

UiO  **Institutt for informatikk**  
Det matematisk-naturvitenskapelige fakultet

# «Chatbot for alle?»

En kvalitativ studie av dyslektikers opplevelse med chatbot

Masteroppgave - Anton Lilleby og Steffen Marstein

VÅR 2020





# «Chatbot for alle?»

En kvalitativ studie av dyslektikers opplevelse med chatbot

Anton Lilleby og Steffen Marstein

© Anton Lilleby og Steffen Marstein  
2020

«Chatbot for alle?» En kvalitativ studie av dyslektikers opplevelse med chatbot

Anton Lilleby og Steffen Marstein

<http://www.duo.uio.no/>

Trykk: Reprosentralen, Universitetet i Oslo



# Forord

Guri. Først og fremst tusen takk for utallige timer med veiledning i verdensklasse. Du har gitt oss solid støtte gjennom hele denne prosessen. Vi er glade for at du har latt oss utforske i stor fleksibilitet, og gitt oss mulighet til å finne en oppgave vi selv synes har vært interessant. Covid-19 har ikke vært en stopper for at du har tatt initiativ til ukentlige diskusjoner over Zoom. I tillegg er din evne til å holde vår motivasjon oppe en bragd i seg selv.

Dysleksi Ungdom. Takk for et lærerikt samarbeid og inspirerende seminar. Vi ønsker også å takke andre deltakere, dette studiet hadde ikke vært noe uten dere.

Vetle og Ivar. Studiehverdagen ville ikke vært det samme uten dere – kjølige-boys.

Familie, samboer og andre som har stått og heiet på oss bokstavelig talt frem til mållinjene. Tusen takk.

**Anton Lilleby og Steffen Marstein**

Universitetet i Oslo

Juni 2020



# Innholdsfortegnelse

<b>1. Introduksjon .....</b>	<b>1</b>
1.1 Motivasjon .....	1
1.2 Avgrensning.....	2
1.3 Forskningsspørsmål .....	3
1.4 Forskningsbidrag .....	3
<b>2. Bakgrunn.....</b>	<b>5</b>
2.1 Spesifikke lese- og skrivevansker.....	5
2.2 Universell utforming.....	8
2.2.1 Bakgrunn og historie.....	8
2.2.2 Universell utforming i kontekst av IKT.....	10
2.2.2.1 Teknisk tilgjengelighet.....	11
2.2.2.2 Brukervennlighet .....	13
2.2.2.3 Dysleksi og tilgjengelighet.....	15
2.3 Digital forvaltning .....	15
2.4. Chatbot.....	18
2.4.1 Chatbot generelt .....	19
2.4.2 Chatbot og universell utforming .....	19
2.4.3 Chatbot innen forvaltningen .....	20
2.4.4 Kunstig intelligens .....	21
2.4.5 Maskinlæring .....	21
2.4.6 Nevrale nettverk.....	22
2.4.7 Chatbot arkitektur .....	22
2.4.8 Regelbasert chatbot.....	22
2.4.9 Korpusbasert chatbot .....	23
2.5 Samtaledesign.....	26
2.5.1 Samtaledesign generelt .....	26
2.5.2 Talehandling - en ytrer alltid noe med et mål .....	27



2.5.3 Samarbeidsprinsippene – en vellykket samtaleprosess.....	28
2.5.4 Turtaking - hvem sin tur er det å snakke?.....	29
2.5.5 Initiativtaking - hvem styrer samtalen?.....	30
2.5.6 Felles enighet - bygge et felles grunnlag .....	30
<b>3. Metodologi.....</b>	<b>32</b>
3.1 Forskningstilnærming .....	32
3.2 Brukersentrert forskningsstrategi .....	32
3.3. Metode og designprosess.....	35
3.3.1 Observasjon.....	36
3.3.2 Ekspert intervju.....	37
3.3.3 Automatisk evaluering .....	39
3.3.4 Brukertestning.....	40
3.3.5 Data analyse .....	44
3.3.6 Metoder og teknikker for design.....	46
3.3.6.1 Brainstorming og skissering.....	46
3.3.6.2 Prototyping .....	47
3.3.6.3 Evaluerings-workshop.....	48
3.3.7 Rekruttering .....	50
3.4 Etske betraktninger .....	50
<b>4. Brukertestning av chatbotene .....</b>	<b>53</b>
4.1 Intervju med deltakerne .....	53
4.1.1 Om deltakerne .....	53
4.1.2 Lese- og skrivestøtteprogram.....	53
4.1.3 Erfaringer med Chat.....	54
4.1.4 Erfaringer med Chatbot.....	54
4.2 Brukertestning rapport.....	55
4.2.1. Skatteetatens chatbot oppsummering.....	57
4.2.2. NAVs chatbot oppsummering.....	64
4.3 Analyse av brukertestningen .....	70
4.3.1 Chatbot brukergrensesnittet .....	70

4.3.2 Problemer i forhold til navigasjon .....	71
4.3.3 Er chatten styrt av et menneske eller en maskin? .....	71
4.3.4 Avbryte leseprosessen.....	72
4.3.5 Svaralternativer og lenker .....	73
4.3.6 Skrivefeil og chatbot predikasjon .....	74
4.3.7 Mengde tekst og Vanskelig språk .....	76
4.3.8 Menneskelig kontakt.....	77
<b>5. Designforslag.....</b>	<b>79</b>
5.1 Prototyper .....	79
5.1.1 Forhindre avbrytelser i lesingen.....	79
5.1.2 Forklar begreper og vanskelige ord .....	81
5.1.3 Større grensesnitt, mindre scrolling .....	83
5.1.4 Les teksten opp for brukeren.....	84
5.1.5 En mer tilgjengelig navigasjon .....	86
5.2 Anbefalinger .....	87
5.2.1 Mer bruk av svaralternativer .....	87
5.2.2 Vurder en annen toleranse for stavefeil .....	88
5.2.3 La en bruker korrigere stavefeil.....	89
5.2.4 Redusere unødvendige chatbot meldinger .....	89
<b>6. Evaluering .....</b>	<b>91</b>
6.1 En mer tilgjengelig navigasjon .....	91
6.2 Større grensesnitt, mindre scrolling.....	91
6.3 Forhindre avbrytelser i lesingen .....	92
6.4 Forklar begreper og vanskelige ord .....	92
6.5 Mer bruk av svaralternativer.....	92
6.6 Les teksten opp for brukeren .....	92
<b>7. Diskusjon.....</b>	<b>94</b>
7.1 Brukertestning som metode .....	94
7.2 Bruk av støtteprogram .....	94

7.3 Har vi tilstrekkelig antall deltakere?.....	95
7.4 Rekruttere deltakere fra en sårbar gruppe.....	96
7.5 Arbeidet med designforslag og universell utforming .....	97
7.6 Samtaledesign som teori .....	100
7.7 Fra menneskelig kontakt til chatbot.....	101
<b>8. Konklusjon .....</b>	<b>104</b>
8.1 Videre forskning .....	106
<b>9. Litteratur.....</b>	<b>108</b>
<b>10. Vedlegg .....</b>	<b>119</b>
Vedlegg A – Samtykkeskjema brukertesting .....	119
Vedlegg B – Samtykkeskjema intervju .....	121
Vedlegg C – Automatisk evaluering Skatteetaten .....	124
Vedlegg D – Automatisk evaluering NAV.....	124
Vedlegg E – Testplan .....	125
Vedlegg F – Funka .....	125
Vedlegg G – Dysleksi Norge.....	127
Vedlegg I – Skatteetaten.....	128



# Liste av figurer

<b>Figur 1.1.</b> Typer lesevansker.....	7
<b>Figur 2.1.</b> Gap-modellen.....	9
<b>Figur 2.2.</b> Trender for kontakt henvendelser i NAV.....	20
<b>Figur 2.3:</b> Venn-diagram over chatbot arkitekturer og kunstig intelligens.....	22
<b>Figur 2.4.</b> Språkforståelse komponenter i korpusbaserte chatboter.....	24
<b>Figur 2.5.</b> En kanonisk interaksjonsstruktur for chatbot innen kundeservice.....	29
<b>Figur 3.1.</b> Standardmodellen ISO 9241-210.....	34
<b>Figur 3.2.</b> Oversikt over forskning- og designprosessen i dette studiet.....	35
<b>Figur 3.3.</b> Funn og beskrivelser fra brukertesten.....	44
<b>Figur 3.4.</b> Organisering og kategorisering av data.....	45
<b>Figur 3.5.</b> Dele inn i temaer.....	45
<b>Figur 3.5.</b> Brainstorming.....	47
<b>Figur 3.6.</b> Skisseringer.....	47
<b>Figur 3.7.</b> Presentasjon av evalueringsworkshopen.....	49
<b>Figur 4.1.</b> Skatteetatens chatbot landingsside.....	57
<b>Figur 4.2.</b> Skatteetatens chatbot predikerer noe av intensjonen rett.....	58
<b>Figur 4.3.</b> Skatteetatens chatbot predikerer feil.....	58
<b>Figur 4.4.</b> Skatteetatens chatbot med svaralternativer.....	60
<b>Figur 4.5.</b> Skatteetatens chatbot mye tekst.....	60
<b>Figur 4.6.</b> Skatteetatens chatbot predikerer noe av intensjonen rett.....	61
<b>Figur 4.7.</b> Skatteetatens chatbot blir forstått feil.....	61
<b>Figur 4.8.</b> NAVs chatbot landingsside.....	64
<b>Figur 4.9.</b> NAVs chatbot predikerer intensjonen rett.....	65
<b>Figur 4.10.</b> NAVs chatbot med mange svarmeldinger.....	65
<b>Figur 4.11.</b> NAVs chatbot vurderer at brukerintensjonen allerede er besvart.....	66
<b>Figur 4.12.</b> NAVs chatbot med svaralternativer.....	66
<b>Figur 4.13.</b> NAVs chatbot viser et svaralternativ.....	68
<b>Figur 4.14.</b> NAVs chatbot predikerer intensjonen feil, men gir rett valg.....	68
<b>Figur 4.15.</b> Avataren til Skatteetatens og NAVs chatbot.....	71
<b>Figur 4.16.</b> Illustrasjon av avbrytelse i leseprosess i NAVs chatbot.....	72
<b>Figur 5.1.</b> Prototype av en chatbot som forhindrer avbrytelse i leseprosessen.....	79
<b>Figur 5.2.</b> Prototype av en chatbot som forklarer begreper og vanskelige ord.....	81
<b>Figur 5.3.</b> Prototype av en forstørret versjon av chatbotgrensesnittet.....	83
<b>Figur 5.4.</b> Prototype av tekst-til-tale funksjonalitet i øvre del av chatbot.....	84
<b>Figur 5.5.</b> Prototype av tekst-til-tale funksjonalitet i chatbot meldingene.....	84
<b>Figur 5.6.</b> Prototype av en klarer navigasjon på NAVs nettside.....	86



# 1. Introduksjon

Skatteetaten og Norges Arbeids- og velferdsetaten (NAV) er to av de største organene i offentlig forvaltning. Digitalisering av forvaltningsorganer slik som Skatteetaten og NAV har pågått over lengre tid. Samtidig som at brukerne skal bli mer selvhjulpne på nett, så blir digital kommunikasjon mellom innbyggere og offentlig forvaltning prioritert. Med dagens avansement i maskinlæring representerer teknologi nye muligheter for kommunikasjon med brukere. Chatbot er et dataprogram som kan simulere menneskelig samtale som automatiserer samtaler tradisjonelt gjort mellom mennesker til samtaler mellom dataprogram og menneske. Flere organer i forvaltningen tilbyr en chatbot for å svare på brukerhenvendelser. Det er derimot lite kunnskap om denne informasjonskanalen er tilgjengelig og brukervennlig for mennesker med nedsatt funksjonsevne. For å bidra med økt forståelse, vil vi derfor i dette studiet utforske hvordan det oppleves å interagere med Skatteetaten og NAV sine chatboter for personer med spesifikke lese- og skrivevansker, også kjent som dysleksi, og hvordan man kan designe chatboter for å imøtekomme dyslektikernes behov.

## 1.1 Motivasjon

Digitale kommunikasjonskanaler har blitt en viktig forutsetning for en demokratisk deltakelse i samfunnet. Nesten alle nordmenn mellom 16 og 79 år bruker internett, to tredjedeler bruker sosiale medier og 79 prosent henter informasjon fra offentlige nettsider (Statistisk sentralbyrå [SSB], 2017, 2019). Vedtak viser at digital kommunikasjon er hovedregelen befolkningen skal forholde seg til med offentlig forvaltning (Kommunal- og moderniseringsdepartementet, 2016, s. 40), og de siste årene har flere offentlige organer begynt å ta i bruk tekstbasert chatbot for digital kommunikasjon med sine brukere. Men mange sliter med å lese og skrive, særlig personer med dysleksi. Dysleksi er spesifikke lese- og skrivevansker som det antas at opptil 10 prosent av Norges befolkning har (Dysleksi Norge, u.d.; Rongved, 2015). Det er en fare for at denne gruppen i befolkningen havner utenfor samfunnsutviklingen når slike digitale kommunikasjonsformer blir mer utbredt.

Fra kommunikasjonskanaler der man tidligere har snakket med et menneske vil man nå kunne oppleve å komme i kontakt med en chatbot. NAV og Skatteetaten er blant noen offentlige etater som gradvis har lansert chatbotløsninger på sine nettsider. Det antas at chatbot vil

reduere kontakt ved telefon, chat og publikumsmottak betraktelig frem mot 2025 i NAV (T. Simonsen, 2019). Dette er noen av årsakene til at chatboter har fått mye oppmerksomhet den siste tiden. Man vet derimot lite om hvordan personer med funksjonsnedsettelse, som for eksempel dyslektikere, opplever å bruke teknologien. Tekst er tilstedeværende i de fleste digitale plattformer og det inkluderer sosiale medier, kommentarfelt og private meldinger, her blir dyslektikere sine skrivefeil fort påpekt av andre de snakker med (Kahrs & Flølo, 2019) og i ytterste konsekvens stopper dialog. Slike tilfeller viser seg å ha betydelig innvirkning på dyslektikere. Studier dokumenterer blant annet mangel på selvtillit og stigmatisering assosiert med dysleksi (McNulty, 2003; Mortimore & Crozier, 2006).

