nettbrett, samt som en mobilapplikasjon.
- Støtte skjermlesere og tastaturnavigasjon på alle enheter, slik at det blir tilgjengelig for brukere med ulike funksjonshemninger.
- Verktøyet skal kunne håndtere store mengder tekst uten at det går utover responstiden, slik at brukeren kan bruke det effektivt uavhengig av tekstmengden.

6.2 Sammenligne automatisk tekstforenkling av ti nyhetsartikler

Artikkelen som ble brukt i prototypen ble spesifikt valgt med tanke på brukertestene (se delkapittel 5.4). Imidlertid er det viktig å merke seg at nyhetsartikler kan variere betydelig i skrivestil, avhengig av avisen og innholdet de tar for seg. For å oppnå en bredere sammenligning og undersøke om det var andre utfordringer med språkmodellen som ikke ble avdekket i prototypen, ble det valgt ut ti forskjellige artikler. Utvalget inkluderte ulike nettaviser, en nettside om geoforskning og et nettsted for kjendis- og underholdningsnyheter. Samlingen bestod av åtte nyhetsartikler, en kronikk og en fagartikkel. Nyhetsartiklene dekket ulike emner som sport, politikk, kjendisnyheter, korrupsjon og kunstig intelligens. Videre var det betydelig variasjon i lengden på tekstene, med den korteste artikkelen på 195 ord og den lengste på 1407 ord. Ved å inkludere denne variasjonen kunne vi få et mer nyansert bilde av språkmodellens ytelse og utfordringer i forskjellige kontekster.

Forenklingen av artiklene ble utført på samme måte som i den tredje iterasjonen av prototypen. Prosessen involverte å dele teksten omtrent i to like lange deler, med unntak av to av tekstene som var under 210 ord hver. Disse artiklene ble ikke delt på grunn av deres korte lengde. Deretter ble tekstene kjørt gjennom språkmodellen med de samme instruksjonene som tidligere for de ulike nivåene (se delkapittel 0). For nivå én og to ble maks lengden satt til 1500 tokens, mens for nivå tre var maks lengden 1200 tokens. Resultatene av den automatiske forenklingen ble sammenlignet ved å legge de forenklede versjonene inn i et Excel-dokument. Videre ble de genererte tekstene gjennomgått manuelt for å identifisere eventuelle brudd på klarspråksregler. Setninger og deler av teksten ble deretter fargekodet basert på egenskaper som gjør teksten mindre tilgjengelig for målgruppen. Denne fargekodingen bidro til å identifisere områder som krever ytterligere forenkling eller justering for å oppnå bedre lesbarhet for personer med dysleksi.

Det var flere likhetstrekk som gikk igjen i mange av de forenklede tekstene. Alle forenklingsnivåene inneholdt setninger som var for lange for målgruppen og burde vært delt inn i to eller flere setninger. Det var også lange eller vanskelige ord i de fleste eksempeltekstene som enten burde erstattes med enklere synonymer eller få tildelt ordforklaring slik som i prototypen. Det var mindre grad av vanskelige ord i tekstene på nivå tre enn i de andre nivåene. Flere av de forenklede versjonene inneholdt setninger som var vanskelig å forstå uten å ha lest den originale artikkelen. Et eksempel på dette er et avsnitt fra

forenklingsnivå én av en artikkel om lanseringen av billigbøker på begynnelsen av 1900-tallet: «På begynnelsen av 1900-tallet ville noen borgere gi arbeiderklassen verktøy for befrielse. Olof Palme forlot borgerskapet og gikk over til venstresida. I dag handler det om å konsumere bøker, ikke dannelse.». Delen av teksten som dette er hentet fra, består av en rekke sitater der det veksles mellom intervjudeltaker og journalist. Dette var gjennomgående, at i artikler som inneholdt sitater, ble det i den forenklete versjonen flere logiske brister og setninger som var usammenhengende. Når informasjonen blir kortet ned og gjenfortalt i en forenklet versjon vil naturligvis noe av innholdet forsvinne, og kanskje spesielt i tekster med mange sitater. I videre arbeid med prototypen ville dette vært aktuelt å se nærmere på.

Det var i tillegg flere eksempler på at språkmodellen feiltolket informasjon og gjenga innhold som ikke var riktig. Dette ble spesielt tydelig i en fagartikkel om hydrotermiske skorsteiner der man beskriver hvordan disse fungerer. Her blir gjengivelsen upresis og vanskelig å forstå. Det er også flere eksempler på informasjon som blir gjengitt på en måte som gjør at strukturen på setningene ikke fungerer godt. Et eksempel på dette er fra en artikkel om utrykningskjøretøy som opplever at de blir hindret i trafikken: «Tom Løvskogen kjører med brannbil og sirener, blålys og horn. En brøytebil ser ikke etter ham. Bjørn Sveen svinger med ambulansebil.». Her blir innholdet fra den originale saken mer eller mindre beholdt, men formidlingen er «klønete» og mindre tilgjengelig.

Retningslinjer for klarspråk nevner blant annet at dersom man bruker henvisningsord som «de» og «den» er det viktig at det er klart for leseren hvem eller hva man viser til (Språkrådet, 2013). I artiklene som ble forenklet var det et par tekster som ikke synes å stå til dette kravet. Et eksempel der bruk av pronomenet «de» virket utydelig i teksten, er i tidligere nevnte artikkel om undersjøiske skorsteiner: «Ofte er det de første tingene de ser når de nærmer seg de varme kildene på havbunnen. De kan strekke seg opptil flere meter.». Her blir «de» brukt fem ganger, men det siktes til forskjellige ting. Innholdet fremstår dermed utydelig, og man blir avhengig av å ha lest den originale artikkelen for å forstå teksten.

En observasjon som kom tydeligere frem i sammenligningen av de ti artiklene er at i halvparten av eksempeltekstene er den totale tekstlengden til nivå to lenger enn nivå én. Det er ingen tilfeller der den totale tekstlengden til nivå tre er lenger enn nivå to. Det er likevel tre tilfeller der den ene halvdel av teksten i nivå tre inneholder flere ord enn tilsvarende halvdel

i nivå to. En annen observasjon er at i nivå to deler språkmodellen ofte opp teksten i mindre avsnitt. Dette blir ikke gjort i nivå én og ikke i like stor grad i nivå tre.

6.2.1 Problem med sitater

Vær Varsom-plakaten skildrer presseetiske normer som må følges innen journalistikken. Punkt 3.7 sier følgende: «Kilder skal gjengis korrekt. Pressen har plikt til å gjengi meningsinnholdet i det som brukes av intervjuobjektets uttalelser. Ingen kilder har krav på å bli gjengitt i sin helhet.» (Pressens Faglige Utvalg, 2021). I nyhetsartikler er det vanlig å bruke sitatstrek for å bearbeide ordlyden i utsagn og gjøre budskapet tydeligere og mer effektivt. Anførselstegn brukes derimot kun for å gjengi mest mulig ordrette sitater (Språkrådet, 2023). Det er også mulig å gjengi utsagn med indirekte sitat der innholdet blir gjenfortalt med andre ord (Søk og Skriv, 2023). Et eksempel på dette kan være: «Ifølge Henriksen, har den norske kronen aldri vært svakere». Ni av artiklene som ble forenklet inneholdt sitater, både med anførselstegn og sitatstrek. Dersom innholdet fra et sitat ble gjengitt i den forenklete versjonen, var det som regel indirekte sitat. Ved to tilfeller ble et sitat som var markert med anførselstegn, gjengitt med en annen ordlyd, men anførselstegnene ble likevel beholdt. Dette skjedde i artikkelen fra Klassekampen der sitatet «pumpet befolkningen full av falsk bevissthet» ble til «fylte folk med feil oppfatning» i forenklingsnivå én og «fylte befolkningen med feil bevissthet» i nivå to. Det siste eksempelet inneholder også en skrivefeil. Det var også eksempler der sitater med personlige meninger blir gjengitt som indirekte sitat som en bekreftet sannhet og blir dermed misvisende.