«Med universell utforming menes at utforming eller tilrettelegging av hovedløsningen i informasjons- og kommunikasjonsteknologi er slik at virksomhetens alminnelige funksjon kan benyttes av flest mulig», heter det i forskriften om universell utforming (Forskrift om universell utforming av IKT-løsninger, 2013, § 1). Universell utforming er en designfilosofi som handler om å utforme omgivelser og produkter slik at et bredt mangfold kan bruke samme hovedløsning, uavhengig av funksjonsevne, inkludert personer med dysleksi. Digitaliseringsdirektoratet er kontrollorganet som bedømmer om digitale løsninger følger forskriften om universell utforming blir ivarettatt. Direktoratet har per i dag ikke undersøkt om chatbot oppfyller retningslinjene for universell utforming. Asbjørn Følstad seniorforsker i SINTEF sier at man trenger mer forskning på dette området. Torres, Franklin og Martins (2018) forsterker denne oppfatning som i sin litteraturstudie vurderer at «chatbot og tilgjengelighet» er et område som mangler akademiske studier. McCarthy og Swierenga (2010) hevder også at man vet lite når det gjelder utfordringer dyslektikere kan ha med nettbaserte grensesnitt. På bakgrunn av et behov for kunnskap i universell utforming av chatbot og dyslektikere sine opplevelser på nett vil vi undersøke hvordan dyslektikere opplever bruk av chatbot i Skatteetaten og NAV. Studiet har også en hensikt å utforske hvordan man kan imøtekomme dyslektikere sine behov ved denne teknologien.

## 1.2 Avgrensning

Studiekonteksten er Norge der universell utforming blir brukt som begrep i forvaltningen. Universell utforming av nettløsninger handler om å øke tilgjengeligheten for et bredt mangfold gjennom en hovedløsning, inkludert mennesker med funksjonsnedsettelse. Det



finnes imidlertid svært mange funksjonsnedsettelse under kategoriene kognitive, sensoriske og motoriske. I denne studien avgrensner vi til å kun se på personer med dysleksi som kategoriseres som en kognitiv funksjonsnedsettelse. Vi har valgt å avgrense til å se på Skatteetatens og NAVs chatbot. De fleste offentlige organer i Norge benytter til vår forståelse samme chatbotleverandør, som er teknologiselskapet Boost.AI. Forskningsspørsmålet har blitt utforsket med en kvalitativ tilnærming. Testbare kriterier som forekommer i standarder og retningslinjer har derfor ikke vært hovedfokus, selv om de ofte blir brukt i sammenheng med universell utforming.

### 1.3 Forskningsspørsmål

Vi har hatt en overordnet interesse til å utforske digital kommunikasjon med offentlig forvaltning for sårbare brukergrupper. Denne interessen har vi delt inn i to spesifikke forskningsspørsmål som vi videre gjør rede for.

**1. Det første forskningsspørsmålet vi stiller er:** Hvordan oppleves chatbotene til Skatteetaten og NAV for dyslektikere?

Med dette spørsmålet ønsker vi å etablere en innsikt i hvordan personer med dysleksi opplever med chatbotene til to av de største etatene i offentlig forvaltning.

**2. Det andre forskningsspørsmålet vi stiller er:** Hvordan designe en chatbot for å imøtekomme dyslektikere sine behov?

Med dette spørsmålet ønsker vi å utforske behovene til dyslektikere nærmere. Vi mener behov er forankret i vår analyse av bakgrunns litteratur, ekspert- og brukerinnsikt. Med design ønsker vi å utforske om man kan øke tilgjengeligheten i chatbot for dyslektikere, men også for ikke-dyslektikere.

### 1.4 Forskningsbidrag

Vi håper dette studiet kan bidra med en tidlig innsikt i hvordan chatboter innen forvaltningen oppleves og fungerer for personer med funksjonsnedsettelsen dysleksi. Forbedringsforslag når det gjelder utfordringer har blitt identifisert og evaluert. Som et steg for å etablere kunnskap om å øke tilgjengeligheten for dyslektikere, som også kan gagne andre brukergrupper.

I dag mangler det innsikt til hvordan chatboter oppleves for mennesker med funksjonsnedsettelse. En målsetting med dette studiet er å gi grunnlag for å etablere en innsikt, og dermed adressere noe av gapet i kunnskap om chatbot tilgjengelighet. Dette studiet har hatt dyslektikere som en brukergruppe. I litteraturen blir det rapportert at dyslektikere sine opplevelser på nett er noe man vet lite om (McCarthy & Swierenga, 2010). Vi anser temaet som relevant innenfor feltet Human-Computer Interaction (HCI) som ser på brukermangfold og tilgjengelige grensesnitt (Sandnes, 2011).

Forskningsgruppen design av informasjonssystemer (DESIGN) ved Universitet i Oslo, Funka, SINTEF, Skatteetaten, NAV, Boost.AI og Digitaliseringsdirektoratet håper vi også kan dra nytte av denne studien.

## 2. Bakgrunn

I dette kapitlet presenterer vi bakgrunns litteratur som er relevant for det vi undersøker. Kapitlet er delt inn i fem deler. Først del omhandler spesifikke lese- og skrivevansker. Deretter ser vi på universell utforming som et begrep for tilgjengelighet. Videre forklarer vi hva en digital forvaltning er, og motivet for at teknologien chatbot begynner å bli fremtredende i samfunnet. Til slutt ser vi på hva som muliggjør chatbot i dag og hvordan man kan designe for en naturlig samtale i denne teknologien.

### 2.1 Spesifikke lese- og skrivevansker

Dysleksi er sammensatt av de to greske ordene «dys» som betyr vanske og «leksi» som betyr ord, derav betyr dysleksi vansker med ord (Dysleksi Norge, u.d). Det er flere faktorer som kan føre til utvikling av dysleksi. Dysleksi Norge sier på sine nettsider at dysleksi er et resultat av et samspill mellom arv og miljø. Om en forelder har dysleksi er det 38 – 50% sjanse for at barnet arver gener som kan gi disposisjon for dysleksi, men hvordan dysleksien utvikler seg kan også være bestemt av visse miljøbetingelser (Dysleksi Norge, u.d.; Klinkenberg, 2017).

Forskere innen spesifikke lese- og skrivevansker har problemer med å samle seg under en felles definisjon for dysleksi, og professor Finn Egil Tønnessen ved Universitetet i Stavanger mener det ikke finnes en klar vitenskapelig definisjon på hva dysleksi er (Tønnessen, 2015).

Med utgangspunkt i operasjonelle definisjoner utarbeidet av International Dyslexia Association (IDA) og British Dyslexia Association (BDA) har Dysleksi Norge laget en egen definisjon til dysleksi, som er som følger:

Dysleksi er en spesifikk lærevanske som gjør det vanskelig å tilegne seg funksjonell lese- og skriveferdighet. Typiske kjennetegn er derfor 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. Noen har også vansker med prosesseringshastigheten og

automatiseringsevnen. Vanskene avviker fra personens øvrige kognitive ferdigheter.  
(Dysleksi Norge, 2017)

Sandnes (2011) identifiserer noen av vanskene nevnt i Dysleksi Norges definisjon:

- Oppfattelse av informasjon
- Huske og bearbeide fonetiske lyder (språklyd)
- Skille ut lydene i ord
- Bygge opp ord fra fonetiske grunnkomponenter

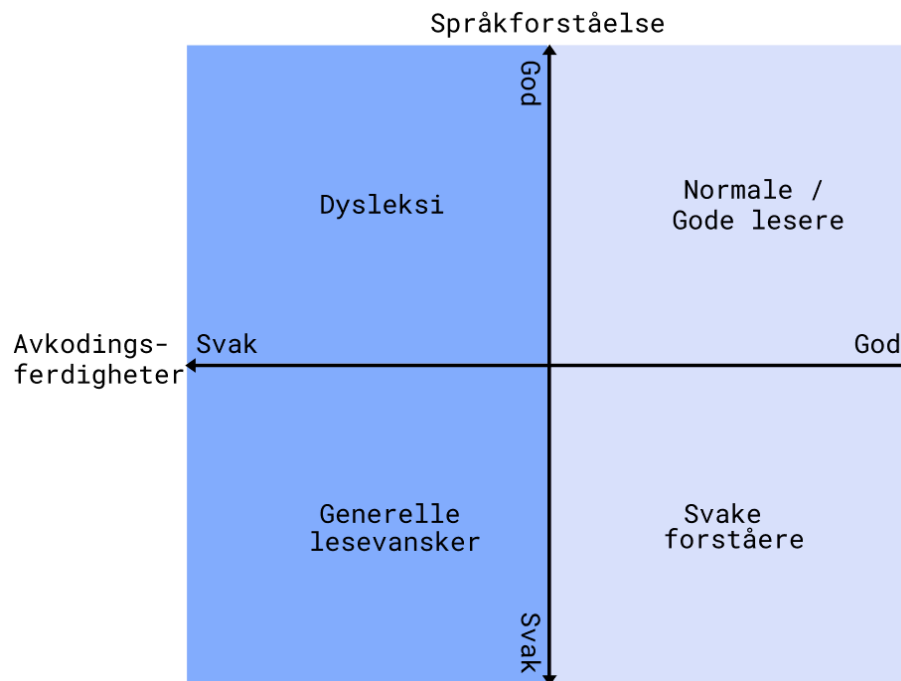
Mange har både lese- og skrivevansker og dette forekommer av ulike årsaker som for eksempel evner, konsentrasjonsvansker og mangelfull opplæring. Slike generelle vansker kan, med målrettet trening, forsvinne helt. Dysleksi er derimot en spesifikk vanske, og det skyldes ikke generelle årsaker som nevnt over, og selv om funksjonsvansker kan reduseres forsvinner de aldri helt. Det er om lag 5% av befolkningen som har dysleksi, noen forskere oppgir også inntil 10% (Dysleksi Norge, u.d; Rongved, 2015; Ahissar, 2007; Sandnes, 2011, s. 192).

Dysleksi må forstås i et livsløpsperspektiv da lese- og skrivevanskene følger personen livet ut (Swanson & Hsieh, 2009). Det er ikke uvanlig med komorbiditet mellom dysleksi og andre funksjonsnedsettelse. Forskning viser sammenheng med Attention Deficit Hyperactivity Disorder (ADHD), Attention Deficit Disorder (ADD) og dyskalkuli (Willcutt et al., 2010; Germanò, Gagliano & Curatolo, 2010).

Vanskene en person kan ha på bakgrunn av dysleksi er individuelt og forskjellig, men ofte innebærer utfordringene svikt i det fonologiske systemet og det manifesterer seg i form av lese- og skrivevansker. For eksempel kan noen være trege lesere, men ha en god forståelse av innholdet, mens andre leser raskt, men er unøyaktig når det gjelder å få med seg innholdet. Vansker med ordavkodning og leseforståelse finnes som regel i alle former for dysleksi (Shaywitz & Shaywitz, 2005).

Gough og Tunmer (1986) utviklet en enkel og mye brukt formel for lesing, «The simple view of reading»: Lesing = avkodning x forståelse. Hensikten med å lese er ofte for å hente informasjon av en tekst. Denne modellen viser to komponenter man må ha for å kunne lese og forstå en tekst. Man må kunne avkode ordene samt forstå det språklige innholdet. Om en av faktorene i formelen, avkodning eller forståelse, ikke er til stede blir også resultatet null. Det

vil si, om du ikke har forståelsen, men bare har avkodet ordene, har man i grunnen ikke lest. Avkoding og forståelse er derfor en forutsetning for at lesing skal finne sted. For dyslektikere ligger ikke problemet på forståelsen, men avkodingen av selve ordet.



**Figur 1.1.** Viser en oversikt over forskjellige typer lesevansker. Basert på Klinkenberg (2017).

Sandnes (2011, s. 192) sier at dyslektikere også har andre tilleggssymptomer som omhandler, nedsatt arbeidsminne, lav konsentrasjonsevne, nedsatt motorikk og sekvensieringsvansker. Det er også veldokumentert at slike vansker fører til at mange dyslektikere sliter med selvfølelsen og at man i enkelte situasjoner mister motet (Carawan, Nalavany & Jenkins, 2014).

Den digitale utviklingen har også endret premissene på noe av funksjonsvanskene når det kommer til å lese og skrive. Digitale støtteprogrammer som er tilpasset dyslektikere har blitt mer utbredt. Disse hjelpemidlene kan være skriveprogrammer, talesyntese og digitale lydbøker. To eksempler er Lingdys og Textpilot som kan brukes for å støtte input og output av tekstlig innhold på nett. Disse programmene retter feilstavede ord og kommer med fullføringsforslag kontinuerlig. Det er også mulig å lytte til tekst med tekst-til-tale funksjonalitet, samt få hjelp til sammenslåing av ord, noe kan være problematisk og vanskelig med det norske språk (Sandnes, 2011)

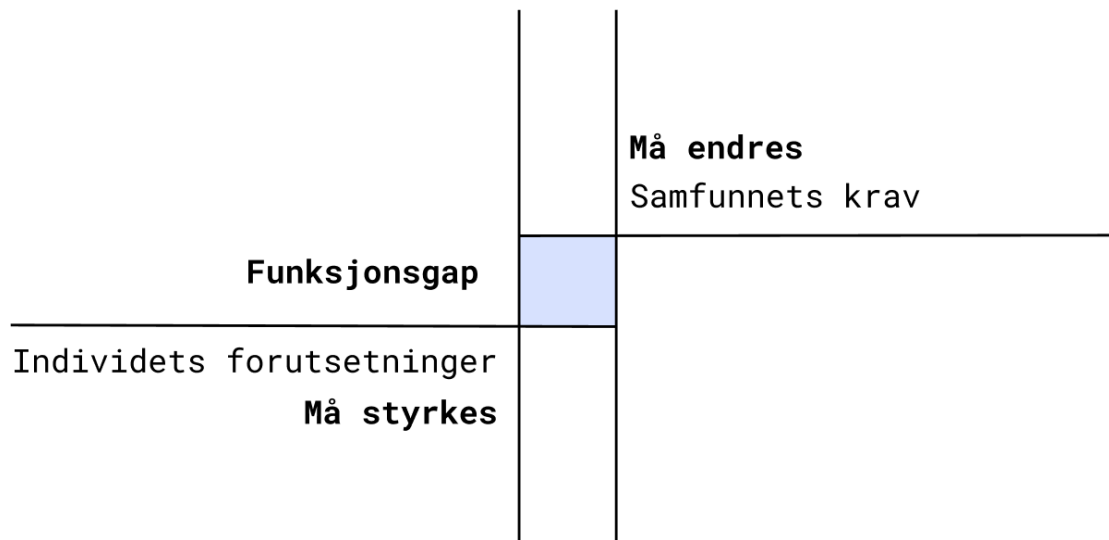
Generalsekretær i Dysleksi Norge, Caroline Solem (2017), sier at utviklingen av slike støtteprogram har vært enorm i de siste årene, og at slike program kan være med på å bidra til mestring og motivasjon samt gi en reell sjanse for dyslektikere til å henge med i et mer skriftlig samfunn.

## 2.2 Universell utforming

Siden dette studiet er utført i Norge der lovgivningen bruker begrepet universell utforming, ser vi det som relevant å inkludere noe av bakgrunnen og historien som har ført til den gjeldende lovgivning. Det neste avsnitt presenterer først en bakgrunn til universell utforming, og videre en forklaring om hvordan strategien blir brukt i kontekst av informasjons- og kommunikasjonsteknologi (IKT), som vi anser som relevant med tanke på at chatbot er en IKT-løsning.

### 2.2.1 Bakgrunn og historie

Fra et historisk perspektiv har behandling og oppfatninger av funksjonshemming gjennomgått flere endringer. Dette har skjedd i stor grad fordi mennesker med nedsatt funksjonsevne har krevd disse endringene selv. På 1950-tallet startet borgerrettsbevegelsen, en protest for medborgerlige rettigheter i USA. Dette førte senere til et forbud mot diskriminering på bakgrunn av rase, farge, religion, kjønn eller nasjonalitet i 1964 (Vavik & Gheerawo, 2009). Et forbud mot diskriminering av funksjonshemmede kom visstnok først i 1990 i USA (Persson, Åhman, Yngling & Gulliksen, 2015), men mange hendelser, vedtak og engasjerte enkeltmennesker har formet denne utviklingen. Sånn sett må lovgivningen sees i sammenheng med den økende bevisstgjøringen av diskrimineringen av funksjonshemmede som begynte med borgerrettsbevegelsen. Noen viktige årsaker til bevisstgjøring er blant annet (1) borgerrettsbevegelser sine gjennombrudd, (2) borgere som protesterte på behandlingen de fikk som funksjonshemmede i etterkrigstiden, og at (3) Forente Nasjoner (FN) vedtok at året 1976 skulle bli det internasjonale året for funksjonshemmede (Persson et al., 2015; Vavik & Gheerawo, 2009).



**Figur 2.1.** Gap-modellen illustrerer at funksjonshemming er en årsak av et misforhold mellom kravene i samfunnet og forutsetningene til individet. Med utgangspunkt i Sandnes (2011, s. 24).

Likeverd og like muligheter for alle borgere har vært en prioritert del av norsk politikk siden slutten av 1970-årene. Stortinget fastslo allerede da at samfunnet hadde et ansvar for å tilrettelegge for personer med nedsatt funksjonsevne, og på 1990-tallet erkjente de at funksjonsevne måtte sees i forhold til det samfunnet som omgis en (Sosialdepartementet, 1977, s. 10-11; Sosial- og helsedepartementet, 1997, s. 6). At begrepet funksjonshemming ble forstått som situasjonsbestemt var et viktig paradigmeskifte fordi begrepet hadde tidligere blitt vurdert som en biologisk eller medisinsk egenskap ved et individ (NOU 2005: 8, 2005, s. 21). I dag betyr ikke en funksjonsnedsettelse nødvendigvis at man er funksjonshemmet. En funksjonsnedsettelse forstås som tap av eller skade på en kroppsdel eller i en av kroppens funksjoner (Sandnes, 2011). Den såkalte gap-modellen, som illustrert i figur 2.1, viser at funksjonshemming er en årsak av et misforhold mellom kravene i samfunnet og forutsetningene til individet. Formålet med gap-modellen er å identifisere, redusere og eliminere samfunnsskapt barrierer som virker hemmende og diskriminerende (Sandnes, 2011). Norge har i dag Diskriminerings- og tilgjengelighetsloven (2008, § 1) som nettopp handler om å nedbygge samfunnsskapt barrierer og hindre at nye skapes.

Stortinget i slutten av 1990-tallet viser implisitt til universell utforming som en prosess for å løse sentrale utfordringer for mennesker med nedsatt funksjonsevne på visse områder i samfunnet, blant annet tilgjengelighet i bygninger og uteareal (Sosial- og helsedepartementet, 1997, s. 58-60). Universell utforming har sine røtter fra arkitekturen med mål om å designe

bygninger og uteområder barrierefritt, og tilgjengelig. Begrepet ble først definert av arkitekten Ronald Mace i 1985, som definerte universell utforming som: «The design of products and environments to be usable to the greatest extent possible by people of all ages and abilities» (Story, Mueller & Mace, 1988, s. 2).

Definisjonen fremmes også av syv anerkjente prinsipper for universell utforming som ble laget av The Center for Universal Design i 1997 i USA. De syv prinsippene er (1) like muligheter for bruk, (2) fleksibel bruk, (3) enkel og intuitiv bruk, (4) forståelig bruk, (5) toleranse for feil, (6) lav fysisk anstrengelse, og (7) størrelse og plass for tilgang og bruk (Story et al., 1998). Prinsippene viser til en ideell målsetting for universell utforming og hvilke egenskaper som er viktige for å oppnå universell utforming i produkter og omgivelser. Konseptet til universell utforming var først ment for fysiske omgivelser, men har senere blitt brukt i sammenheng med design av IKT løsninger (Fuglerud, 2014), som de neste avsnittene tar sikte på å forklare nærmere.

## 2.2.2 Universell utforming i kontekst av IKT

Universell utforming i kontekst av IKT fokuserer på å designe for tilgjengelig løsninger for flest mulig. Det inkluderer et bredt mangfold av forskjellige mennesker, blant annet kjønn, fattige, eldre, kulturer og funksjonshemmede (Berget, Herstad & Sandnes, 2016). I dag er det akseptert at mennesker med funksjonsnedsettelse har behov for å benytte støtteprogram sammen med hovedløsningen. På samme måte som at rullestolbrukere skal kunne ferdes med mobilitet i tilpassede uteområder og i bygninger, skal også IKT-løsninger legge til rette for støtteprogram (Fuglerud, Skotkjerra & Halbach, 2015). FNs konvensjon om rettighetene til mennesker med nedsatt funksjonsevne utarbeidet en definisjon som tar høyde for det, og er kanskje mer nyansert enn Ronald Mace sin definisjon på 1980-tallet:

Universal design means the design of products, environments, programmes and services to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design. Universal design shall not exclude assistive devices for particular groups of persons with disabilities where this is needed. (Forente Nasjoner [FN], 2006, s. 4)



Tidligere var det ikke fremmed at alternative versjoner av nettløsninger ble laget for personer med nedsatt funksjonsevne (Berget, 2016, s. 5). De alternative versjonene hadde som regel et annet design enn det originale nettstedet. Et typisk eksempel er en tekstbasert versjon av en nettside, som trolig førte tilgang til en mengde informasjon for blinde, men førte til andre vilkår for en seende person (Story et al., 1998, s. 71; Fuglerud et al., 2015). Konsekvensen var at blinde ble utelukket i hovedløsningen og seende måtte forholde seg til tilleggsfunksjonalitet de hadde mindre nytte av. Kontrasten er universell utforming som inkluderer et bredt mangfold av mennesker i en og samme hovedløsning.

Det er flere designtilnærminger som har latt seg inspirere for å utforme IKT løsninger som kan brukes av et bredt mangfold av mennesker. Eksempler på det er universell tilgjengelighet, design for alle, barrierefritt design og inkluderende design. Tilnærmingene blir brukt i forskjellige deler av verden på bakgrunn av deres opprinnelse, og de bærer flere likheter med universell utforming. Persson et al. (2015, s. 20) hevder at disse tilnærmingene deler alle det overordnede målet om at IKT løsninger skal være tilgjengelig, brukbart og effektivt for så mange som mulig, uansett hvilke utfordringer brukerne kan møte.

Det er prosessen av å øke tilgjengeligheten i IKT for forskjellige brukergrupper som kalles universell utforming (Sandnes, 2011). Tilgjengelighet i IKT-løsninger innebærer for flere forskere: (1) teknisk tilgjengelighet slik at man har tilgang og kan oppfatte informasjonen til tjenesten, og (2) brukervennlighet slik at informasjonen i tjenesten er intuitiv, effektiv og forståelig (Paddison & Englefield, 2003; Fuglerud, 2014; Sandnes, 2011; Fuglerud et al., 2015).

#### 2.2.2.1 Teknisk tilgjengelighet

Dersom informasjon på nett presenteres i en form som sansene våre kan oppfatte har man det man kaller en teknisk tilgjengelighet. Målet med denne tilgjengeligheten er å sørge for at mennesker med et bredt spekter av muligheter har tilgang til informasjonen og funksjonaliteten til løsningen (Fuglerud et al., 2015; Fuglerud, 2014, s. 43). Den tekniske tilgjengeligheten er forankret i Forskrift om universell utforming av IKT-løsninger (2013) som krever at nettbaserte løsninger rettet mot allmennheten skal være i samsvar med Web Content Accessibility Guidelines 2.0 (WCAG 2.0).

WCAG 2.0 er et dokument med et sett av tekniske anbefalinger som ble utarbeidet av World Wide Web Consortium (W3C) i 1999 som et initiativ for å gjøre netttinnhold mer tilgjengelig for mennesker med nedsatt funksjonsevne (Persson et al., 2015). Dokumentet nevner blindhet, hørselstap, lære- og talevansker, kognitive begrensninger og begrenset bevegelse som noen eksempler på funksjonsnedsettelse. Dokumentet er delt inn i fire prinsipper som tar for seg at netttinnhold skal være oppfattelig, opererbar, forståelig og robust. Hvert prinsipp har 12 tilhørende retningslinjer som videre er delt inn i 61 suksesskriterier, som er mer spesifikke og testbare utsagn. Hvert kriterium har også et nivå som representerer effekten det har på design og den visuelle presentasjonen. Nivåene er A til AAA, der AAA er det høyeste nivået og har den største effekten på designet (Caldwell, Cooper, Reid & Vanderheiden, 2008).

Digitaliseringsdirektoratet har det praktiske forvaltningsansvaret for at nettløsninger følger den overnevnte forskriften. De gjennomfører kontroller og statusmålinger der målet er å avdekke digitale barrierer. I løpet av 2018 ble det til sammen kontrollert 278 private og offentlige nettsider med høyt bruksvolum viser en oppsummering fra Digitaliseringsdirektoratet (2018). Videre viser oppsummeringen at det ble totalt gjennomført 26 818 tester hvor 40% av nettstedene ikke var i samsvar med WCAG 2.0 retningslinjene. Digitaliseringsdirektoratet stiller krav om at nettsider må minst oppfylle 35 av 61 suksesskriterier til WCAG 2.0 på et A - AA nivå. Det er versjon 2.0 av WCAG som kontrolleres i dag. Det er forventet at utgivelsen av WCAG 2.1 i nærmeste fremtid blir å følge ettersom et oppdatert regelverk fra et EU-direktiv behandles i departementet, som stiller ytterligere krav til WCAG.

Den tekniske tilgjengeligheten kan oppnås ved å overholde WCAG 2.0 sine retningslinjer for tilgjengelighet. Dette kan til en viss grad testes ved bruk av både automatiske og ekspert evalueringer (Fuglerud et al., 2015).

Automatiske evalueringer gjennomføres med tilgjengelighetsverktøy. Suksesskriterier som for eksempel kan evalueres med slike verktøy er alternativ tekst på bilder og om fargekontraster er tilstrekkelig på nettsider. Fuglerud og Røssvoll (2012) hadde for eksempel en studie på tilgjengelighet i nettbaserte valgsystemer, der automatiske verktøy ble brukt på markeringsspråk (HTML 5), stil (CSS 3) og tilgjengelighet (WCAG 2.0). Tidligere forskning (Vigo, Brown & Conway, 2013) viser imidlertid at automatiske verktøy som evaluerer

suksesskriterier kan avdekke høyst halvparten av avvik i WCAG som man kan finne ved å gjøre ekspertevaluering. Det kan derfor være nødvendig å supplere automatiske evalueringer med ekspertevaluering.

Ekspertevaluering tar for seg at man manuelt evaluerer tilgjengeligheten i nettløsninger. Slike gjennomganger er ifølge Fuglerud et al. (2015) typisk basert på retningslinjer slik som WCAG, men også heuristikker og personas kan bli brukt, som forklares i kapittel 2.2.2.2. Et eksempel på en manuell evaluering av retningslinjer er å evaluere om tastaturnavigering kan nå all funksjonalitet, og at forstørrelse av tekst ikke mister funksjonalitet. Ekspertevaluering er igjen en manuell analyse av grensesnitt som kan sikre teknisk tilgjengelighet. I likhet med automatisk tester tyder også tidligere undersøkelser at en ekspertevaluering ikke avdekker alle problemene brukere kan ha i løsningen (Power, Freire, Petrie & Swallow, 2012). Videre viser de i undersøkelsen at ekspertevaluering kan finne inntil halvparten av bruksproblemer som man finner gjennom brukertester med personer med nedsatt funksjonsevne.

#### 2.2.2.2 Brukervennlighet

Når man designer produkter ment for menneskelig interaksjon er det sentralt å se på brukervennligheten (Sharp, Preece & Rogers, 2019). Med brukervennlighet referer vi til det engelske ordet «usability», som også omtales som brukskvalitet og brukbarhet (Sandnes, 2011; Språkrådet, u.d.). Standarden International Organization for Standardization (ISO, 2018) definerer brukervennlighet slik: «extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use».

Brukervennlighet er et begrep som beskriver produkter som er effektive å bruke, enkle å lære og gir en behagelig opplevelse sett fra et brukerperspektiv (Sharp et al., 2019). Begrepet blir sett på som et verktøy eller middel for interaksjonsdesignere for å vurdere interaktive produkter og brukeropplevelsen på et mer detaljert nivå. Sharp et al. (2019) argumenterer at det er viktig med et slikt designprinsipp da det ikke er tilstrekkelig å kun spørre om systemet er enkelt å lære eller bruke. Noen prinsipper er med andre ord nødvendig for å lede designet. Nielsen (2012) argumenterer for at brukervennlighet er en forutsetning for å overleve på nettet, for brukere vil forlate nettsiden dersom de opplever den som vanskelig å bruke. Nielsen (2012) beskriver begrepet brukervennlighet grundigere ved bruk av fem kvalitetsattributter:

- Learnability: hvor enkelt er det for brukeren å gjennomføre enkle oppgaver når de tester ut grensesnittet for første gang?
- Efficiency: hvor raskt kan erfarne brukere gjennomføre en oppgave?
- Memorability: hvor enkelt kan brukere returnere til designet etter en periode med inaktivitet.
- Errors: kan brukeren gjøre mange feil uten at disse får konsekvenser for bruken, samt det er enkelt å komme videre selv om feil har blitt gjort.
- Satisfaction: designet bør også appellere til brukeren samt skape en positiv brukeropplevelse.