7 DISKUSJON

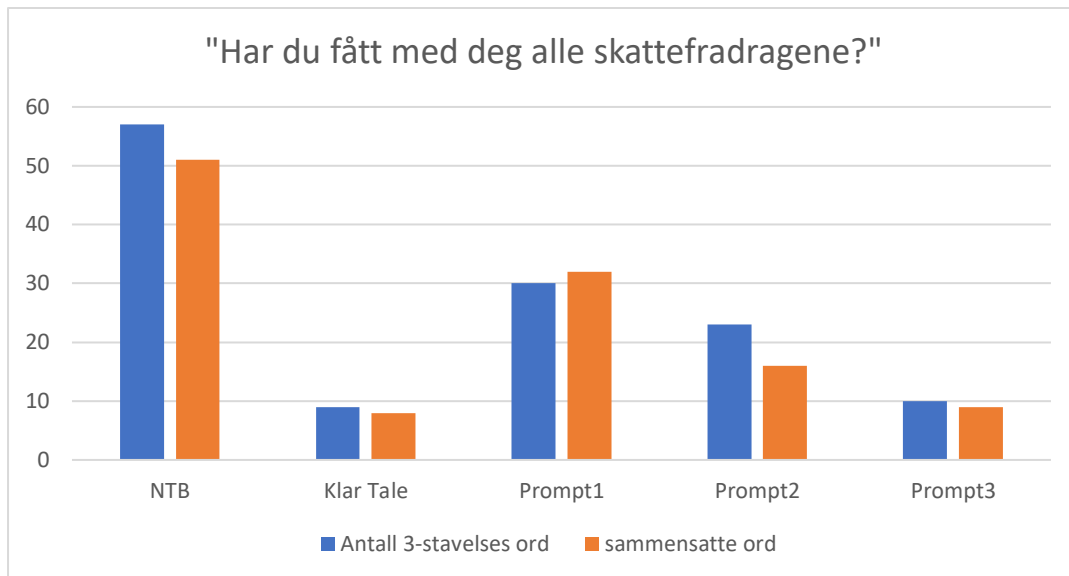
Formålet med denne studien har vært å, gjennom iterasjoner, utvikle en prototype som automatisk forenkler nyhetsartikler tilpasset personer med dysleksi. I dette kapitlet diskuterer jeg tekstforenklingstilnærmingen som har blitt gjort og effektiviteten av den.

7.1 Automatisk tekstforenkling

I dette delkapitlet vil jeg gjennomføre en grundigere evaluering av tekstforenkling ved å sammenligne den med en artikkel fra NTB og en fra Klar Tale. Videre blir tekstforenklingen i prototypen og promptene analysert. Til slutt oppsummeres gjennomgående utfordringene i de genererte tekstene og kvaliteten vurderes.

7.1.1 En sammenligning av tekstforenklingstilnæringer: NTB, Klar Tale og GPT-3

For å videre teste kvaliteten på tekstforenkling utført av GPT-3, ble en nyhetsartikkel fra NTB forenklet. Samme artikkel var allerede forenklet av avisen Klar Tale, og dette ble brukt til sammenligning. Artikkelen handler om at mange ikke har levert skattemeldingen sin enda og at fristen for å levere snart går ut. Det ble utført tre nivåer av forenkling med GPT-3 og det ble utført med samme metode som i tredje iterasjon av prototypen, se delkapittel 0. Den originale teksten er på 604 ord og Klar Tales versjon er på 180 ord. De tre forenklingsnivåene er på henholdsvis 390 ord, 307 ord og 111 ord. Se **Error! Reference source not found.** for en sammenligning av antall tre-stavellesord og sammensatte ord i hver artikkel.



Figur 8: Graf som viser antall trestavelserord og sammensatte ord i de respektive versjonene av artikkelen om skattefradrag.

I tekstene som ble generert av GPT-3 hadde alle forenklingnivåene setninger med “klønete” språk som gjør tekstene mindre tilgjengelige. Et eksempel på dette er denne setningen fra nivå to: «Dette er større tall enn samme tid i fjor.». Innholdet er korrekt, men måten setningen er skrevet gjør at flyten stopper opp. Et annet eksempel er setningen fra forenklingnivå tre: «Leveringsfristen er 2.mai, hvis du trenger forsinkelse, må du søke innen 30.april» (se Tabell 3). Her har ordet «utsettelse» fra NTB-artikkelen blitt erstattet med synonymet «forsinkelse». Selv om ordene har tilsvarende betydning, blir det i denne sammenhengen grammatisk ukorrekt. Alle de genererte tekstene inneholder lange setninger som burde deles opp. Dette var spesielt tydelig for de to første forenklingnivåene.

Forenklingnivå tre og artikkelen av Klar Tale er mest lik innholdsmessig. Selv om Klar Tale-artikkelen er lenger enn nivå tre, er den mer sammenhengende og enklere å forstå. I nivå tre-teksten er for eksempel ordet “forhåndsutfylt” brukt slik som i NTB-artikkelen. Dette kunne med fordel ha blitt erstattet og skrevet på en lettere måte. I versjonen til Klar Tale er dette gjort. Selv om nivå-tre-teksten ikke fungerer på alle områder, er den viktigste informasjonen bevart og det er ikke nødvendig for mottakeren å ha lest versjonen til NTB for å forstå innholdet. Det er likevel nødvendig med noen manuelle justeringer. I nivå tre blir det brukt anførselstegn for å sitere Skattedirektøren slik: «Sjekk at opplysningene er riktige og husk å legg til det som ikke er forhåndsutfylt.». Informasjonen er hentet fra et indirekte sitat, men ordlyden er endret og fremstilt som et direkte sitat. Dette bryter med presseetiske regler for gjengivelse av sitater og bruk av anførselstegn (se delkapittel 6.2.1).

Klar Tale	Forenklingsnivå 3
<p>Skatteetaten sier at mange ennå ikke har sjekket skattemeldingen sin. 30. april er siste frist for å sjekke at all informasjon er riktig.</p> <p>Rundt 1,5 millioner har allerede fått skatteoppgjøret sitt. Mens 2,6 millioner har levert skattemeldingen.</p> <p>Mens mange har enten levert og fått igjen penger, er det også noen som ennå ikke har gjort det. Skatteetaten sier at det er flere hundre tusen som ikke har åpnet den skattemeldingen.</p> <p>– Det er ditt ansvar å sjekke at opplysningene er riktig. Det sier Nina Schanke Funnemark. Hun er skattedirektør.</p> <p>Hun sier det er viktig å ikke bare sjekke om du får igjen penger eller om du må betale tilbake. Du må også sjekke at all informasjon er riktig.</p> <p>– Det kan være at du skal ha mer penger til gode enn hva skattemeldingen viser, sier hun.</p> <p>Blant annet må du selv fylle ut fradrag fra reiser. Det samme gjelder informasjon om bolig eller salg av bolig.</p> <p>Siden fristen er på en søndag, er det heller ikke for sent om du leverer tirsdag 2. mai. Du kan også søke om å utsette fristen (Skipperø & NTB nyheter, 2023)</p>	<p>Rundt 2,6 millioner har levert skattemeldingen sin og 1,5 millioner har fått utbetalt skatt til gode.</p> <p>Skattemeldingen er sendt til 4,2 millioner personer og 3,8 millioner har åpnet den.</p> <p>Noen har glemt å legge til opplysninger og risikerer å få mindre penger til gode. Skattedirektøren sier: «Sjekk om opplysningene er riktige og husk å legge til det som ikke er forhåndsutfyllt».</p> <p>Hvis du trenger mer tid til å gjøre dette, kan du søke om utsettelse av fristen. Leveringsfristen er 2.mai, hvis du trenger forsinkelse, må du søke innen 30.april.</p> <p>Når skattemeldingen er levert, lager Skatteetaten et skatteoppgjør.</p> <p>De fleste vil ha fått skatteoppgjøret innen juni, men noen må vente til høsten.</p>

Tabell 3: Artikkel av Klar Tale og forenklingsnivå tre. Setningen fra eksempelet er markert med gult.

7.1.2 To iterasjoner av automatisk tekstforenkling: Fremgangsmåte og refleksjoner

For å forenkle artiklene valgte vi å bruke GPT-3 (se delkapittel 5.1.1), for at artiklene skulle bli mest mulig lesbare for målgruppen. Vi var imidlertid forberedt på å gjøre manuelle endringer i den genererte teksten. Vi brukte retningslinjer for klartekst som utgangspunkt for endringene. Når det gjelder artikkelen vi brukte i brukertesten hadde vi en del krav (se delkapittel 5.4) som ble utformet med tanke på brukertestene.

Første iterasjon av prototypen

I første iterasjon av prototypen brukte vi den tekniske prototypen til å utforske hvordan vi skulle forenkle tekster. Etter å ha kjørt tekstene gjennom GPT-3 så vi at den forenklete versjonen ble for kort og manglet mye informasjon. Dette skyldes blant annet lengden på nyhetsartikkelen. Davinci modellen (se delkapittel 5.1.1) kan prosessere opp til 2049 tokens

per «request» (OpenAI, u.å.-b). Det betyr at man må ta i betraktning mengde input i forhold til ønsket output. Dersom inputen er lang, vil man bruke opp mesteparten av tokenene på å prosessere teksten og dermed få en relativt kort tekst tilbake. Vi ville ha tre nivåer av forenkling. Derfor måtte vi finne en måte å løse det på slik at den genererte teksten hadde «passende» lengde i de tre nivåene. Vi forsøkte først å dele teksten opp i deler, ulike prompts og ulike maks lengder for å finne den riktige metoden. Det ble gjort flere tester med ulik lengde av tokens for generert tekst. Dersom teksten ble delt i to, på omtrent likt antall ord, ble den genererte teksten lengre. Det gjorde det lettere å få til tre distinkte nivåer. Vi endte med å sette maks lengde til 1500 for det første nivået. Dette gjorde at den nedkortede versjonen var merkbart kortere uten å miste «for mye» informasjon.