Før vi kan vurdere om informasjonen som presenteres er brukervennlig må det imidlertid være mulig å oppfatte informasjonen. Dette handler igjen om den tekniske tilgjengeligheten, er det for eksempel mulig å oppfatte informasjonen dersom man har sensoriske eller kognitive nedsettelse? En teknisk tilgjengelighet må derfor være til stede før man kan snakke om det er mulig å forstå informasjonen som presenteres, altså brukervennlighet. Nielsen (2012) kvalitetsattributter handler stort sett om hvordan informasjon på nettet bør presenteres, slik at vi kan forstå informasjonen. Dette er også i tråd med det Fuglerud et al. (2015) hevder, nemlig at brukervennlighet er individuelt basert på hvordan informasjon som presenteres blir forstått. Brukervennlighet og kognitiv tilgjengelighet blir brukt om hverandre av Fuglerud et al. (2015). Man snakker om kognitiv tilgjengelighet fordi mennesker bruker kognitive ferdigheter slik som hukommelse, problemløsning, oppmerksomhet og gjenkjennelse for å forstå informasjon som presenteres.

Det er flere tilnæringer for å sikre brukervennlighet i grensesnittet for mennesker med nedsatt funksjonsevne, blant annet ved ekspertevaluering med heuristikker, personas eller brukertesting (Fuglerud et al., 2015). Heuristisk evaluering tar for seg å undersøke og bedømme overholdelsen av et sett med heuristikker. Nielsen (1994) hevder at heuristikker er ment som en metode for blant annet å evaluere brukervennligheten i løsninger, men ikke retningslinjer for hvordan nettløsningen bør bygges opp. En annen tilnærming for å evaluere brukervennlighet er gjennomgang med personas. Personas utformes basert på informasjon om brukere eller potensielle brukere av løsningen. En persona gjennomgang, innebærer at eksperter forsøker å løse oppgaver fra perspektivet til en persona (Fuglerud et al., 2015).

Både oppfattelse og forståelse, eller mer spesifikt teknisk tilgjengelighet og brukervennlighet, må være til stede dersom man skal kunne bruke løsningen. Dette er en sentral del av universell utforming av IKT (Fuglerud et al., 2015). Fuglerud (2014) hevder at det ikke er tilstrekkelig å kun basere seg på retningslinjene til WCAG 2.0 da kompleksiteten på grensesnittene blir mer informasjonsbelastet og dynamiske, noe som kan gjøre det kognitivt krevende. Det er derfor ikke tilstrekkelig å kun fokusere på teknisk tilgjengelighet for noen brukergrupper, slik som dysleksi.

### 2.2.2.3 Dysleksi og tilgjengelighet

Det er fortsatt et behov for mer forskning om tilgjengelighet på nett for dyslektikere. McCarthy og Swierenga (2010, s. 6) hevder at forskning på dysleksi og tilgjengelighet på nett er mangelfull og bør bli mer forsket på siden bevis indikerer at nettløsninger som øker tilgjengeligheten for dyslektikere også kan komme ikke-dyslektikere til gode.

Støtteprogram som brukes av dyslektikere kan redusere flere utfordringer med skriving og lesing på nett, men ikke alle. For å designe for tilgjengelighet har det blitt utformet flere retningslinjer rettet mot dysleksi, og det inkluderer å gjøre lesbarhet, navigasjon, skriftstil, fargekontrast, tydelig og klart i nettløsninger (The British Dyslexia Association, 2018). Et sett med suksesskriterier i WCAG 2.0 omfatter dysleksi, men de ser ikke ut til å imøtekomme brukere med dysleksi i stor grad (Berget et al., 2016). Forskning på dette området har virket til å handle mer om generelle temaer slik som klartekst, navigasjon og fonter. Men de siste årene har forskning på dette området også tatt for seg tilgjengelighet i søkesystemer (Berget & Sandnes, 2015; 2016), som er noe overførbart til chatbot teknologien vi ser på i dette studiet.

## 2.3 Digital forvaltning

Skatteetaten og Norges Arbeids- og velferdsforvaltning (NAV) er to av de største organene i offentlig forvaltning. NAV har ansvaret for organisering og finansiering gjennom ulike tiltak som arbeid, trygd og sosialhjelp. NAVs hovedoppgave er å bidra til å skape et inkluderende samfunn og arbeidsliv. Etaten skal ivareta vanskeligstilte gruppers behov i tillegg til å sikre inntekt ved arbeidsløshet, svangerskap og fødsel, aleneomsorg for barn, sykdom og skade, uførhet, alderdom og dødsfall (Arbeids- og sosialdepartementet, u.d.). Skatteetatens samfunnsoppdrag er å tilrettelegge for inntekter til det offentlige gjennom å påse at skatter og

avgifter blir fastsatt og betalt på riktig måte. I tillegg har de et ansvar om å oppdatere og forvalte Folkeregisteret som en del av sin funksjon (Finansdepartementet, u.d.).

Digitalisering av forvaltningsorganer slik som Skatteetaten og NAV har pågått over lengre tid. Dette blir ofte kalt digital forvaltning, også kalt e-forvaltning, og omhandler at forvaltningen gjør bruk av IKT (Schartum, Tranvik & Jansen, 2017). IKT på 1950-tallet ble hovedsakelig brukt internt i forvaltningen. Et eksempel er regnemaskinen EMMA som i 1958 ble brukt for å beregne ligningen av skatt i befolkning. Senere ble IKT brukt i forvaltningen for oppbevaring av elektronisk opplysninger, statistisk analyse, tekstbehandling og regneark. Ekstern kontakt med forvaltning ble først mulig via internett på 1980-tallet. Det elektroniske fortollingssystem kalt TVINN var en av de første IKT løsningene for ekstern kontakt i 1988 (NOU 2019: 5, 2019). Etter dette har IKT-løsninger blitt synlig og utbredt i flere offentlige organer, og flere etater kommuniserer nå digitalt med brukerne sine.

Digitalisering blir brukt som et begrep i forbindelse med blant annet (1) overgangen fra analoge til digitale løsninger, (2) videreutvikling av eksisterende digitale løsninger, og at (3) digitale løsninger effektiviserer og endrer måten vi gjør ting på (Dvergsdal, 2019). Det er vesentlige forskjeller mellom de enkelte forvaltningsorganene i graden av digitalisering. Skatteetaten kan trekkes frem som et organ med en relativt høy digitaliseringsgrad (NOU 2019: 5, 2019). Stort sett er sakene de behandler relativt likeartede, og regelverket de forholder seg til egner seg for automatisering. NAV kan trekkes frem som et organ med både høy og lav digitaliseringsgrad. Noen ytelser egner seg for automatisering, for eksempel barnetrygd, pensjonsordninger og foreldrepenger. Derimot bærer sosialhjelp og utenlandssaker ofte et preg av særregler som krever skjønn og personlig oppfølging i NAV, som videre fører til manuell saksbehandling og mindre automatiseringsgrad i forvaltningsorganet (NAV's omverdensanalyse, 2019).

I de fleste tilfeller innebærer digitalisering en omstilling. Schartum et al. (2017) hevder at det lenge har vært kjent at elektronisk databehandling betyr endret organisering av arbeidsprosesser. For eksempel lansering av nettbanker førte til færre bankfilialer, algoritmer har tatt over noe av jobben en aksjemegler gjør og reisebyråer må i stor grad vike for selvbetjening på nett. En omstilling hos forvaltning er også til stede. I Meld. St. 27 (Kommunal- og moderniseringsdepartementet, 2016, s. 29) meldes det at forvaltningsorganer

skal fortsette med en digital selvbetjening, som innebærer for eksempel videreutvikling av digitale løsninger som: rapportering- og søknadsskjema, automatisert saksbehandling og kommunikasjonskanaler og mer. Skatteetaten har for eksempel et mål at brukere skal bli mer selvhjulpne (Andersen, Enger & Knutsen, 2019). De ønsker at dette reduserer antall henvendelser, slik at det frigjøres ressurser de kan bruke på dem som trenger støtte for å følge regelverket. Skatteetaten får hvert år mer enn 2 millioner henvendelser (Andersen et al., 2019). Telefon, skattekontorer og Altinn er størst på mottak av henvendelser til dem. Samtidig ser man at personlig oppmøte på skattekontorer hvert år reduseres. I en treårs periode har besøkene på skattekontor blitt redusert med 40 prosent. Deler av reduksjonen kan være på grunn av nedleggelse av flere skattekontorer. Andre årsaker skyldes at deres største informasjonskanal ut til publikum, det vil si nettsiden deres, har blitt en bedre løsning med bedre informasjon (Andersen et al., 2019). Den samme trenden i besøksnedgang ser man også i andre etater. NAV opplever en besøksnedgang på 10 prosent i løpet av en toårsperiode (NAV i tall og fakta, 2018).

I 2016 redegjør Arbeids- og sosialdepartementet for kanalstrategien. Målet med denne strategien er å bidra til at NAV får flere mennesker av stønad og ut i arbeid. Dette innebærer at åpningstidene på publikumsmottak reduseres og at brukerne skal tas bort fra de tradisjonelle tjenestepunktene som skranker og telefon og over på selvbetjeningsløsninger på nett. Årsaken til disse tiltakene er at NAV ønsker å samle ressursene for å i større grad arbeide målrettet mot de brukerne som har behov for arbeidsrettet oppfølging (Arbeids- og sosialdepartementet, 2016).

Kanalstrategien til NAV kan tenkes å være et tiltak for å møte avbyråkratiseringsreformen, som tilsier 5 prosent reell nedgang i budsjettet til NAV og Skatteetaten fram mot 2030 (NAVs omverdensanalyse, 2019). NAVs omverdensanalyse tar for seg denne problemstillingen og forventer at brukervekst øker, noe som tilsier at offentlig sektor må drives 15 prosent mer effektivt i 2030. NAV forventer å møte denne utfordringen med å øke produktiviteten ved hjelp av digitalisering og selvbetjening.

Samtidig som at brukerne skal bli mer selvhjulpne på nett ønsker offentlig forvaltning at kommunikasjon med innbyggere blir digitalt. I Meld. St. 27 (Kommunal- og moderniseringsdepartementet, 2016, s. 40) heter det at offentlig sektor skal ha et digitalt førstevalg for kontakt med brukere. Et digitalt førstevalg betyr at alle brukere, innbyggere eller næringsdrivende, skal først og fremst kommunisere digitalt med offentlig sektor. For

forvaltningen innebærer dette at de er tilgjengelig på nett og har tilhørende nettbaserte tjenester for digital kommunikasjon). Et digitalt førstevalg har blant annet blitt synlig etter elektroniske meldinger ble hovedregelen i 2014 (Kommunal- og moderniseringsdepartementet, 2014). Det vil si at kommunikasjon fra forvaltning skjer først og fremst på nett og ikke brev lenger. Dersom man ønsker brev kommunikasjon, må man aktivt velge å reservere seg mot elektronisk kommunikasjon.

Videre i Meld. St. 27 (Kommunal- og moderniseringsdepartementet, 2016, s. 40) blir det sagt at løsninger for digital kommunikasjon skal være så gode at de aller fleste faktisk ønsker å benytte disse, og at disse blir valgt fremfor manuelle informasjonskanaler, for eksempel telefon, skattekontorer og NAV-publikumsmottak. For å oppnå dette har den nåværende regjeringen en plan om at offentlige virksomheter iverksetter «radikale endringer med mål om å bedre brukeropplevelser» (Kommunal- og moderniseringsdepartementet, 2019, s. 3). Dette innebærer at offentlige organer sine digitale løsninger skal ha et økt fokus på universell utforming og et tydelig språk i kommunikasjon med deres brukere.

Den nåværende digitaliseringspolitikk i offentlig forvaltning gir uttrykk for å handle om tre overordnede mål: mer effektiv forvaltning, bedre tjenester og en brukerorientert utvikling (NOU: 2019: 5, 2019). Samtidig som de ønsker bedre løsninger og mer effektivitet, skal også offentlig forvaltning satse på systemer med kunstig intelligens (Kommunal- og moderniseringsdepartementet, 2020), såkalte chatboter inngår i denne strategien. Denne teknologien har en høy grad av automatisert kommunikasjon som flere offentlige instanser har begynt å bruke i kommunikasjon med brukerne sine de siste årene.

## 2.4. Chatbot

Dette kapitlet forklarer hva slags chatbot man snakker om i nåværende digitaliseringspolitikk og hvordan disse chatbotene er bygget opp. Det første avsnittet omhandler om hvorfor chatbot har blitt en fremtredende teknologi de siste årene og hvordan teknologien stiller når det gjelder tilgjengelighet. Videre redegjør vi for hvilke chatbot vi ser på i dette studiet og dens tekniske oppbygning sett i et historisk perspektiv.



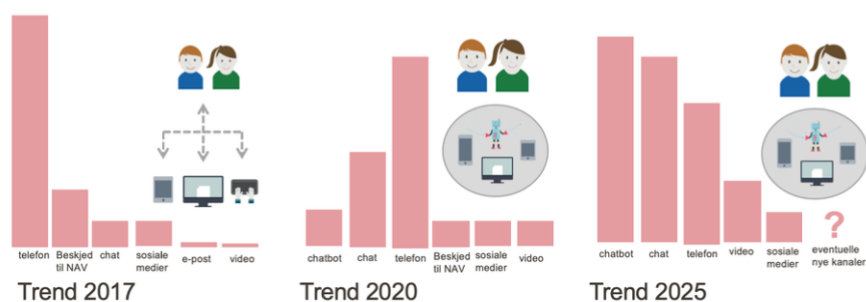
### 2.4.1 Chatbot generelt

En chatbot er et dataprogram som tilbyr kommunikasjon med mennesker i et naturlig språk (Brandtzæg & Følstad, 2018), blant annet i tekst, tale eller begge. Begrepet "chatbot" er bare ett av mange navn som beskriver denne programvaren. På 1990-tallet ble teknologien kalt chatterbotter og i dette studiet forholder vi oss til forkortelsen chatbot som kommer av å ha samtale (chat) med en robot (bot). De første kjente chatbotene har vært gjenstand for forskning og utvikling siden 1960-tallet, men den virkelige gnisten for denne teknologien kom først i 2016, da Microsoft lanserte sin første chatbot og rammeverket ble åpnet for tredjepartsutviklere. Like etter gjorde Facebook det samme på deres meldingsapplikasjon (Brandtzæg & Følstad, 2018). Det gikk bare fire måneder før Facebook hadde 18 000 registrerte chatboter (Dent, 2016), for å sette interessen i perspektiv tok det til sammenligning nesten dobbel så lang tid for Apple å nå slike tall da de lanserte applikasjoner på App Store (Dalrymple, 2009). Den fornyede interessen for chatboter for de store teknologiselskapene ligger i at det har blitt gjort nye fremskritt i kunstig intelligens, maskinlæring og naturlig språkprosessering (Jordan & Mitchell, 2015). For å nå ut til brukere blir chatboter på meldingsapplikasjoner sett på som den neste store teknologien for selskapene (Følstad & Brandtzæg, 2017).

### 2.4.2 Chatbot og universell utforming

Samtidig som chatboter blir mer fremtredende i den digitale verden er kunnskapen om tilgjengeligheten til teknologien begrenset. Lena Drevsjø, ekspert på universell utforming i Funka, sier: «Det er i dag ingen som vet om chatboter følger kravene til universell utforming» (Laurin, 2019). Torres et al. (2018) sin studie som omhandlet å finne litteratur på chatbot og tilgjengelighet spør: «Are Chatbots accessible to blind users? How do these users interact with those interfaces?» (s. 630). De finner imidlertid ingen spesifikk litteratur som gir svar på dette, og hevder at en årsak kan være at det dagens chatbot er et relativt nytt tema. SINTEF forsker Asbjørn Følstad forsterker dette inntrykket. Han forteller til oss at man trenger mer forskning på chatboter når det gjelder tilgjengelighet og universell utforming. Digitaliseringsdirektoratet har per i dag ikke kontrollert om chatbot er en tilgjengelig teknologi som brukes i Norge.

### 2.4.3 Chatbot innen forvaltningen



**Figur 2.2.** Trender for kontakt henvendelser i NAV. Figuren viser ikke størrelser som er reelle. Hentet fra NAV-Kontaktsenter (NKS) presentasjon av T. Simonsen (2019, s. 9).

Chat blir i stigende grad en prioritert kanal for kundeservice, og chatboter blir sett på som et stadig mer relevant supplement til denne kanalen (Følstad, Nordheim & Bjørkli, 2018a). Henvendelser per e-post, telefon og personlig oppmøte på kontor kan være ressurskrevende, ettersom det typisk trenger personlig samhandling med veiledere. En chatbot kan betjene mange brukere samtidig, uansett hvilken tid det er på døgnet. Med innføring av chatboter kan man potensielt senke ressursene og kostnadene, samt gi en mer effektiv og tilgjengelig veiledning (Xu, Liu, Guo, Sinha & Akkiraju, 2017). Et område som derfor virker lovende for offentlig forvaltning er chatbot innen kundeservice. T. Simonsen (2019) i NAV-kontaktsenter ser for seg at chatbottrenden vil øke de neste fem årene, og etter hvert bli den største kommunikasjonskanalen for brukere hos NAV (se figur 2.2).

Chatboter begynner å bli fremtredende i offentlig forvaltning. Det er derfor viktig at forvaltningsorganene fokuserer på universell utforming, som også fremkommer i stortingsmeldingene nevnt i det forrige kapittelet. Skatteetaten og NAV presenterer chatbot som førstekontakt på nettsiden deres og fungerer sånn sett som et digitalt førstevalg for kommunikasjon med brukere. Skatteetatens chatbot har ingen navn, men NAV har kalt sin Frida. Både NAV og Skatteetaten benytter leverandøren Boost.ai sin chatbotløsning. Boost.ai er et skandinavisk programvareselskap som har spesialisert seg på chatbot. De leverer løsninger med kunstig intelligens og er ifølge deres nettside **footnote** basert på maskinlæringsmetoden dyplæring, som de neste avsnittene vil ta for seg grundigere.

#### 2.4.4 Kunstig intelligens

Kunstig intelligens har endret seg med årene, og det har chatbotene også. Kunstig intelligens er et underfelt i informatikk, men hva begrepet konkret innebærer virker det til å være ingen enighet om etter det ble en vitenskapelig disiplin på Dartmouth College-konferansen i 1956 (Russel & Norvig, 2016). McCarthy (1998) som definerte begrepet ett år før konferansen er også i oppfatningen om at det ikke finnes en solid definisjon. Russel og Norvig (2016, s. 2-3) hevder at ulike mennesker legger forskjellig betydning i begrepet, og organiserer begrepet i fire kategorier som er basert på en rekke definisjoner: Kunstig intelligens handler om et systems evne til å enten (1) tenke menneskelig, (2) tenke rasjonelt, (3) handle menneskelig, eller (4) handle rasjonelt.

Chatboter vi ser på i dette studiet er basert på maskinlæring, og i dag vil det som omtales som kunstige intelligente teknologier for det meste være basert på maskinlæringsmetoder (Bratteteig & Verne, 2018).

#### 2.4.5 Maskinlæring

Maskinlæring fokuserer på hvordan man kan bygge datamaskiner som forbedres gjennom sin egen erfaring (Jordan & Mitchell, 2015). Maskinlæring har vært en vitenskapelig disiplin siden 1990-tallet, men det var først et avansement i metodikken i det siste tiåret som førte til at den ble interessant for kommersiell bruk (Jordan & Mitchell, 2015). Den underliggende årsaken er ifølge Holmquist (2017) et gjennombrudd i 2012, da en elektronisk brikke som skulle brukes til å generere hurtig høyoppløselig grafikk på datamaskiner, var den regnekraften man trengte for å behandle data med maskinlæringsmetoden nevrale nettverk.

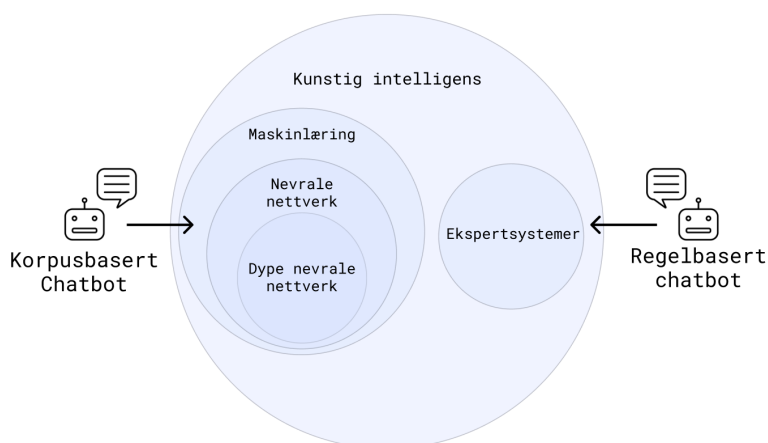
Mekanismene for å lære fra dataene er mønstergjenkjenning basert på statistiske analyser som trenger store datamengder, for å forbedre det statistiske grunnlaget for å gjenkjenne mønstrene i dataen (Bratteteig & Verne, 2018; Holmquist, 2017). Det er hovedsakelig to måter maskinlæringsalgoritmer lærer av data, som påvirker hva slags erfaringer de har når de trener og bygger modellene sine. Enten med (1) en overvåket læring ved at en veileder viser maskinen målet med hva den skal lære for eksempel med en annotasjon i dataen, eller (2) uovervåket læring der algoritmen må lære å gjøre seg kjent med dataene uten veiledning (Goodfellow, Bengio & Courville, 2016).

## 2.4.6 Nevrale nettverk

Såkalt nevrale nettverk er et paradigmeskifte i maskinlæring og er en avansert metode for å lære av data. Nevrale nettverk lages av tilkoblinger av kunstige "nevroner" og er en metafor for det avanserte nevrale nettverket i mennesker eller dyr sine biologiske hjerner (Holmquist, 2017; Goodfellow et al., 2016). Når man utsetter nettverket for stimuli når algoritmen tar det riktige valg blir det mulig å trene nettverket til å ta riktig valg, og dette beskriver en overvåket læring. Med introduksjonen av forsterket regnekraft på 2010-tallet åpnet det seg videre for mer effektive teknikker som dype nevrale nettverk. Dype nevrale nettverk består av mange lag med nevroner mellom input og output i nettverket som brukes for å gjenkjenne flere funksjoner i datasettet som blir brukt (Goodfellow et al., 2016).

## 2.4.7 Chatbot arkitektur

En av de store utfordringene i kunstig intelligens har vært å bygge maskiner med evne til å kommunisere med menneske på et naturlig språk. Blant annet for å overkomme den språklige utfordringen til Alan Turing, kalt Turing Testen. Testen tar for seg å overbevise et menneske om at en maskin kan «tenke», ved at maskinen imiterer et menneske (Turing, 2009). Det er i den sammenheng normalt to arkitekturer for chatbot som brukes, eller har blitt brukt, for å imitere menneskelig atferd, nemlig regel- og korpusbasert arkitektur (Jurafsky & Martin, 2019, s. 491),



**Figur 2.3.** Venn-diagram over chatbot arkitekturer og kunstig intelligens. Basert på Goodfellow, Bengio og Courville (2016, s. 9).

## 2.4.8 Regelbasert chatbot

Dataprogrammer som er regelbasert har en oppførsel som er definert på forhånd. Chatboter med denne tilnærmingen benytter regler for å identifisere både brukerutsagn og systemsvar

(Jurafsky & Martin, 2019). Programmerte regler er typisk logiske kontrollstrukturer (hvis dette – så gjøre dette). Joseph Weizenbaum (1966) programmerte en av de første regelbaserte chatboter herunder navnet ELIZA. Programmet simulerte blant annet en psykolog og i en klientsentrert terapi, og kunne ha samtaler med mennesker, eller pasienter, om ethvert emne. Den baserte seg på programmerte regler, der det «pasienten» skrev inn ble syntaktisk manipulert til et spørsmål, og deretter sendt tilbake til pasienten. Av den grunn trengte ikke ELIZA faktisk kunnskap om emnet for å svare. Flere etterkommere av ELIZA er i større grad basert på domenekunnskap, og disse blir kalt eksperterssystemer. Et eksempel er chatboten Edgar Smith som etterligner en eldre butler i Monserrate palasset i Portugal (Fialho et al., 2013). Edgar sin evne til å svare om palasset ligger i at eksperter som kan en hel del om palasset manuelt har lagt dette inn Edgars kunnskapsbase.

Eksperters sin rolle til å være forbilder for hvordan mennesker resonnerer har imidlertid vært en større utfordring enn man først antok for eksperterssystemer (Bratteteig & Verne, 2018). Menneskets kunnskap om verden er enorm, subjektiv og objektiv, og hvordan man skulle få denne uformelle kunnskap inn i datamaskiner var vanskelig å artikulere på en formell måte. Kunnskapen trenger imidlertid datamaskiner for å oppføre seg intelligente (Goodfellow et al., 2016), noe som den nåværende arkitekturen for kunstige intelligente chatboter er bedre egnet til.

#### 2.4.9 Korpusbasert chatbot

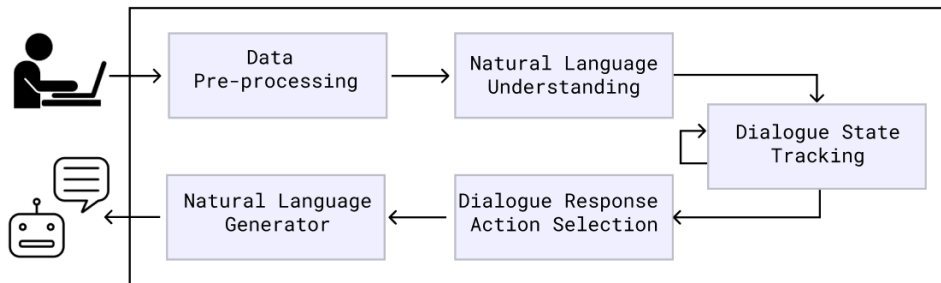
Korpusbasert er en chatbot arkitektur som bygger på maskinlæring hvor data med språklig ytringer er driveren for dataprogrammets svar (Serban, Lowe, Henderson, Charlin & Pineau, 2018). Korpusbasert i kontrast til regelbasert, kan lære fra sin egen erfaring ved behandling av tekstkorpus og lage sine egne regler. I språkvitenskapen blir begrepet korpus brukt som et begrep for en samling av digitale tekster med språklige ytringer (Andersen, 2019). En chatbot som er korpusbaserte er for eksempel Microsoft Xiaoice (Shum, He, & Li, 2018) som er en chatbot ment for generell nettpprat og andre inkluderer for eksempel personlig assistenter slik som Google Assistant og Apples Siri.

Skatteetatens og NAVs chatbot benytter en korpusbasert arkitektur. De er ifølge leverandørens nettside<sup>1</sup>, basert på maskinlæringsmetoden dype nevrale nettverk. De er også

---

<sup>1</sup> <https://www.boost.ai/>

noe regelbaserte, som vi kommer tilbake til senere i dette avsnittet. Med utgangspunkt til korpusbasert chatbotlitteratur forklarer vi videre hvordan chatbotene vi ser på bearbeider tekst og gjør det mulig å respondere sluttbrukeren.



**Figur 2.4.** Språkforståelse komponenter i korpusbaserte chatboter. Basert på Serban, Lowe, Henderson, Charlin og Pineau (2018, s. 3).

Korpusbasert chatbot består av flere språkforståelse komponenter på systemnivå, som illustrert i figur 2.4, og de fleste komponentene kan bli opplært (Serban et al., 2018). Det er såkalte chatbot-trenere som trener og arbeider med chatbotene i Skatteetaten og NAV. Når en bruker sender tekst input til en korpusbasert chatbot, vil først en preprosessering komponent foreta en rekke tekstbehandlingsmetoder for å renske teksten. En rensing reduserer kompleksiteten i teksten, men gjør den lesbart maskinelt og håndterbar for neste komponent. Det er ingen gitt rekkefølge hvordan rensing av tekst gjøres. Thakur (2018) nevner noen metoder som blir benyttet av chatbot leverandøren til Skatteetaten og NAV:

- Tokenisering: omformer setninger til ord
- Språkdeteksjon: finner ut hvilket språk teksten er på
- Stavekontroll: retter opp stavefeil i teksten
- Stoppord filtrering: fjerner ord eller tegn av mindre betydning
- Språkoversettelse: oversetter teksten til engelsk
- Stemming: en algoritme som finner stammen eller rotformen til et ord

Når teksten er rensket vil komponenten «Natural Language Understanding» (NLU) prøve å predikere brukerintensjonen ved å både trekke ut nøkkelord, og andre ord systemer vurderer som viktig (Jurafsky & Martin, 2019). Med en brukerintensjon mener man hvilket mål brukeren prøver å oppnå med teksten som er sendt inn. For eksempel kan målet til en bruker være å finne ut når selvangivelsen kommer eller finne ut hvor man fyller ut søknadsskjema for

foreldrepenger. Før chatboten kan predikere et relevant svar til brukerintensjonen må NLU sine indre modeller bli opplært av maskinlæringsalgoritmer og chatbot-trenere. Noe enkelt forklart er det chatbot-trenere sitt arbeidsområde å trene NLU eller såkalte klassifiseringsmodeller i denne komponenten. En chatbot-trener lager først treningsdata som består av et sett med meldinger med godt formulerte brukerintensjoner som er annotert til å omhandle en intensjon. En annotasjon er for eksempel et nøkkelord i brukerintensjonen. Denne treningsdataen blir deretter sendt til overvåkede maskinlæringsalgoritmer som trener modellen slik at den kan lære seg å predikere intensjonen i treningsdataen. Når en modell har blitt trent på denne dataen lager chatbot-trener testdata som er basert på enkle språkformuleringer funnet i vanlige samtaler. Testdataen sendes videre til dype nevrale nettverk for å teste hvordan modellen predikerer intensjonen som chatbot-trener har arbeidet med (L. Simonsen, 2019; Thakur, 2018).