Andre iterasjon av prototypen

I den andre iterasjonen av prototypen ble det utviklet tre forskjellige nivåer av forenklede versjoner for nyhetsartikkelen. Disse nivåene inkluderer originalteksten (nivå null) samt gradvis mer forenklede versjoner (nivå én til tre). For å finne den mest effektive tilnærmingen testet vi ulike prompts og instruksjoner (se tabell 3). Testene avdekket visse forskjeller i resultatene basert på formuleringen av promptene. Noen instruksjoner resulterte i svært korte sammendrag av artikkelen, men de inneholdt fortsatt noen lange og vanskelige ord. For eksempel ble ord som «bedriftshelsetjeneste» og «oberstløytnant» ikke fjernet av noen av instruksjonene. Basert på disse resultatene konkluderte vi med at prompten «Rewrite the following text in an easy-to-understand language for people with dyslexia:» ga det beste resultatet for nivå én av forenklingen. Resultatet var en nedkortet versjon av teksten som likevel beholdt mye av informasjonen fra den opprinnelige artikkelen. Et viktig hensyn var å sikre at nivå én ikke var for drastisk forskjellig fra originalartikkelen, slik at det var rom for ytterligere forenkling i de påfølgende nivåene. For å bedre oppfylle klarspråksretningslinjene var det imidlertid nødvendig å gjøre visse manuelle endringer i tekstene (se delkapittel 5.4.1).

Rewrite the following text with easy words and short sentences:

Rewrite the following text in an easy-to-understand language for people with dyslexia:

Rewrite the following text in an easy-to-understand language for people with dyslexia and exchange long words with a shorter synonym:

Rewrite the following text in an easy-to-understand language and in short sentences:

Lightly summarize the following text in an easy-to-understand language for people with dyslexia:

Summarize the following text in an easy-to-understand language for people with dyslexia:

Tabell 4: Oversikt over et utvalg prompts som ble testet. Teksten markert i gul er prompten som ble valgt.

I nivå to beholdt vi den samme teksten som i nivå én, men gjorde manuelle endringer som å dele opp lange setninger. Grunnen til at endringen ble gjort for hånd, var for å kontrollere at dette nivået var hakket mer forenklet enn det før, men likevel ikke for kort. Siden dette var første iterasjon av prototypen som skulle vurderes gjennom brukertester, var vi usikre på hvor forenklet nivåene burde være. Vi ville bevare mest mulig av innholdet, og det var vanskelig å vite akkurat hvor vi skulle legge det mellomste nivået. Det var også et poeng at endringene, selv om de ikke ble gjort av en generativ språkmodell, skulle speile endringer som man i teorien kan kode.

Det siste forenklingsnivået viser den mest forenklete versjonen av teksten. For dette nivået ble det brukt en annen prompt: «Rewrite the following text with easy and short words» og 1200 tokens. Dette for å få en enda kortere versjon av teksten. Denne instruksjonen ga ikke den korteste teksten, men den inneholdt korte setninger og gjenga innholdet i originalteksten i et tydelig og enkelt språk. Det var likevel behov for å gjøre noen få manuelle justeringer som å dele en setning i to setninger.

Tredje iterasjon av prototypen

I møte med brukerne viste det seg at forenklingen av teksten ikke var god nok. Nivå én og to var for like, da vi fikk tilbakemelding om at deltakerne ikke så forskjell på disse nivåene. Det var derfor behov for ytterligere forenkling av nivå to. Tilbakemeldingene for nivå tre var også at dette nivået burde forenkles drastisk, spesielt siden dette nivået kanskje er rettet spesielt mot de som har ekstra behov for tilrettelagt tekst. Disse tilbakemeldingene ble brukt som utgangspunkt for videre arbeid med forenklingsnivåene. Den forenklete teksten fra nivå to i den andre iterasjonen ble brukt som nivå én i denne versjonen. Årsaken til dette var på grunn av tilbakemeldinger fra brukerne. De ga uttrykk for at nivå én og to var for like. De var likevel positive til grad av forenkling i forhold til forenklingsnivået. Selv om teksten var mye kortere, inneholdt den fremdeles mye informasjon. Det ble derfor sluttet at nivå to, som hadde kortere setninger, var bedre å bruke som nivå én i tredje iterasjon.

Under testingen av det andre og tredje nivået av forenkling utførte vi flere tester med ulike prompts for å videre forenkles artikkelen. Ingen av de testede promptsene ga et resultat som oppfylte kravene til klarspråk. På lik linje med tidligere iterasjoner av prototypen måtte det utføres manuelle endringer i teksten for å oppnå det vi betraktet som en mer forståelig tekst.

7.1.3 Evaluering av prompts

Gjennom eksperimentering med ulike instruksjoner for å finne den mest effektive måten å be språkmodellen om å forenkle tekstene, ble flere formuleringer testet. Basert på resultatene fra disse testene ble det deretter gjort en manuell vurdering av hvilken instruks som ga de beste tekstene i forhold til idealkteksten. En av utfordringene var at noen av promptene ga for korte tekster som manglet vesentlig informasjon. Eller at informasjonen ble feil gjengitt eller inneholdt merkelige ord som for eksempel «neseblodtilfeller». Det virket som at forsøket på å forenkle teksten gikk ut over formidlingen av innholdet og kvaliteten på språket.

Det er også viktig å være oppmerksom på at det er begrensninger for hvor mange instruksjoner man kan teste. De forenklede tekstene er et direkte resultat av både språkmodellens evne til å beherske norsk språk og kvaliteten på instruksjonene den mottar. Instruksjoner som er på et lavt nivå, kan føre til mindre presise resultater. Når man evaluerer effekten av prompts som brukes til automatisk tekstforenkling, er det derfor viktig å være oppmerksom på potensielle feilkilder. En av disse kan være dårlig formulerte spørsmål eller instruksjoner til språkmodellen. Hvis prompts ikke er klare og tydelige, kan det resultere i unøyaktigheter i den forenklede teksten. Basert på resultatene fra testingen av forskjellige formuleringer av prompts, ser det ut til at valget av prompts har en betydning. Det er mulig at vi kunne ha fått en høyere kvalitet på resultatene av tekstforenklingen dersom vi hadde stilt andre og «bedre» prompts. Da hadde det kanskje vært mindre behov for manuelle justeringer.

7.1.4 Gjennomgående utfordringer med forenklede tekster

En gjennomgående observasjon i de genererte tekstene er tilstedeværelsen av generelle grammatiske feil. Til tross for utprøving av ulike prompts og instruksjoner, har det blitt observert at disse feilene oppstår uavhengig av formuleringen av promptene. Dette indikerer en underliggende utfordring med språkmodellens evne til å produsere grammatisk korrekte setninger gjennom omformulering av eksisterende tekst. For å forbedre lesbarheten og kvaliteten på de genererte tekstene, er det derfor nødvendig med en manuell gjennomgang og redigering.

Utfordringene knyttet til å finne prompts som produserer ønsket resultat kan tyde på at dagens språkmodeller har visse begrensninger. En av disse er den maksimale tokensgrensen, som begrenser promptlengden, artikkellengden og den forenklede teksten. Det er viktig å merke

seg at denne begrensning kan bli mindre relevant etter hvert som teknologien utvikler seg, på samme måte som kostnaden for datamaskinlagring har blitt billigere over tid. Vi kan allerede observere at nyere GPT-modeller har større tokenkapasitet. OpenAI har delt informasjon om GPT-4, som for øyeblikket kun er tilgjengelig for betalende kunder, og har en kapasitet på 8000 tokens. En utvidet modell vil også kunne håndtere opp til 32 000 tokens (OpenAI, u.å.-d). Videre ville det være en fordel om retningslinjer for klarspråk kunne inkorporeres direkte i språkmodellen, uten å bruke opp tokens.

Språk er en kompleks struktur som omfatter flere regler og varierende setningsstrukturer. Språkmodeller kan støte på utfordringer med å opprettholde en konsistent grammatikkstruktur og velge riktige ord for konteksten. I tillegg til disse utfordringene knyttet til språkmodellens generelle evner, oppstår det ytterligere utfordringer ved omformulering av norske tekster. Norsk er et språk med begrenset utvalg av spesialiserte modeller sammenlignet med større språk, noe som kan påvirke kvaliteten på de genererte tekstene (se delkapittel 2.7.1). For å kunne generere tekster med høy språkkvalitet, kreves det trening av modellen på store mengder norsk tekstdata.