Når NLU sine indre modeller er opplært til å predikere brukerintensjoner, vil den neste komponenten bestemme samtalen sin tilstand basert på nåværende og tidligere systemresponser. Det vil si at «Dialogue State Tracker» lager en status basert på hva brukeren har uttrykt så langt og hva brukeren sier på nåværende tidspunkt, og prøver å forutsi samtalekonteksten. «Dialogue Response Selection» baserer seg videre på denne statusen for å velge riktig systemrespons (Serban et al., 2018). De fleste korpusbaserte chatbot vil imidlertid gjøre lite modellering av samtalekonteksten, og fokuser heller på å generere en respons basert på den siste ytringen av brukeren (Jurafsky & Martin, 2019, s. 495).

Systemresponsen som chatbotene vi ser på lager ikke egne svar, det vil si at den ikke har en «Natural Language Generator» (se figur 2.4). Som nevnt tidligere er chatbotene også regelbasert, det vil si at eksperter eller chatbot-trenere legger inn forhåndsskrevne svar som chatbotene selv velger mellom. Dersom chatbotene ikke klarer å predikere et sannsynlig svar vil chatbotene også sende et reservesvar som også blir kalt en «fallback» melding.

Korpusbaserte chatboter som nå har blitt forklart i korthet, introduser en systemadferd man tidligere ikke har vært kjent med. Med maskinlæringsalgoritmer lærer chatbotene av erfaring og systemadferden kan derfor være vanskeligere å forstå i motsetning til tradisjonelle regelbaserte chatboter. Det neste kapittelet tar for seg hvordan oppførselen til systemet som er avhengige av kunstig intelligens kan potensielt gjøre seg forståelig for sluttbrukere.

## 2.5 Samtaledesign

Dette kapitlet tar for seg hvordan samtalen kan gjøres forståelig i samtale med kunstig intelligente chatboter. Mer spesifikt omhandler det språklige konsepter som talehandling, samarbeidsprinsippene, tur- og initiativtaking og siste del tar for seg felles enighet som kan være viktig når man designer samtaler i chatbot.

### 2.5.1 Samtaledesign generelt

Human-Computer Interaction har siden 1980-tallet, utforsket design, bruk og interaksjon med datamaskiner med et fokus på det grafiske grensesnittet mellom mennesker og datamaskiner. Design av brukergrensesnitt har en forhistorie fra kommandolinjegrensesnitt på teleprintere (Moore & Arar, 2018). Den gang måtte brukere lære seg å snakke maskinen sitt språk. Premissene er i dag endret til at chatboter kan lære å snakke brukeren sitt språk. Chatboter som baserer seg på naturlig språk er ikke nytt, og har eksistert i mange tiår (Weizenbaum, 1966). Det er fremtredelsen av korpusbaserte chatboter som nå representerer et potensielt skifte fra å studere design, bruk og interaksjon med grafiske brukergrensesnitt til å studere samtalen som objektet for design (Følstad & Brandtzæg, 2017).

«All conversations are interactions but not all interactions are conversations», sier Moore og Arar (2018, s. 4). De hevder at man bør omfavne kompleksiteten av naturlig språk i design av chatboter istedenfor å produsere en simpel interaksjon med den. Først da kan man snakke med en maskin på en naturlig måte. Samtaledesign av chatbot anbefales derfor til å benytte kunnskap om menneskelige samtaleprosesser, som kan føre til at det blir enklere for mennesker å utnytte teknologien. Dette blir utdypt av Jurafsky og Martin (2019) som sier: «Conversation between humans is an intricate and complex joint activity. Before we attempt to design a conversational agent to converse with humans, it is crucial to understand something about how humans converse with each other» (s. 488).

Ved å se på egenskaper ved samtale mellom mennesker kan det gi en struktur for å lage samtaledesign som oppleves menneskelignende, og dermed mer naturlig å interagere med. Følstad og Taylor (2020) hevder at dagens chatbot design er motivert av blant annet talehandling og samarbeidsprinsipper. I tillegg til disse språkkonseptene er turtaking, initiativtaking og en felles enighet også sett på som relevante prinsipper til samtaledesign i



litteraturen (Jurafsky & Martin, 2019). En tilnærming til å lage kunstig intelligente chatboter er å starte med å forstå samtaleprosesser hos mennesker, som blir videre forklart.

### 2.5.2 Talehandling - en ytrer alltid noe med et mål

Filosofer som Wittgenstein, Austin og hans student Searle sitt arbeid innenfor pragmatikken står her sentralt og omhandler hvordan språklige uttrykk gir mening gjennom bruk. Searle (1969) sin talehandlingsteori tar for seg at en ytring er en handling med ord, der språk blir sett på å være en handling like mye som en fysisk handling. Hver tur i en samtale, eller hver melding i en chatbot, kan ses på som en handling. En som taler ytrer alltid noe med et mål i tankene. Det kan for eksempel være ordre, påstander, klager, beskrivelser, spørsmål og løfter med flere (Svendsen, 2018). Disse handlingene blir kalt talehandlinger og det finnes ulike klassifiseringer av dette. For eksempel Searle (1976) sine fem følgende kategorier når en taler ytrer seg:

1. Assertives: uttaler seg om tilstanden i verden (foreslår og konkluderer)
2. Commissives: forplikter seg til en handling (lover og planlegger)
3. Directives: prøver å få lytteren til å utføre en handling (spør og inviterer)
4. Expressives: uttrykker sine synspunkter om ting i verden (takker og unnskylder)
5. Declarations: endrer tilstanden i verden ved hjelp av ytringen (du har sparken)

Searle (1976) sine kategorier er relevant både for hvordan chatboten predikerer og gjør seg forstått til brukeren. En chatbot som ikke er tilstrekkelig trent vil antageligvis misforstå hva målet til taleren er. Det er en nødvendighet at chatboter klarer å predikere brukerintensjoner for at brukere skal kunne oppnå sine mål i samtalen. Alle kategoriene er ikke nødvendigvis relevante for chatbot innen kundeservice, men å foreslå, spørre eller takke er ganske konvensjonelle talehandlinger innenfor dette området. Det er også nødvendig at chatboten gjør seg forstått for brukeren. Det er som tidligere nevnt chatbot-trenere som lager predefinerte svar som chatboten benytter for å svare brukeren, og Searle sine kategorier kan være et godt utgangspunkt for disse svarene.

### 2.5.3 Samarbeidsprinsippene – en vellykket samtaleprosess

Samarbeidsprinsippet tar for seg at man skal opptre samarbeidende for å etablere en vellykket samtaleprosess. Filosofen Grice (1975) introduserte dette prinsippet og delte det inn i fire mer spesifikke samtalemaksimer:

- Maxim of Quality: innebærer si hva du mener er sant
- Maxim of Quantity: innebærer å si så mye informasjon som er nødvendig, men ikke mer enn nødvendig
- Maxim of Relation: innebærer å snakke om hva som er relevant for samtalen
- Maxim of Manner: innebærer å være tydelig og forklare på en måte som gir mening for andre

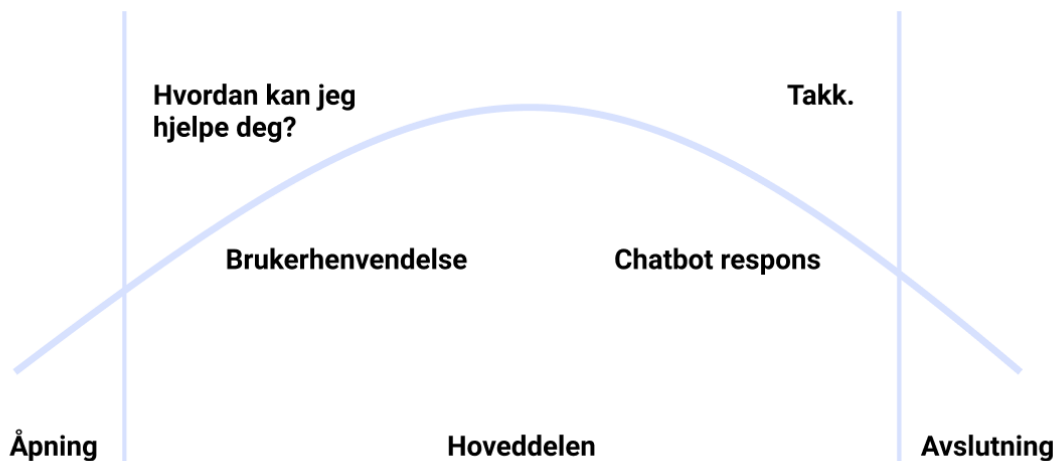
Grice (1975) sin teori handler om hvordan mennesker bruker språk i kommunikasjon. Grice var ifølge Falkum og Kjöll (2014) den første som hevdet at samtaler er en slutningsbasert prosess. Som betyr at den som lytter til en som taler lager antagelser om talerens hensikt med ytringen. Når en samtale utreder seg mellom mennesker antar lytteren at ytringen generelt vil være sann, ha rett mengde med informasjon, være relevant og gi mening. Dersom denne modellen ikke samsvarer med forventningene, vil lytteren prøve å finne sannhet i det taleren formidlet (Falkum & Kjöll, 2014). Dette er en forklaring på samarbeidsprinsippet, og har flere likheter med det Norman (1988) tenker er den vanskelige delen i design, det vil si når ting går galt:

If a person says something that is not understandable, we ask for clarification. If a person says something that we believe to be false, we question and debate. We don't issue a warning signal. We don't beep. We don't give error messages. We ask for more information and engage in mutual dialogue to reach an understanding. (Norman, 1988, s. 198)

På samme måte som mennesker blir forvirret eller frustrert når maksimene ikke følges, kan chatboter også forårsake lignende problemer. Mennesker er godt vant med en rekke sosiale normer. For eksempel å hilse, er en av de vanligste uskrevne reglene. Chatboter er ikke

mennesker, men det kan være fordelaktig at de følger grunnleggende sosiale konvensjoner for at mennesker skal forstå dem. De fire maksimene til Grice kan være et godt utgangspunkt for å gjøre seg forstått.

#### 2.5.4 Turtaking - hvem sin tur er det å snakke?



**Figur 2.5.** En kanonisk interaksjonsstruktur for chatbot innen kundeservice. Basert på Szymanski og Moore (2018, s. 21).

Noe som kan virke trivielt i samtaleprosesser er turtaking. Det er en beskrivelse av hvordan turer fordeles mellom samtalepartnere. En samtale vil oppleves naturlig dersom man tar turen på det riktige tidspunktet, slik at det ikke forekommer overlapp, taushet eller andre anormale pauser i samtalen. Det kan som sagt virke åpenbart, men det krever en analyse av en rekke turtakingssignaler, blant annet tid, tonenivå og kroppssignaler må betraktes (Levinson & Torreira, 2015). Dette gjør tur-taking sammensatt, men samtaler som baserer seg på tekst er noe begrenset når det gjelder dette. Både tone og kroppsspråk er fravikende i tekstbaserte chatboter.

Samtaler innen kundeservice har to bestemte roller. Det består av en bruker som forespør en tjeneste og den andre parten i samtalen, en veileder, leverer tjenesten på vegne av organisasjonen. Det er normalt at veileder starter samtalen med en åpning, vanligvis en hilsen, identifiserer seg selv, identifiserer organisasjon og åpner opp for tilbydelse (Moore & Arar, 2018). En vanlig tilnærming innen chatbot for kundeservice er derfor å la chatboten ta turen ved åpningen av samtalen og etter hver brukerhenvendelse den mottar (som vist i figur 2.5)

### 2.5.5 Initiativtaking - hvem styrer samtalen?

En talehandling som fører til at noe inntreer sees på som et initiativ. En som styrer samtalen på et tidspunkt er en initiativtaker. En initiativtaker kan starte en samtale og bestemme hvor samtalen skal gå. Hvor mye samtalen styres, avhenger av hvor mye initiativ som blir tatt. Noen ganger kan man oppleve at samtalen blir kontrollert av en initiativtaker som i for eksempel et intervju. Dette er imidlertid ikke normen i menneskelige samtaler. Det er vanligvis normalt at partene i en samtale skifter på hvem som tar initiativet og slike interaksjoner blir kalt mixed-initiativ (Jurafsky & Martin, 2019). Passive initiativ interaksjoner som system-initiativ og user-initiativ har tradisjonelt vært enklere å bruke i systemer, siden de kun responderer på det brukeren spør om. De må dermed ikke forholde seg til kompleksiteten av initiativ interaksjoner slik som i mixed-initiative. Et system som tar full kontroll i samtalen, og godtar kun gyldige svar, er system-initiativ. På den andre siden av dette finner man user-initiativ tilnærmingen, som er i hovedsak styrt av brukeren (Jurafsky & Martin, 2019).

Chatbot innen kundeservice er normalt basert på user-initiative, det vil si at det er bruker som oppretter kontakt for å be om svar på en brukerintensjon. Det er chatboten sin oppgave å identifisere brukeren sin intensjon, tilby mulige løsninger og muligens iverksette handlinger. Muligheten til å iverksette handlinger er imidlertid noe som typisk mangler i disse chatbotene siden de ikke er integrert med kunderelasjonsprogrammer (Følstad, Skjuve & Brandtzæg, 2018b). Chatbotene vi undersøker har ikke dette integrert, og de er stort sett basert på user-initiativ, men de er også noe mixed-initiativ siden de starter samtalen og spør av og til om brukeren er tilsted.