Det har vært flere nyhetsartikler som diskuterer bruken av kunstig intelligens for å besvare eksamensoppgaver. I en artikkel fra Dagsavisen beskrives det hvordan instituttlederen benyttet ChatGPT, som er basert på GPT-3, til å skrive en eksamensoppgave og oppnådde karakteren A (Bjerke, 2023). Dette indikerer at GPT-3 er i stand til å produsere tekster på norsk med både god språkkvalitet og innhold. Men, på tross av at bruk av ChatGPT kan gi resultatet A på eksamen, har vi i arbeidet med prototypen sett at tekstene som produseres har mange logiske brister, skrivefeil og grammatiske feil. Etter samtaler med Klar Tale har det blitt understreket at omskriving av eksisterende artikler er mer utfordrende enn å produsere en tekst fra bunnen av (se delkapittel 5.3.2). Dette understreker kompleksiteten av å omformulere eksisterende tekster. Selv de som har mye erfaring med det, synes det er tidkrevende å manuelt gjøre det arbeidet vi har forsøkt å gjøre via GPT-3. Det er derfor viktig å ta hensyn til de spesifikke utfordringene knyttet til omformulering av eksisterende tekster og sikre en grundig evaluering av resultatene. I tillegg til regler for å skrive tekster i klarspråk, er det kanskje nødvendig å ha spesifikke regler for hvordan omformulere tekster til klarspråk.

En annen utfordring er problemet med gjengivelse av sitater, som sett i delkapittel 6.2.1. Det er imidlertid mulig å løse problemene med sitatgjengivelse i de forenklede tekstene ved å for eksempel «hardkode» regler for gjengivelse av sitater slik at de følger pressens etiske regler. Når det gjelder forskjeller mellom forenklingsnivåene kan det se ut som at det er mindre utfordringer i nivå tre enn i nivå én og to. Nivå tre unngår også i større grad problemer med usammenhengende setninger i forhold til de to andre nivåene.

Gjennom sammenligningen av ti artikler kom det frem at halvparten av tekstene generert for nivå to, er lenger enn nivå én. Dette tilsier at prompten eller antall tokens for dette nivået ikke er godt nok. To av eksempelartiklene som ble brukt, var under 210 ord og forskjellen mellom forenklingsnivåene var derfor ikke så stor. Dette kan indikere at det kanskje ikke er behov for flere nivåer av forenkling for korte artikler. Det er likevel behov videre testing av kortere artikler.

7.1.5 Lesbarhetsindeks

For å evaluere kvaliteten på tekstforenklingen ble det gjennomført testing av liksskår for de ulike nivåene i den siste iterasjonen av prototypen. Liksskår, også kjent som lesbarhetsindeks, er en metode som gir en indikasjon på hvor lettlest en tekst er (se delkapittel 3.3.2). Både den originale GPT-3-tekstversjonen og den manuelt redigerte teksten ble testet. Vi brukte likskalkulatoren fra skriftlig.no for å beregne liksskårene. Resultatene er presentert i Tabell 5: *Oversikt over liksskår for NRK-artikkel og alle forenklingsnivåene fra siste iterasjon av prototypen.*

og gir en indikasjon på lesbarhetsnivået for de ulike forenklede versjonene i forhold til den opprinnelige NRK-artikkelen. Skriftlig.no bruker følgende skala for lesbarhet:

- «< 30: Veldig lettlest, som barnebøker
- 30–40: Lettlest, som skjønnlitteratur eller ukeblader
- 40–50: Middels vanskelig, som vanlig avistekst
- 50–60: Vanskelig, vanlig verdi for offisielle tekster
- 60: Veldig tunglest byråkratspråk» (Skriftlig.no, u.å.).

NRK-artikkel	Nivå 1	Nivå 1 – GPT-3	Nivå 2	Nivå 2 – GPT-3	Nivå 3	Nivå 3 – GPT-3
35	35	40	28	35	20	25

Tabell 5: Oversikt over liksskår for NRK-artikkel og alle forenklingsnivåene fra siste iterasjon av prototypen.

Tabellen viser at teksten generert av GPT-3 blir gradvis mer forenklet. Det første forenklingsnivået får imidlertid en høyere skår enn den originale teksten. Det er også stor forskjell mellom de genererte tekstene og den manuelt redigerte teksten. Nivå én og tre har fem poeng mindre og nivå to har syv poeng mindre enn GPT-3-versjonen. Noe som igjen viser at GPT-3 ikke forenkler tekstene tilstrekkelig. Dette er likevel bare én artikkel og som sett gjennom sammenligningen av andre artikler, er det store forskjeller på de genererte tekstene. Det ble derfor også undersøkt liksskår for NTB-artikkelen og de andre versjonene se Tabell 6.

NTB-artikkel	Klar Tale	Nivå 1	Nivå 2	Nivå 3
37	25	33	34	34

Tabell 6: Oversikt over liksskår for NTB-artikkel, Klar Tale og GPT-3

Den andre tabellen viser ganske andre tall enn det første eksempelet. Her har alle forenklingsnivåene en lavere skår enn den originale artikkelen, men det er ikke en gradvis progresjon slik som den andre artikkelen. Skåren til nivå to og tre er den samme, og ett hakk over nivå en. Klar Tales versjon av teksten har desidert lavest skår og kommer best ut av testen. Nivå tre er kortere enn Klar Tales versjon, men får likevel en mye høyere skår. Selv det ble brukt samme prompt for begge artikler blir resultatet vidt forskjellig. Liksskår er kun ment å gi en indikasjon basert på beregninger, og vil ikke være en like god metode for evaluering som en manuell vurdering på bakgrunn av klarspråkregler. Siden liksskår er noe som ble nevnt av alle journalistene vi snakket med, kan det tyde på at de vurderer det som et nyttig hjelpemiddel.

7.1.6 Retningslinjer for klarspråk: den ideelle teksten vs. den genererte teksten

For å evaluere tekstene som ble generert av GPT-3, er det benyttet retningslinjer for klarspråk som et viktig verktøy. Disse retningslinjene har til hensikt å sikre at teksten er enkel å forstå og lese for målgruppen, og dermed oppfyller kravene til klart språk.

I vurderingen av retningslinjenes effektivitet for automatisk tekstforenkling er det viktig å vurdere om de treffer brukerne og oppnår sitt formål. Vi har erfart at retningslinjene kan være nyttige som veiledning i en manuell tekstproduksjon da de gir en referanse og et rammeverk for å sikre enkelhet og forståelighet. Ved å følge retningslinjene kan man produsere tekster som er tilpasset målgruppen og oppfyller kravene til klarspråk. Det er imidlertid viktig å merke seg at effekten av retningslinjene kan variere avhengig av konteksten og målgruppen. Enkelte retningslinjer kan være mindre relevante eller ikke alltid nødvendigvis treffe brukerne på riktig måte. Det kan være situasjoner der bruk av passive setninger faktisk kan være hensiktsmessig eller der andre språklige valg bedre tilpasser seg målgruppens behov. Retningslinjene for klarspråk er ikke nødvendigvis en absolutt mal, men heller et verktøy som kan brukes som en veiledning. De gir en grundig forståelse av hva som kan bidra til å forenkle tekster og gjøre dem mer leservennlige, men det krever også en vurdering av konteksten og behovene til målgruppen. Ved automatisk tekstforenkling er det derfor viktig å balansere bruken av retningslinjene med en kritisk og kontekstuell tilnærming.

Den genererte teksten viste seg å ikke være tilstrekkelig forenklet, de fulgte ikke retningslinjer for klarspråk. Fordi tekstene skulle brukes i prototypen og testes med personer i målgruppen, vurderte vi det som nødvendig å gjøre manuelle endringer. Det er viktig å merke seg at «idealk tekstene» i enkelte tilfeller bryter med retningslinjene. Et eksempel på dette er at en setning ble omgjort fra aktiv til passiv. Dette var ikke tilsiktet og viser at selv når man forsøker å følge retningslinjene aktivt, kan det skje feil. På den andre siden er det ikke alltid at passive konstruksjoner gjør teksten vanskelig å lese. Klarspråkreglene er nettopp retningslinjer og ikke rigide regler som må følges i alle tilfeller.

7.2 Brukersentrert design

7.2.1 Datainnsamling

I arbeidet med oppgaven samlet vi inn data om dysleksi, automatisk tekstforenkling og kunstig intelligens. Det ble også undersøkt hvilke produkter som allerede eksisterer og hvordan hva som er tilgjengelig på norsk. Dette ble gjort i starten av prosjektet og la grunnlag for oppgaven og hvilken retning prototypen tok.

For å samle kvalitative data ble det gjennomført semi-strukturerte intervjuer med personer med dysleksi, journalister med erfaring fra klarspråk og kunstig intelligens. Gjennom intervjuene ble det samlet mye nyttig informasjon. Ved å velge denne type intervjuer er det rom for å stille oppfølgingsspørsmål og ha en mer naturlig samtale. Dette gjorde at det ble nevnt informasjon som ikke nødvendigvis hadde kommet frem dersom det hadde vært et mer strukturert intervju. Under intervjuene ble flere av de forhåndsbestemte spørsmålene besvart i løpet av samtalen.