### 2.5.6 Felles enighet - bygge et felles grunnlag

En samtale er en kollektiv handling som utføres sammen mellom en taler og en lytter. Som alle andre handlinger som gjøres i felleskap er det viktig å etablere hva begge partene er enige om, og dette blir kalt felles enighet eller «common ground» (Jurafsky & Martin, 2017). En felles enighet oppnås ved at taleren anerkjenner at den som lytter forstår og godtar det taleren snakker om. Mennesker erkjenner hverandres ytringer hele tiden, blant annet med bekreftende eller avkreftende svarord eller ved å repetere det lytteren har tidligere sagt (Jurafsky &

Martin, 2017). En slik erkjenning er ikke kun forbeholdt til språklige handlinger, men også ved menneske-maskin interaksjon og andre interaksjoner. Når mennesker interagerer med maskiner erkjenner maskiner ved å gi tilbakemeldinger, og når maskinen ikke gir tilbakemeldinger kan resultatet til handlingen ende med flere misforståelser (Norman, 1988). For å forhindre at chatbotsamtaler ikke ender med misforståelser er det derfor viktig at de anerkjenner brukeren med svarord eller repetere det som blir sagt.

## 3. Metodologi

I dette kapitlet redegjøres studiens metodologiske tilnærming og filosofiske antagelser. Først vil vi ta for oss forskningstilnærmingen og forskningsstrategien. Deretter blir den overordnede forskning- og designprosessen forklart nærmere. Beskrivelser av valgte metoder og hvordan de ble brukt til å samle inn data blir videre forklart, og til slutt redegjør vi for etiske betraktninger som er gjort i løpet av forskningsprosessen.

### 3.1 Forskningstilnærming

All forskning opererer ut fra paradigmer eller underliggende filosofiske antagelser om hva som innebærer gyldig forskning og hvilke forskningsmetoder som er passende (Myers, 1997). I forskning av informasjonssystemer er det vanlig å posisjonere seg innenfor følgende filosofiske perspektiver: positivistisk, fortolkende og kritisk (Myers, 1997). Dette studiet baserer seg på det fortolkende paradigme. Fortolkende forskning baserer seg generelt på antagelsen om at mennesker har deres egen subjektive mening av verden, og meninger mellom mennesker og forskerne utgjør en intersubjektiv sannhet. Med en fortolkende tilnærming er vi ikke ute etter et forskningsresultat med en absolutt sannhet slik som vurderes mulig i det positivistiske paradigmet, men heller en subjektiv sannhet. Denne tilnærmingen gir oss en tilgang til en virkelighet som kun kan komme gjennom meningsinnholdet vi undersøker, som for eksempel språklige ytringer, delte meninger og bevissthet (Myers, 1997). I dette studiet utforsker vi dyslektikere sine opplevelser med chatboter, der vi prøver å forstå og beskrive opplevelsen gjennom betydningene deltakerne tillegger meningsinnholdet. Dermed kan vi i henhold til Myers (1997) si at dette studiet befinner seg innenfor det fortolkende paradigmet.

### 3.2 Brukersentrert forskningsstrategi

At vi betegner dette studiet under det fortolkende paradigme har betydning for hvilke metodologi og metoder vi velger, og hvordan vi når frem til et forskningsresultat. Metodologier er forskningsstrategier eller en oppskrift som veileder forskning. I dette studiet er brukersentrert design valgt som forskningsstrategi. Brukersentrert design ble introdusert av Norman på 1980-tallet og er et konsept innen forskningsfeltet for Human-Computer Interaction (Abrams, Maloney-Krichmar, & Preece, 2004). Brukersentrert design fremstiller en

designfilosofi med en hensikt om å lage og teste produkter med fokus på brukernes behov, i motsetning til at brukere endrer oppførsel og tilpasser seg produktene (Abrams et al., 2004). Noen funksjonsvansker ved bruk av chatbot for mennesker med dysleksi kan løses med hjelp av støtteprogram og trening, men utfordringer forårsaket av mangelfull design bør heller løses ved å endre dataprogrammet, som er i henhold til den brukersentrerte designfilosofien (Berget, 2016)

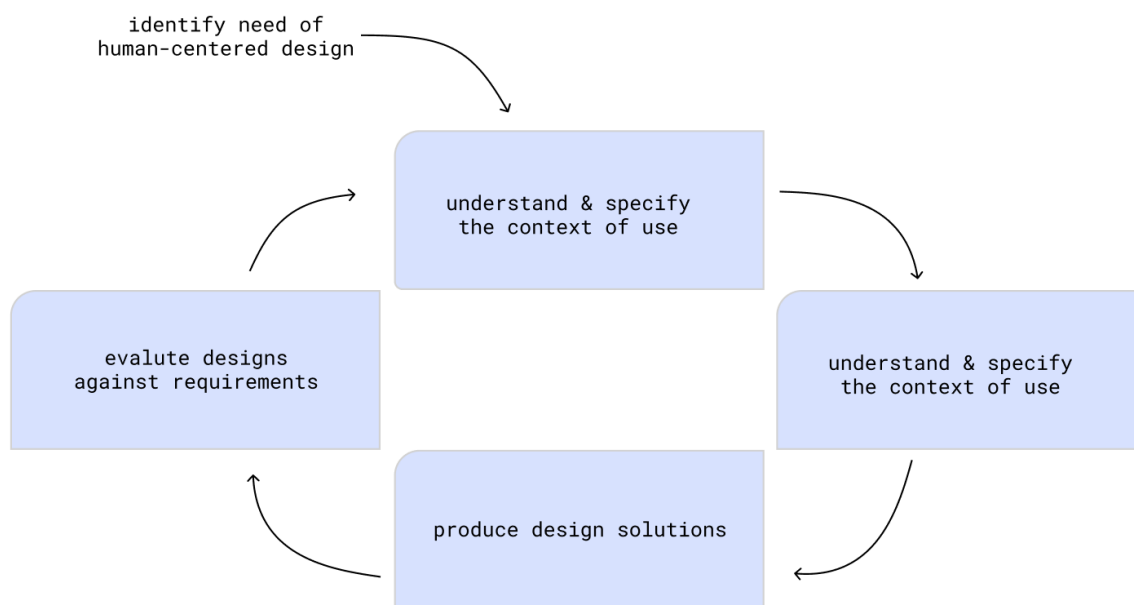
Når feltet for Human-Computer Interaction ble etablert, utarbeidet Gould og Lewis (1985) tre prinsipper for en brukersentrert designtilnærming, og prinsippene er (1) et tidlig fokus på brukere og oppgaver, (2) empiriske målinger, og (3) iterativ design. Med det første prinsippet vektlegges det å forstå hvem brukerne er, som innebærer å involvere og observere brukere tidlig i designprosessen. Det andre prinsippet handler om å la brukere, eller potensielle brukere, interagere med prototyper eller simulering av systemet som utvikles, og observere samt analysere deres ytelse og reaksjoner. Deretter bør systemet redesignes i forhold til problemer knyttet til brukervennlighet som blir identifisert. Det siste prinsippet om iterativ design tar for seg å utforme prototyper, brukerteste, og redesigne i en syklisk prosess så lenge dette er nødvendig (Sharp et al., 2019, s. 48).

Det er forskjellige perspektiver innen brukersentrert design, for eksempel mainstream design, design av systemer eksklusivt for brukere med funksjonshemninger eller universell utforming (Berget, 2016). Vi tar sikte på å undersøke universell utforming perspektivet, med et fokus på å øke tilgjengeligheten gjennom en hovedløsning. Innen universell utforming perspektivet er det særlig et element i brukersentrert design som kommer til uttrykk for å være anbefalt av flere forskere. Nemlig å evaluere med brukere med nedsatt funksjonsevne for å sikre at løsningen er brukervennlig (Fuglerud, 2014, s. 13).

En kilde til inspirasjon for vårt valg av forskningsstrategi er Fuglerud (2014) sin doktoravhandling «Inclusive design of ICT: The challenge of diversity». Avhandlingen tar for seg en rekke prosjekter innen universell utforming hvor forskningsstrategien er brukersentrert design komplementert med brukertesting og empiriske evalueringer. Prosjektene inkluderer personer med ulike funksjonsnedsettelse, både på eksisterende løsninger og prototyper (Fuglerud, 2005; Fuglerud, Reinertsen, Fritsch & Dale, 2009; Fuglerud, 2011). En lignende tilnærming anbefales av en rekke forskere. Paddison og Engefield (2004) hevder for eksempel: «It is not enough to follow accessible guidelines and make the appropriate

technical accessibility changes. People with special accessibility needs have a distinct user profile with unique requirements that need to be addressed within a user-centred design process» (s. 508).

Begnum (2016) identifiserer flere metodologiske tilnærminger brukt i universell utforming: brukersensitiv inkluderende design, inkluderende design og deltakende design. Dette studiet tar for seg en brukersentrert design metodikk der vi designer for brukere, noe som skiller seg fra den sistnevnte metodikken deltakende design der man i større grad designer med brukerne (Sanders & Stappers, 2002).



**Figur 3.1.** Standardmodellen ISO 9241. Basert på Fuglerud (2014, s. 37).

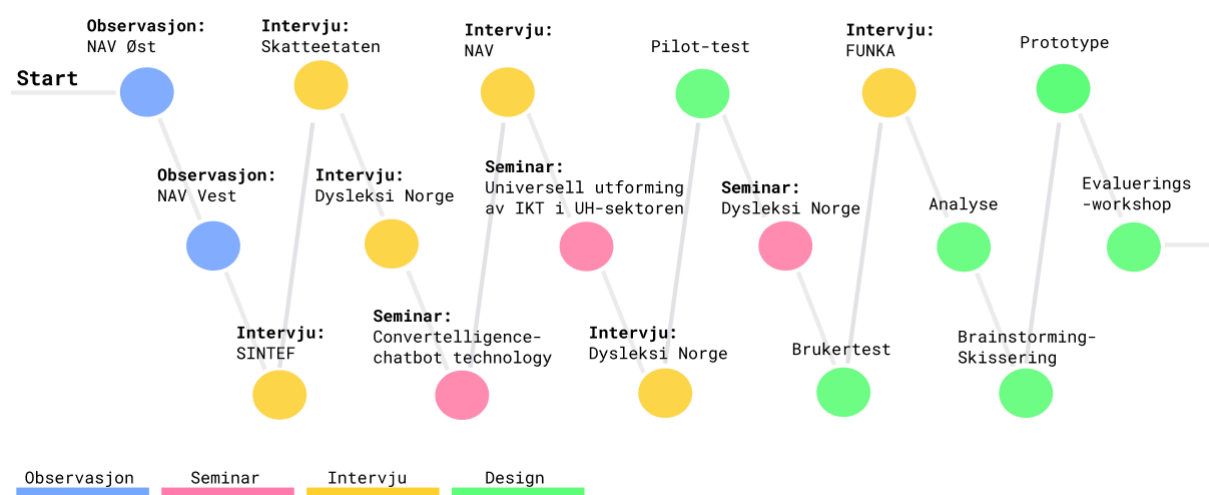
En brukersentrert designprosess har typiske faser som man finner i såkalte livssyklusmodeller. En typisk brukersentrert designprosessen tar for seg fire aktiviteter: (1) oppdage krav, (2) designe alternativer, (3) prototype, og (4) evaluere (Sharp et al., 2019, s. 52). Dette er aktiviteter som også inkluderer de tre prinsippene for brukersentrert design av Gould og Lewis (1985) som forklart tidligere. Aktivitetene i prosessen er til for å bli repetert mange ganger fordi de informerer hverandre. For eksempel kan det være nødvendig å bearbeide brukerkrav etter evaluering. Nøyaktig hvordan prosessen skjer varierer fra prosjekt til prosjekt, men utføres repetitivt inntil løsningen oppfyller krav som kan for eksempel omhandle kriterier om brukervennlighet eller brukeropplevelse (Sharp et al., 2019). Det er flere livssyklusmodeller som har blitt utviklet for dette, tre kjente modeller er Star model, Google Design Sprints og standardmodellen ISO 9241-210 som vises i figur 3.1.



Som en oppsummering baserer vi dette studiet på det fortolkende paradigmet der vi har valgt en brukersentrert designtilnærming for å produsere svar på forskningsspørsmålene vi stiller.

### 3.3. Metode og designprosess

Vi vil følgende utrede hvilke metoder som har blitt brukt og hvordan disse har blitt brukt i forskning- og designprosessen. Rekruttering og etiske betraktninger med tanke på metoder og datainnsamling blir gjennomgått avslutningsvis.



**Figur 3.2.** En oversikt over den overordnede forskning- og designprosessen i dette studiet.

Studiet startet med en divergerende designprosess med observasjon, som vist i figur 3.2. Begrepet divergerende beskrives av Löwgren og Stolterman (2004, s. 29) som en prosess der designere utvider sine ideer til å finne alternativer, utforske muligheter og spørsmål. Mer spesifikt betyr det at mer informasjon, alternativer og spørsmål blir skapt i denne prosessen enn det man får svar på. Konvergerende er det motsatte av divergering, og tar for seg å fokusere på en spesifikk løsning av flere ideer skapt fra den divergerende prosessen (Löwgren & Stolterman, 2004). Vi mener at den tidlige designprosessen kan i henhold til Sanders & Stappers (2008) kjennetegnes som en «Fuzzy-front-end». Begrepet referer til at den tidlige designprosessen, også kalt front-end, er både tvetydig og kaotisk fordi prosessen typisk inneholder aktiviteter som undersøker åpne spørsmål, men hva som undersøkes er ikke konkret bestemt. Den tidlige designprosessen kan beskrives som aktiviteter som har informert og inspirert oss i letingen til det spesifikke området vi til slutt undersøkte.

Senere i studiet hadde vi en konvergerende prosess der vi beveget oss fra det store og de mange ideene, til å studere universell utforming av chatbot. I den sammenheng hadde vi intervjuer med eksperter på chatboter, universell utforming og dysleksi. I tillegg deltok vi på flere seminarer og presentasjoner i forbindelse med disse temaene. Den brukersentrerte designprosessen i dette studiet har tatt for seg å (1) tidlig involvere deltakere, (2) brukerteste eksisterende chatboter, (3) analysere resultatene, (4) designe prototyper av designforslag, og (5) evaluere prototypene med deltakere. Livssyklusmodeller i brukersentrert designtilnærming har flere elementer som passer vårt prosjekt, men slike modeller tar utgangspunkt at systemet ikke er et sluttprodukt. Normalt ville man for eksempel startet med en behovsanalyse og samle inn brukerkrav, men siden vi starter med å brukerteste eksisterende systemer, begynner vi tidligere på en evaluering enn man vanligvis ville gjort med tanke på livssyklusmodellen (se figur 3.1).

### 3.3.1 Observasjon

Hatch (2002, s. 72) hevder at målet med observasjon er å forstå omgivelser, kulturer eller å se verden gjennom øynene til de man undersøker. Som en utforskende prosess valgte vi å gjøre feltobservasjon i to av NAVs publikumsmottak i Oslo. Denne metoden ble brukt for å samle innsikt om hvordan omgivelsene er utformet og hvilke handlinger som foretas i omgivelsene. Hvor mye man involverer seg i observasjon varierer i forhold til hva man ønsker å undersøke, for eksempel kan man være passiv, moderat eller en aktiv observatør (Hatch, 2002). Siden vi observerer i et offentlig rom uten å direkte involvere oss mer enn vår tilstedeværelse, betrakter vi oss som passive observatører som i liten grad påvirker flyten av hendelser.