7.2.2 Iterativ utvikling

Dette prosjektet har fulgt en iterativ produksjonsprosess. Det vil si at vi kontinuerlig har fått tilbakemelding gjennom utviklingsprosessen. Det har gjort at vi har kunne evaluere og justere funksjoner og tilpasse prototypen til målgruppen. Med tanke på tekstforenkling hadde dette en stor innvirkning da vi fikk tilbakemelding fra brukerne underveis om at nivåene av forenkling var for lite differensierte. Det var behov for videre forenkling. Disse tilbakemeldingene var avgjørende for det videre arbeidet med prototypen.

7.2.3 Evaluering

Evalueringen av prototypen innebar flere viktige aspekter. Først og fremst gjennomførte vi flere iterasjoner med brukertester og justeringer basert på tilbakemeldinger fra brukere med dysleksi. Dette ga oss verdifulle innsikter i hvordan prototypen fungerte i praksis og hvilke endringer som var nødvendige for å forbedre lesbarheten og tilpasningen av nyhetsartikler. I tillegg til brukertester ble det også gjennomført en siste evaluering der vi involverte eksperter innen feltet. Disse ekspertene ga oss verdifulle tilbakemeldinger og vurderinger av prototypens evne til å forenkling tekstene i tråd med klarspråksprinsippene. Deres ekspertise og

innsikt bidro til en dypere forståelse av prototypens styrker og begrensninger, og hjalp oss med å identifisere områder som krever ytterligere forbedringer. En viktig observasjon under evalueringen var betydningen av promptene som ble brukt til å instruere språkmodellen. Enkelte instruksjoner resulterte i svært korte oppsummeringer med noen vanskelige ord fortsatt til stede, mens andre instruksjoner ikke klarte å fjerne spesifikke termer som «bedriftshelsetjeneste» og «oberstløytnant». Denne observasjonen understreker viktigheten av å utvikle klare og presise instruksjoner for språkmodellen, slik at den kan generere mer nøyaktige og målrettede forenklinger.

Samlet sett har evalueringen av prototypen gitt oss verdifulle innsikter og retninger for videreutvikling. Resultatene av brukertester og tilbakemeldinger fra eksperter har bidratt til en forbedret forståelse av prototypens styrker og svakheter. Disse funnene vil være avgjørende for å kunne videreutvikle og forbedre prototypen med sikte på å skape et verktøy som kan bidra til bedre tilgjengelighet og forståelighet av nyhetsartikler for personer med dysleksi.

7.3 Hvordan kan generative språkmodeller tilpasses for å forenkle nyhetsartikler for personer med dysleksi?

For å svare på forskningsspørsmålet «Hvordan kan generative språkmodeller tilpasses for å forenkle nyhetsartikler for personer med dysleksi?» er det forsøkt å utvikle en prototype som forenkler tekster i ulike nivåer. Flere tilnærminger til tekstforenkling er testet gjennom å instruere GPT-3 modellen med ulike prompts. I tillegg har promptsene blitt testet på et større utvalg artikler enn den som ble brukt i prototypen. Etter å ha evaluert resultatene av de ulike promptsene endte vi opp med tre distinkte formuleringer, se Tabell 8.

«Rewrite the following text in an easy-to-understand language for people with dyslexia»

«Rewrite the following text in an easy-to-understand language for people with dyslexia and remove long, difficult words»

«Rewrite the following text in a short, easy-to-understand language for people with dyslexia and remove long, difficult words»

Tabell 7: Oversikt over prompts brukt for tekstforenkling for forenklingsnivå 1 - 3.

Disse promptene ble valgt basert på en evaluering av ulike resultater fra flere prompter. På bakgrunn av diverse kriterier for klarspråk, forenklet disse promptene teksten mer

tilfredsstillende enn de andre. Her handlet det om å finne balansen mellom det som var enkelt nok, men der tilstrekkelig informasjon ble bevart, i tillegg til at teksten ikke trengte for mange manuelle justeringer. Basert på en samlet vurdering av de forenklete versjonene av artiklene som er diskutert i denne oppgaven, indikerer det at språkmodellen som ble brukt ikke er god nok slik den fungerer i dag til dette formålet. Tekstene ble ikke tilstrekkelig forenklet for målgruppen. Selv om språkmodellen i dag kan brukes til å generere gode tekster fra bunnen av, er teknologien ikke tilpasset formålet for denne oppgaven, å forenkle nyhetsartikler tilpasset personer med dysleksi. For at det skal være mulig å benytte denne teknologien til et slikt mål, er det nødvendig å gjøre videre arbeid.

Som nevnt i delkapittel 7.1.3, kan formuleringen av promptene som er brukt i prototypen være mangelfulle eller ikke presise nok. I videre arbeid med denne typen verktøy vil det være viktig å utforme spesifikke instruksjoner og prompts. Disse bør instruere språkmodellen til å generere enklere tekster ved å bruke enkle ord, unngå kompleks terminologi og redusere setningslengden. Videre kan språkmodellens trening tilpasses ved å inkludere tekster som er spesielt utformet for personer med dysleksi, slik at den lærer seg å generere mer tilpassede og lettfattelige tekster for denne målgruppen. Kontekstuelle instruksjoner kan også benyttes for å informere språkmodellen om behovene til personer med dysleksi og utfordringer de møter ved lesing og forståelse av komplekse tekster. En iterativ tilnærming med testing, evaluering og justering av språkmodellen er viktig for å sikre at den produserer tilfredsstillende forenklete tekster. I tillegg kan en kombinasjon av automatisert forenkling og menneskelig redigering være nødvendig for å sikre høy kvalitet og nøyaktighet i tekstene. Gjennom disse tilpasningene kan generative språkmodeller bidra til å skape mer tilgjengelige og leservennlige nyhetsartikler for personer med dysleksi, og dermed forbedre deres leseopplevelse.

7.4 Hvordan kan vi utnytte graderte forenklingsnivå for tilpassing av nyhetsartikler for personer med dysleksi?

For å utnytte graderte forenklingsnivå for tilpassing av nyhetsartikler for personer med dysleksi, er det viktig å ta hensyn til brukernes preferanser og behov. I 5.5.2 så vi at tilbakemeldingene fra brukertester viste at det er en variasjon i hvordan personer med dysleksi oppfatter og foretrekker de ulike forenklingsnivåene. Noen brukere foretrakk å lese den originale versjonen av teksten for å unngå å miste viktig informasjon, mens andre oppfattet

essensen av teksten på det tredje forenklingsnivået og likte den korte og enkle setningsstrukturen. Dette indikerer at tilpasning av nyhetsartikler for personer med dysleksi bør inkludere flere nivåer av forenkling for å imøtekomme forskjellige preferanser og behov.

En mulig tilnærming er å tilby en svært kort punktvis versjon på det laveste forenklingsnivået, som gir en essensiell oversikt, etterfulgt av gradvis mer detaljerte nivåer for de som ønsker mer informasjon. Dette kan gjøre det lettere for personer med dysleksi å få ut essensen av teksten samtidig som de har mulighet til å lese mer detaljert hvis de ønsker det.

Videre kan gradert forenkling av nyhetsartikler være spesielt nyttig i situasjoner der brukere trenger å håndtere omfattende og komplekse tekster. Dette kan omfatte forskningsartikler, der forenklingen kan hjelpe journalister med å forstå og formidle forskernes budskap på en mer tilgjengelig måte. Det kan også gjelde byråkratiske tekster i offentlige etater, der forenklingen kan bidra til at mottakerne lettere forstår budskapet. Gjennom brukertestene kom det også frem at noen brukere kanskje ikke ville brukt tekstforenklingen til vanlig, men heller i skoleoppgaver eller lignende. De så verdien av forenklingen når det gjaldt lange tekster med mye informasjon. Dette indikerer at tekstforenklingen kan være spesielt nyttig i situasjoner der brukere trenger å håndtere omfattende og komplekse tekster.

For å vurdere om den graderte forenklingen er «bra nok» bør en kombinasjon av faktorer tas i betraktning. Dette inkluderer tilbakemeldinger fra brukere med dysleksi for å evaluere deres opplevelse og behov, vurdering av formidling og lesbarhet av den forenklete teksten, samarbeid med eksperter på dysleksi og klarspråk for validasjon, samt testing og iterasjon for å forbedre forenklingen over tid. Ved å kontinuerlig evaluere og tilpasse tilnærmingen basert på brukernes tilbakemeldinger, kan man sikre at forenklingen er optimal og tilfredsstillende for behovene til personer med dysleksi når det gjelder tilpassede nyhetsartikler.

Videre kan gradert forenkling av nyhetsartikler ha utilsiktede fordeler som å spare tid ved at teksten blir kortere og mer konsis. Dette er spesielt verdifullt i en tid der informasjonsmengden er overveldende. Ikke bare for personer med dysleksi, men også for andre lesere som ønsker å raskt få essensen av en nyhetsartikkel. Denne tilnærmingen kan bidra til å bryte den onde sirkelen der personer med dysleksi leser mindre på grunn av langsommere lesing. Ved å gi tilgang til mer tilgjengelig og forenklet informasjon kan det hjelpe dem med å subjektivt oppleve lesing som lettere. Under samtalene med Klar Tale ble

det tydelig at forenkling av tekster kan ha en betydelig innvirkning, og et konkret eksempel er da forenklingen av partiprogrammet resulterte i at noen stemte ved valget. Dette demonstrerer at forenkling av tekster kan utgjøre en reell forskjell.

8 AVSLUTNING

Dette kapittelet oppsummerer hovedfunnene i oppgaven. Deretter nevnes flere begrensinger ved utviklingen og tilnærmingen til automatisk tekstforenkling som er gjort i dette prosjektet. Til slutt diskuteres muligheter for videre utvikling av verktøyet.

8.1 Resultater

I denne masteroppgaven er det utforsket og evaluert automatisert tekstforenkling som et verktøy for å forbedre tilgjengeligheten av nyhetsartikler for personer med dysleksi. Det er undersøkt ulike tilnærminger og metoder, og gjennomført brukertester for å evaluere effektiviteten og brukervennligheten av en prototype basert på GPT-3-språkmodellen. Funnene indikerer at gradert forenkling av nyhetsartikler kan være en verdifull tilpasning for personer med dysleksi. Ved å tilby flere nivåer av forenkling, kan brukerne velge det nivået som passer best for deres leseferdigheter og behov. Videre har vi identifisert flere utfordringer knyttet til språkmodellens genererte tekster, inkludert grammatikalske feil, unødvendig repetisjon og manglende forståelse av konteksten. Disse utfordringene kan begrense nøyaktigheten og kvaliteten av tekstforenklingen og peker på behovet for videre forbedringer og finjusteringer av språkmodeller og prompter.

Det er viktig å merke seg at automatisert tekstforenkling ikke erstatter behovet for redaksjonell innsats og menneskelig vurdering. En kombinasjon av maskinell forenkling og manuell redigering vil kunne gi de beste resultatene i å skape tilgjengelige og leservennlige nyhetsartikler for personer med dysleksi. Videre forskning kan utforske mer spesifikke tilnærminger for å forbedre tekstforenkling, inkludert bruk av kontekstsensitive modeller og mer avanserte instruksjoner og prompter. Det er også behov for å undersøke hvordan tekstforenkling kan tilpasses til ulike typer tekster og sjangere, og hvordan det kan integreres i eksisterende nyhetsredaksjoner og plattformer.

I sum representerer denne masteroppgaven et bidrag til feltet med tilgjengelig kommunikasjon og teksttilpasning for personer med dysleksi. Ved å kombinere teknologiske

fremskritt med brukerfokuserte tilnæringer, kan vi arbeide mot å skape et mer inkluderende og tilgjengelig informasjonssamfunn for alle lesere, uavhengig av deres leseferdigheter.

8.2 Studiens begrensninger og videre utvikling

Det er flere begrensninger ved denne oppgaven som er viktig å nevne og som bør tas i betraktning ved tolkning av resultatene. En begrensning er at vi brukte samme nyhetsartikkel i alle iterasjonene av prototypen. Det ville vært hensiktsmessig å teste prototypen med et bredere utvalg tekster for å evaluere lesbarheten sammen med deltakere i brukertester. Det ble utført to runder med brukertester, og fire av deltakerne hadde sett artikkelen en gang før. Dette kan ha innvirkning på oppfattelsen på den ytterligere forenklingen av nivåene. Videre kan det være hensiktsmessig å involvere flere eksperter på klarspråk og foreta en grundigere analyse av prototypens evne til å følge klarspråksretningslinjene og produsere tekster med høy lesbarhet.

En begrensning knyttet til forskningsmetoden er at vi benyttet oss av en kvalitativ tilnærming. Selv om denne tilnærmingen ga oss innsikt i brukernes opplevelser og perspektiver, har den også sine begrensninger når det gjelder generaliserbarhet og kvantifisering av resultatene. Videre kan det være diskusjon rundt påliteligheten av intervjuanalysen. Selv om vi har fulgt anerkjente metoder for analyse, er det alltid en viss grad av tolkning og subjektivitet som kan påvirke resultatene.

En av utfordringene i dette prosjektet var å finne informanter som ville dele sine erfaringer med dysleksi. Vi brukte forskjellige strategier for å rekruttere deltakere, blant annet gjennom personlige kontakter, plakater i Bergen, og sosiale medier. Dysleksi Bergen og Dysleksi Ungdom var spesielt hjelpsomme med å spre informasjon om prosjektet og formidle kontakt med potensielle informanter. Vi er klar over at utvalget vårt kan være skjevt, siden det var lettere å få tak i studenter enn andre grupper. I tillegg hadde de fleste av deltakerne en mild grad av dysleksi. Det kan påvirke resultatene våre, og vi kan ikke generalisere funnene våre til alle med dysleksi. Vi har likevel fått noen innblikk i hvordan dysleksi kan variere mellom personer, både gjennom deltakernes egne opplevelser og gjennom deres familiære bakgrunn.

Det bør også nevnes at prototypen i denne oppgaven var en enkel teknologisk implementering. Videre arbeid kan involvere utvikling av en mer fullstendig prototype som

utfører tekstforenklingen direkte, både med tanke på funksjonaliteten og brukergrensesnittet. Det vil også gjøre det mulig å implementere flere funksjoner i prototypen som kan testes videre. Til slutt er det viktig å være oppmerksom på at begrensningene i denne oppgaven er spesifikke for norsk og nyhetsartikler. Det kan være andre begrensninger og utfordringer knyttet til andre språk og tekstsjangre.

Videre utvikling

Videre utvikling av tekstforenklingsteknologien har et stort potensial til å forbedre tilgjengeligheten til nyhetsartikler. For å ta teknologien til nye høyder, er det flere områder som kan utforskes. Først og fremst er det viktig å fortsette å forbedre og optimalisere språkmodellene som brukes til tekstforenkling. Dette kan omfatte trening av modellene på mer spesifikke tekstdata, slik at de bedre kan tilpasses ulike typer tekster. Videre kan utforskning av metoder for å integrere kontekstuell informasjon i språkmodellene være en vei å gå. Dette kan bidra til å produsere mer presise og sammenhengende forenklete tekster. I tillegg bør det forskes på å utvikle mer avanserte og effektive algoritmer og metoder for automatisk evaluering av tekstforenkling. Dette kan hjelpe til med å kvantifisere kvaliteten på de genererte tekstene og gi verdifull innsikt i eventuelle forbedringsområder. Gjennom kontinuerlig forskning og innovasjon kan tekstforenklingsteknologien fortsette å utvikle seg og tilpasse seg behovene til målgruppen.

9 REFERANSER

- Abid, A., Farooqi, M., & Zou, J. (2021). Persistent Anti-Muslim Bias in Large Language Models. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 298–306. <https://doi.org/10.1145/3461702.3462624>
- Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for text-based emotion detection: A review of BERT-based approaches. *The Artificial Intelligence Review*, 54(8), 5789–5829. <https://doi.org/10.1007/s10462-021-09958-2>
- Al-Thanyyan, S. S., & Azmi, A. M. (2022). Automated Text Simplification: A Survey. *ACM Computing Surveys*, 54(2), 1–36. <https://doi.org/10.1145/3442695>
- Amazon. (u.å.). *What is Word Wise? - Amazon Customer Service*. Hentet 23. mai 2023, fra <https://www.amazon.com/gp/help/customer/display.html?nodeId=201645250>
- Anderson, D. (2021, januar 13). *When AI writes poetry*. Humanise.AI. <https://www.humanise.ai/blog/ai-writes-poetry/>
- Anderson, J. (1983). Lix and Rix: Variations on a Little-known Readability Index. *Journal of Reading*, 26(6), 490–496.
- Benassi, J. L. G., Amaral, D. C., & Ferreira Junior, L. D. (2011). Product vision management: Concept and models evaluation. *Product Management & Development*, 9(2), 163–172. <https://doi.org/10.4322/pmd.2012.008>
- Berget, G., Herstad, J., & Sandnes, F. E. (2016). *Search, Read and Write: An Inquiry into*. <http://dx.doi.org/10.3233/978-1-61499-684-2-450>
- Bjerke, Y. L. (2023, mars 29). *Instituttlederen gikk undercover og brukte ChatGPT på eksamen – fikk A*. Dagsavisen. <https://www.dagsavisen.no/demokraten/navn-i-nyhetene/2023/03/29/instituttlederen-gikk-undercover-og-brukte-chatgpt-pa-eksamen-fikk-a/>
- Bordia, S., & Bowman, S. R. (2019). *Identifying and Reducing Gender Bias in Word-Level Language Models* (arXiv:1904.03035). arXiv. <http://arxiv.org/abs/1904.03035>
- Britannica. (u.å.). *History of publishing—Newspapers, Newsletters, Censorship, and Freedom of Press | Britannica*. Hentet 31. mai 2023, fra <https://www.britannica.com/topic/publishing/The-first-newspapers>
- British Dyslexia Association. (u.å.-a). *About the British Dyslexia Association*. British Dyslexia Association. Hentet 22. april 2023, fra <https://www.bdadyslexia.org.uk/about/about-the-british-dyslexia-association>
- British Dyslexia Association. (u.å.-b). *Dyslexia Style Guide 2023*.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A.,

Shyam, P., Sastry, G., Askill, 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* (arXiv:2005.14165). arXiv. <http://arxiv.org/abs/2005.14165>

Carlini, N., Tramèr, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., Erlingsson, Ú., Oprea, A., & Raffel, C. (2021). *Extracting Training Data from Large Language Models*.

Chowdhary, K. R. (2020). Natural Language Processing. I K. R. Chowdhary (Red.), *Fundamentals of Artificial Intelligence* (s. 603–649). Springer India. https://doi.org/10.1007/978-81-322-3972-7_19

Dastin, J. (2018, oktober 10). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>

Denny, E., & Weckesser, A. (2022). How to do qualitative research? *BJOG: An International Journal of Obstetrics & Gynaecology*, 129(7), 1166–1167. <https://doi.org/10.1111/1471-0528.17150>

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. <http://arxiv.org/abs/1810.04805>

Dysleksi Norge. (2017). *Faglige retningslinjer for kartlegging, utredning og oppfølging av elever med dysleksi*. <https://dysleksinorge.no/wp-content/uploads/2017/08/faglige-retningslinjer-versjon-23.pdf>

Ebling, S., Battisti, A., Kostrzewa, M., Pfützte, D., Rios, A., Säuberli, A., & Spring, N. (2022). Automatic Text Simplification for German. *Frontiers in Communication*, 7. <https://www.frontiersin.org/articles/10.3389/fcomm.2022.706718>

Forstner, S. L. (Under utvikling). *Designing a Dyslexia-Friendly Interaction With News Articles*. University of Bergen.

Frantz, C. (2020, september 23). I wrote a book with AI. *Medium*. <https://medium.com/@FrantzLight/i-wrote-a-book-with-ai-200abbccc533>

Funnemark, N. S., & Aagesen, G. (2019). *Lånekassens innspill til nasjonal strategi for kunstig intelligens*. <https://www.regjeringen.no/contentassets/0e36c85fcfe143a5b626c53cf292cb3b/lanekassens-innspill-til-nasjonal-strategi-for-kunstig-intelligens.pdf>

Grammarly. (u.å.). *About Us | Grammarly*. Hentet 1. juni 2023, fra <https://www.grammarly.com/about>

Gregor, P., Dickinson, A., Macaffer, A., & Andreasen, P. (2003). SeeWord—A personal word processing environment for dyslexic computer users. *British Journal of Educational Technology*, 34(3), 341–355. <https://doi.org/10.1111/1467-8535.00331>

Helland, T. (2021). Dysleksi. I *Store norske leksikon*. <http://snl.no/dysleksi>

Hemingway Editor. (u.å.). *Hemingway Help*. Hentet 1. juni 2023, fra <https://hemingwayapp.com/help.html>

Hero. (u.å.). *Hero · Reading tool—Faq*. Hero · Reading tool. Hentet 1. juni 2023, fra <http://beta.heroapp.ai/p17298226/>

Johansen, N. (2022, januar 28). *Stadier i lese- og skriveutviklingen*. <https://www.kvafjord.kommune.no/stadier-i lese-og-skriveutviklingen.535626.no.html>

Kalyan, K. S., Rajasekharan, A., & Sangeetha, S. (2021). *AMMUS: A Survey of Transformer-based Pretrained Models in Natural Language Processing* (arXiv:2108.05542). arXiv. <http://arxiv.org/abs/2108.05542>

Kommunal- og moderniseringsdepartementet. (2020, januar 14). *Nasjonal strategi for kunstig intelligens* [Plan]. Regjeringen.no; regjeringen.no. <https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/?ch=3>

Kutuzov, A., Barnes, J., Velldal, E., Øvrelid, L., & Oepen, S. (2021). *Large-Scale Contextualised Language Modelling for Norwegian* (arXiv:2104.06546). arXiv. <http://arxiv.org/abs/2104.06546>

Kvadsheim, S. S. (2022, november 8). *Ny undersøkelse: - Norske bedrifter mangler kompetanse på kunstig intelligens* | *Finansavisen*. https://www.finansavisen.no/nyheter/arbeidsliv/2022/11/08/7958038/ny-undersokelse-norske-bedrifter-mangler-kompetanse-pa-kunstig-intelligens?internal_source=sistenytt&zephr_sso_ott=vTbbs0

Lancaster, A. (2023, mars 20). Beyond Chatbots: The Rise Of Large Language Models. *Forbes*. <https://www.forbes.com/sites/forbestechcouncil/2023/03/20/beyond-chatbots-the-rise-of-large-language-models/>

Lindland, L. A. (2023, mars 12). Kunstig intelligens og ChatGPT tvinger utdanningssektoren til å tenke nytt, men ingen nye regler nå. *NRK*. https://www.nrk.no/rogaland/kunstig-intelligens-og-chatgpt-tvinger-utdanningssektoren-til-a-tenke-nytt_-men-ingen-nye-regler-na-1.16298342

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of

the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, 74(5), 570–581. <https://doi.org/10.1002/asi.24750>

Lånekassen. (2016). *Årsrapport 2015*. <https://lanekassen.no/siteassets/om-organisasjonen/arsrapporter/arsrapport-2015-versjon-1.0.pdf>

Microsoft. (u.å.). *Use Immersive Reader in Microsoft Edge*. Hentet 23. april 2023, fra <https://support.microsoft.com/en-us/topic/use-immersive-reader-in-microsoft-edge-78a7a17d-52e1-47ee-b0ac-eff8539015e1>

moderniseringsdepartementet, K. (2017, november 2). *Klarspråksjuryen* [Styreradutvalg]. Regjeringen.no; regjeringen.no. <https://www.regjeringen.no/no/dep/kdd/org/styrer-rad-og-utvalg/klarspraksjuryen/id2577471/>

National Center on Improving Literacy. (2018, januar 24). *Phonemic Awareness*. National Center on Improving Literacy. <https://improvingliteracy.org/glossary/phonemic-awareness>

National Center on Improving Literacy. (2022, juni 21). *Phonological Awareness: What is it and how does it relate to phonemic awareness?* National Center on Improving Literacy. <https://improvingliteracy.org/brief/phonological-awareness-what-it-and-how-does-it-relate-phonemic-awareness>

Nicolson, R. I., & Fawcett, A. J. (2011). Dyslexia, dysgraphia, procedural learning and the cerebellum. *Cortex*, 47(1), 117–127. <https://doi.org/10.1016/j.cortex.2009.08.016>

NTB. (u.å.). *Nynorskroboten*. NTB. Hentet 1. juni 2023, fra <https://www.ntb.no/kommunikasjon/nynorskroboten>

OpenAI. (u.å.-a). *Embeddings*. Hentet 1. mai 2023, fra <https://platform.openai.com>

OpenAI. (u.å.-b). *Models—OpenAI API*. Hentet 1. mai 2023, fra <https://platform.openai.com/docs/models/gpt-3>

OpenAI. (u.å.-c). *What are tokens and how to count them? | OpenAI Help Center*. Hentet 1. mai 2023, fra <https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them>

OpenAI. (u.å.-d). *What is the difference between the GPT-4 models? | OpenAI Help Center*. Hentet 31. mai 2023, fra <https://help.openai.com/en/articles/7127966-what-is-the-difference-between-the-gpt-4-models>

Pan, X., Zhang, M., Ji, S., & Yang, M. (2020). Privacy Risks of General-Purpose Language Models. *2020 IEEE Symposium on Security and Privacy (SP)*, 1314–1331. <https://doi.org/10.1109/SP40000.2020.00095>

Pressens Faglige Utvalg. (2021, januar 1). *Vær Varsom-plakaten*. Presse.no. <https://presse.no/pfu/etiske-regler/vaer-varsom-plakaten/>

Rello, L., Baeza-Yates, R., Dempere-Marco, L., & Saggion, H. (2013). Frequent Words Improve Readability and Short Words Improve Understandability for People with Dyslexia. I P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, & M. Winckler (Red.), *Human-Computer Interaction – INTERACT 2013* (Bd. 8120, s. 203–219). Springer Berlin Heidelberg.
https://doi.org/10.1007/978-3-642-40498-6_15

Rello, L., & Bigham, J. P. (2017). Good Background Colors for Readers: A Study of People with and without Dyslexia. *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 72–80. <https://doi.org/10.1145/3132525.3132546>

Rewordify. (u.å.). *Rewordify.com | Understand what you read*. Hentet 23. mai 2023, fra <https://rewordify.com/>

Sharp, H., Rogers, Y., & Preece, J. (2019). *Interaction design 5e*. John Wiley and Sons.

SIL International. (2003). *Phoneme*. Glossary of Linguistic Terms.
<https://glossary.sil.org/term/phoneme>

Skipperø, P. J. & NTB nyheter. (2023, april 27). *Flere hundre tusen nordmenn har ennå ikke åpnet skattemeldingen*. Klar Tale. <https://www.klartale.no/norge/2023/04/27/flere-hundre-tusen-nordmenn-har-enna-ikke-apnet-skattemeldingen/>

Skriftlig.no. (u.å.). *Liks*. Skriftlig.no. Hentet 31. mai 2023, fra <https://skriftlig.no/lik/>

Språkrådet. (2013). *Klarspråk*.
<https://www.sprakradet.no/upload/Klarspr%c3%a5k/Dokumenter/2013%20Klarspr%c3%a5k%20bm.pdf>

Språkrådet. (2023, mai 8). *Anførselstegn eller replikkstrek for ungdomsskoleelever?*

Språkrådet. <http://www.sprakradet.no/svardatabase/sporsmal-og-svar/anforselstegn-eller-replikkstrek-for-ungdomsskoleelever/>

Staiano, F. (2022). *Designing and Prototyping Interfaces with Figma: Learn essential UX/UI design principles by creating interactive prototypes for mobile, tablet, and desktop*. Packt Publishing Ltd.

Statped. (2020, juni 2). *Hva er dysleksi?* <https://www.statped.no/lese--og-skrivevansker/lese--og-skrivevansker/hva-er-dysleksi/>

Søk og Skriv. (2023, mai 2). *Korleis skal ein referere? | Søk & Skriv*.
<https://www.sokogskriv.no/kjeldebruk/korleis-skal-ein-referere.html#ved-endringar-i-sitatet>

Web Content Accessibility Guidelines (WCAG) 2.1. (u.å.). Hentet 22. april 2023, fra <https://www.w3.org/TR/WCAG21/>

Williamson, K., & Johanson, G. (Red.). (2018). *Research methods: Information, systems, and contexts* (Second edition). Chandos Publishing.

WordHero. (u.å.). *#1 AI Writing Software / AI Writer & Assistant / WordHero*. Hentet 1. juni 2023, fra <https://wordhero.co/>

Wordtune Team. (u.å.). *Announcing Wordtune API*. Hentet 23. april 2023, fra <https://www.wordtune.com/blog/announcing-wordtune-api>

Zaman, F., Shardlow, M., Hassan, S.-U., Aljohani, N. R., & Nawaz, R. (2020). HTSS: A novel hybrid text summarisation and simplification architecture. *Information Processing & Management*, 57(6), 102351. <https://doi.org/10.1016/j.ipm.2020.102351>

10 VEDLEGG

VEDLEGG A: INTERVJUGUIDE TIL SEMI-STRUKTURERTE INTERVJUER MED MÅLGRUPPEN

- Greet the participant
 - Introduce yourself
 - Thank the participant for taking part in the interview
 - Give a short introduction into the topic of the interview
 - Let participant sign consent form
 - Start recording
1. Har du lest noen nyheter i dag? Kan du fortelle litt om hva det handlet om?
 2. Hva slags nyheter er du mest interessert i?
 - a. Fortell mer?
 3. Hva er ditt forhold til nyheter? Hvor ofte oppsøker du nyheter?
 4. Hvordan oppsøker du nyheter (nettavis/podcast/papir/mobil/pc, osv.)?
 - a. Hva er det som gjør at du velger denne måten å oppsøke nyheter?
 - b. Hvis flere: hva bruker du mest/foretrekker du? Hvorfor?
 5. Hva pleier du å gjøre når du leser en artikkel?
 - a. Eks. se på overskrifter/bilder osv.
 - b. Skumlese
 - c. Lese hele artikkelen
 6. Er det noe du ikke liker eller syns er hindrende når du leser en artikkel?
 - a. Eksempel?
 - b. Eks. lange/kompliserte setninger, lange ord, lang tekst og kompliserte ord

7. Hva gjør at en artikkel føles interessant og lett å lese?
 - a. Eksempel?
 - b. Eks. oppsett av artikkelen, setningsstruktur, osv.
 1. Hva gjør at du slutter å lese en artikkel?
 2. Har du prøvd tillegg i nettleseren som gjør noe med teksten?
 - a. Hvilke tillegg og hva gjorde de? (eks. lese opp tekst)
 - b. Bruker du det fortsatt, evt hva fikk deg til å slutte å bruke det?
 3. Er det noe mer du vil legge til eller har du noen spørsmål?
- Stop the recording
 - Thank the participant for taking part in the interview

VEDLEGG B: PYTHON-KODE OG HTML-KODE FRA TEKNISK PROTOTYPE

```
# Import the necessary libraries
from flask import Flask, request, render_template
import openai

# Insert valid OpenAI API key here
openai.api_key = "API_key"

# Create a Flask app
app = Flask(__name__)

# Define the route for the home page
@app.route('/')
def home():
    return render_template("home.html")

# Define the route for the result page
@app.route('/result', methods=['POST'])
def result():
    # Get the input text from the form
    input_text = request.form["input_text"]
```

```

# Use the OpenAI GPT-3 model to rewrite the text
completion = openai.Completion.create(
    engine="text-davinci-003",
    prompt=f"Rewrite the following text in an easy-to-understand language for people with dyslexia: {input_text}\n",
    temperature=0.5,
    max_tokens=1500,
    top_p=1,
    frequency_penalty=0,
    presence_penalty=0
)

# Get the rewritten text from the Completion object
dyslexia_friendly_text = completion.get('choices')[0].get('text')

# Render the result template and pass the rewritten text to it
return render_template('result.html', dyslexia_friendly_text=dyslexia_friendly_text)

# Run the app
if __name__ == '__main__':
    app.run(debug=True)

```

Figur 9: Python-kode fra teknisk prototype.

```

<!DOCTYPE html>
<html>
  <head>
    <meta name="viewport" content="width=device-width">
    <link rel="stylesheet" type="text/css" href="style.css" />
    <title>Tekstforenkling for dysleksi</title>
  </head>
  <body>
    <div class="page">
      <div class="heading">
        <div class="page_text">
          <p id="headline">Tekstforenkling</h1>
        </div>
        <div class="help-tip">
          <p>Skriv teksten din i boksen under og klikk på "Skriv om" for å gjøre teksten mer lettlest. </p>
        </div>
      </div>
    <div class="inp">

```


```
<form action="../../templates/result.html" method="post">
  <div class="textbox" contenteditable="true" name="input_text" rows="10" cols="50"> </div>
  <input class="input_button" type="submit" value="Skriv om">
</form>
</div>
</div>
</body>
</html>
```

Figur 10: HTML-kode fra teknisk prototype.

VEDLEGG C: TEKSTFORENKLINGER

Dette vedlegget er lagt til som et eksternt Excel-dokument.

VEDLEGG D: PERSONAS



Ingrid Evensen
24 years old, Ålesund

Studies & Job

- 5th bachelor semester in nautical science at NTNU
- previously quit biotechnology after 2 semesters
- has been working part-time at a café for 2 years

Background

Ingrid grew up in Stordal near Ålesund and has struggled with reading from early on, but she only got a diagnosis with dyslexia in 10th grade. She usually has more difficulties with reading text than writing it.

She moved out of her parents' home after finishing school and did a year of military in Northern Norway. Then, she went to Ålesund for her studies where she currently lives in a shared apartment with three other students. One of them also studies nautical science with her.

The dyslexia affects her on a daily basis at the university, for example when she has to read scientific papers or books from her syllabus as well as when writing essays. Ingrid tries to work around it by using tools that read the text out loud, but she finds those often annoying after some time because the voice does not sound very natural. Also, she uses a tool that suggests word and phrases to help her with writing.

News behaviour

- consumes news mostly via podcasts and tv, thinks it takes too long time to read articles
- thinks it's takes too long time to read articles, only worth for very interesting topics like marine life or current global news

Hobbies

- sailing since childhood
- interest in marine-related topics
- goes to the gym in her freetime

Figur 11: Persona A



Harald Ødegård

46 years old, Voss

Job

- Electrician

Background

Harald is 43 years old and has worked as an electrician for over 20 years. He loves his job and really likes that every day at work is different. Harald is married and has two children aged 15 and 12. He is very involved with his family and likes to go hiking on the weekends. Often has to drive his children to and from activities.

Harald has always struggled with reading and writing but has not formally been diagnosed with dyslexia. He doesn't want to get diagnosed because he does not think it will make a difference. He avoids reading and writing when possible, for example calling instead of messaging.

News behaviour

- He is most interested in local and national news, but keeps up with world news that the Norwegian media writes about.
- Always listen to the radio on his way to/from work and at work
- Sometimes reads the news during his lunch break
- Reads news on phone (NRK app) and sometimes local news paper (paper version)

Hobbies

- Hiking
- Skiing
- Working on projects around the house
- TV and movies

Figur 12: Persona B