#### **Gjennomføring**

Observasjon som metode ble brukt innledningsvis i dette studiet da vi søkte etter en problemstilling vi kunne se nærmere på. Målet med metoden var å utforske temaet digital forvaltning. For å få et innblikk i hva klienter foretar seg og hvordan omgivelsene er utformet på fysiske NAV-publikumsmottak, valgte vi å være passive observatører på to mottak, henholdsvis øst og vest i Oslo. Vi satt oss ned på benker og stoler med god oversikt av lokalene i ca. 1 time, men noterte ikke for vi ønsket å unngå oppmerksomhet rundt vår tilstedeværelse. Etter observasjonene lagde vi enkle utkast av lokalene, diskuterte og noterte hva vi hadde sett og opplevd. Noen av inntrykkene våre ser vi fortsatt som relevant og presenterer disse videre.

Vi la merke til flere likheter enn ulikheter ved mottakene. De hadde begge til felles at man ble møtt av flere datamaskiner når man entret mottaket. Disse ble hyppig brukt da vi observerte, og vi antar at de trolig blir brukt i sammenheng med å gå inn på NAVs nettside. Denne antagelsen gjorde vi ettersom vi var på deres publikumsmottak, men også fordi plakater og TV-skjermer ved mottakene foreslo å benytte nettstedet for selvbetjening. Andre likheter var at mottakene hadde automat for kølapp, skranker, kontorer, utskriftsmaskiner og sitteplasser.

Ved mottaket på øst så det ut til at det var noe ventetid for å få hjelp ved skrankene og fra det vi observerte var det bare en ansatt som betjente. På øst så vi også at skjemaer, brosjyrer og andre papir lå noe utilgjengelig og lengre inn i lokalet. Vi observerte at papirene ved begge mottakene lå rotete til i en hylle og dette antydte for oss at dette var et område som ikke skulle være særlig tilbydende for de besøkende.

I ettertid av observasjonene hadde vi generert nye inntrykk når det gjelder en digital forvaltning. Besøkene forsterket vårt inntrykk om at NAV prøver å få deres brukere mer selvhjulpne digitalt. Dette var en av de overordnede årsakene til at vi nettopp begynte å skifte retning mot chatbot brukt i forvaltningen.

### 3.3.2 Ekspert intervju

Intervju er en metode som ofte blir brukt i anledninger der man ønsker å gå i dybden. Innen fagfeltet til brukersentrert design brukes det blant forskere for å bygge forståelse om behov, preferanser, bekymringer og holdninger hos de man intervjuer (Lazar, Feng & Hochheiser, 2017, s. 189). Før et intervju bør man ta stilling til hvilken struktur intervjuet skal ha, og her kan man velge mellom strukturert, semistrukturert og ustrukturert. I dette studiet har vi forholdt oss til semistrukturerte intervjuer med forhåndsskrevne intervjuguider med åpne spørsmål. Siden vi ikke har erfaring og domenekunnskap om utvalget vi intervjuer, anser vi en slik tilnærming som passende. I kontrast til strukturerte intervjuer åpner en semistrukturerte tilnærming opp for fleksibilitet til å be om avklaringer, stille tilleggs spørsmål og få utbroderende svar fra intervjuetobjektet (Lazar et al., 2017).

Datamaterialet som blir generert er muntlig informasjon som er forankret i fortolkninger av de vi intervjuer. Intervjuene har for det meste blitt gjort med opptak av lyd, med et par tilfeller

hvor vi kun tok notater. Ved notattaking er det utfordrende å få med all informasjon, og noe kan bli oversett som følge av dette. Ved å benytte opptak av lyd, kunne vi i større grad være til stede i gjennomføring av intervjuet uten at faktorer som skriving tok vekk oppmerksomheten. Det er også en kost-nytte-verdi av lydopptak. Transkribering er en tidkrevende prosess, i tillegg til at lydopptak kan virke stressende eller pressende for intervjuobjektet, og kanskje spesielt når man snakker om personlige erfaringer med dyslektikere.

## **Gjennomføring**

Totalt har vi gjennomført 11 intervjuer, hvor 5 av intervjuerne var en del av brukertesting. De formelle intervjuene har vært med Dysleksi Norge, Skatteetaten, NAV, SINTEF og Funka (se figur 3.2). De involverte personene har domenekunnskap på universell utforming, chatboter og dysleksi.

Vi har intervjuet seniorforsker Asbjørn Følstad ved SINTEF. Følstad har en bakgrunn fra psykologi og har bidratt i mange forskningsprosjekter som ser på brukersentrert utvikling i digitale løsninger som chatboter. I tillegg har vi intervjuet en ekspert på universell utforming, Kristian Munter. Munter har nylig hatt et samarbeid med organisasjonen Funka om et prosjekt som gransket hvordan noen utvalgte implementasjoner av chatboter fungerer for mennesker med funksjonsnedsettelse. Vi har intervjuet avdelingen for brukskvalitet i Skatteetaten og avdeling for universell utforming i NAV. Vi har også intervjuet representanter fra Dysleksi Norge for å få en bedre forståelse om hvordan det er å leve med dysleksi og da spesielt i et digitalt samfunn.

Intervjuene med eksperter foregikk på deres arbeidsplass, mens de resterende intervjuene ble gjort på Institutt for informatikk ved Universitetet i Oslo. Alle ekspertintervju hadde en varighet på ca. 1 time. Vi utformet intervjuguiden til hvert intervju, og disse guidene ble oppdatert med temaer som oppsto underveis som vi synes var interessante og viktige å utforske videre. Foruten de to første intervjuene ble samtlige tatt opp med lydopptaker. Vi opplevde det å notere uten opptak var en distraksjon som forringet kvaliteten på intervjuet, og valgte derfor å ta opptak samt transkribere kort tid etter utførelse.

Intervjuene kan deles opp i tre faser:

1. I første fase intervjuet vi eksperter da vi ønsket å få kunnskap om chatboter, universell utforming og den nåværende situasjonen til digitaliseringsprosjektet av etatene.
2. I andre fase ble domeneeksperter på dysleksi fra Dysleksi Norge og medvirkende i Funkas prosjekt vedrørende «Universell utforming av chatbot» intervjuet. Dysleksi Norge kunne tilføre oss nyttig innsikt i hvordan det er å ha kognitive utfordringer som skrive- og lesevansker samt demonstrere typiske digitale støtteprogram for dyslektikere og Funka oppsummerte resultatene fra deres studie og kom med innsikt vedrørende chatboter og universell utforming.
3. I siste fase hadde vi intervjuer i forbindelse med brukertesting. Her fikk vi personlige erfaringer fra ulike typer utfordringer man kan ha med skrive- og lesevansker og vi fikk en bedre forståelse for individuelle behov. I tillegg var det nyttig som kartlegging av hvordan de manøvrer seg i den digitale verden ved hjelp av støtteprogrammer og teknikker.

### 3.3.3 Automatisk evaluering

Før man involverer personer med funksjonsnedsettelse i en brukertest anbefales det av Fuglerud et al. (2015) at man utfører en automatisk evaluering av tilgjengeligheten i brukergrensesnittet. Dette gjør man for sikre at grensesnittet kan brukes av personer med funksjonsnedsettelse, og at det er kompatibelt med støtteprogram. En automatisk evaluering kan til en viss grad evaluere om grensesnittet har en teknisk tilgjengelighet med tanke på WCAG 2.0. En teknisk tilgjengelighet er som nevnt at man har tilgang til å bruke løsningen (se kapittel 2.2.2.1). Det finnes en rekke automatisk testverktøy for eksempel Achecker, Siteimprove og axe. Siteimprove<sup>2</sup> var programmet som vi benyttet for å kontrollere chatbotene, og dette verktøyet er tilgjengelig som en utvidelse i nettleseren Google Chrome.

### Gjennomføring

Før brukertesten gjennomførte vi en automatisk evaluering av chatbotene for å få en indikasjon om det var en tilgang til bruk for dyslektikere. Chatbotene ble testet isolert fra nettsiden i Google Chrome. Dette ble gjort ved å fjerne alle HyperText Markup Language (HTML) elementer som ikke hadde med chatbotene å gjøre i nettleserkonsollen. Da vi utførte testen fikk vi noen indikasjoner på at det var en tilstrekkelig brukstilgang for dyslektikere. Erfaringer fra den automatiske evalueringen kommer vi tilbake til i diskusjonskapittelet.

---

<sup>2</sup> <https://siteimprove.com/>

### 3.3.4 Brukertesting

En velkjent metode for å evaluere brukervennlighet, tilgjengelighet og brukerbehov i et grensesnitt er brukertesting. Ifølge Dumas og Redish (1999, s. 22) er det mange forskjellige måter man kan brukerteste, men alle metodene deler ofte disse fem karakteristikene:

- Hovedmålet er å forbedre brukervennligheten
- Deltakerne er virkelige brukere
- Deltakerne gjør oppgaver som reflekterer virkeligheten
- Man observerer og noterer det brukeren sier og gjør
- Man analyserer dataene, finner problemområder og kommer med løsninger for å fikse disse

Et annet mål er å forbedre prosessen i hvordan produktet som testes blir designet og utviklet, så ikke problemområdene man finner gjentar seg. I tillegg må man også se på hva som allerede fungerer bra, så man ikke fjerner optimal funksjonalitet i et eventuelt redesign (Lazar et al., 2017, s. 264).

Brukertesting er vanlig å benytte for å teste prototyper eller utviklede grensesnitt.

Grensesnittene som testes er ofte skjermer enten på datamaskin, mobil eller nettbrett (Lazar et al. 2017, s. 263). Det er mange ulike stadier og måter man kan anvende brukertesting av et grensesnitt, for eksempel:

- Lavoppløselige prototyper og papirprototyper
- Høyoppløselige og klikkbare prototyper
- Ferdigutviklet grensesnitt før implementering og levering
- Grensesnitt i eksisterende systemer

Ifølge Nielsen (1993, s. 165) er brukertesting en fundamental metode for å få en forståelse av hvordan brukerne anvender brukergrensesnittet, og hvilke konkrete problemer som oppstår ved interaksjon. Det skilles ofte mellom to evalueringstyper når man skal se på et grensesnitt: (1) formativ evaluering, og (2) summativ evaluering. Formativ evaluering ser for eksempel på hvilke aspekter i et grensesnitt som fungerer og hva som ikke fungerer, og hvorfor det er slik. Det finnes flere typer formative evalueringsmetoder, for eksempel heuristikker og kognitive walk-through, men den mest brukte og vanlige formative evalueringsmetoden er brukertester (Nielsen, 1993, s. 170). Ved brukertesting lager man oppgaver som brukere skal gjennomføre,

mens det ofte oppfordres til å tenke høyt mens det observeres hva de gjør og hva de sier. Summativ evaluering sammenligner i større grad metrikker fra tidligere design eller konkurrerende løsninger der en typisk metode ville ha vært i form av målinger som tid eller klikk (Joyce, 2019). Vi er derimot mer interessert i de rike beskrivelsene fra utvalget, og har derfor valgt formativ evaluering av grensesnittet.

Vår brukertest fokuserer som nevnt på eksisterende chatbotgrensesnitt hos etatene NAV og Skatteetaten. Vårt hovedfokus var å se på brukervennlighetsbarrier og andre tilgjengelighetsproblemer dyslektikere kan støte på ved interaksjon med disse.

Nielsen (2000) mener at fem brukere er tilstrekkelig, og at rundt 80% av problemer knyttet til brukervennlighet blir funnet i brukergrensesnittet med dette antallet. Jakob Nielsen viser til mange prosjekter når de poengterer at en typisk bruker vil oppdage rundt 31% av brukervennlighetsproblemene i et grensesnitt. Bruker nummer to og tre vil finne mye av det samme som første bruker, og hva vi lærer av testene overlapper. Etter bruker fem mener de at man kaster bort tid da man i stor grad vil oppdage de samme problemene og får dermed lite ny innsikt (Nielsen, 2000).

En testplan bør være formulert før selve brukertesten starter. Testplanen skal i hovedsak klargjøre hva vi ønsker å finne ut av, samt praktisk informasjon som blant annet deltakere, prosedyre, utstyr og lokasjon.

Tradisjonelt innen brukertesting skiller man felt- og laborietesting. Historisk har brukertesting av forskjellige produkter og løsninger blitt testet i et kontrollert miljø, ofte kalt laboratorier, for å se på hvor «brukervennlig» produktet er. Fordelen med å teste i et laboriemiljø er at man i større grad er i kontroll over eksterne faktorer som støy, bevegelse, sosiale hindringer og at lignende skal spille inn på resultatene av testen. Man kan argumentere for at et slikt miljø ikke simulerer en kontekst brukeren naturlig ville ha befunnet seg i, noe som da ikke blir en reell brukssituasjon.

Denne brukertesten vil ta for seg to chatbotgrensesnitt ved bruk av en bærbar datamaskin i et delvis kontrollert miljø. Fuglerud (2011) argumenterer for å la brukerne benytte eget utstyr når man brukertester med mennesker som kan ha behov for støtteprogram. Hun poengterer at det er et hav av forskjellige kombinasjoner av programvarer og teknologiske hjelpemidler for å støtte brukernes behov, noe som gjør det utfordrende å optimalisere i et fastsatt testmiljø. Vår

tilnærming var derfor å tilrettelegge i størst mulig grad for deltakerne. Vi fasiliterte slik at deltakerne selv kunne bestemme om de ønsket å bruke eget utstyr, samt hvor de ønsket at testen skulle finne sted. Dette var også et strategisk valg med tanke på å forenkle rekrutteringsprosessen, noe som vil bli diskutert i kapittel 3.3.7.

### **Gjennomføring**

Før vi gjennomførte brukertestene ønsket vi å gå gjennom testplanen (se vedlegg E) i en pilot-test for å finne eventuelle feil, mangler og utfordringer vi ikke hadde tatt hensyn til. En pilot-test er en fullstendig gjennomgang av de fleste punktene i testplanen. Her testes alt utstyr, oppgaver, programvare og rollene til testpersonell før man brukertester med de reelle deltakerne. En slik gjennomgang gjør også at vi får en felles forståelse for hendelsesforløpet i testen, i tillegg til at vi blir mer fortrolig med alle praktiske elementer det innebærer i en brukertest. Sandnes (2011) poengterer at en godt planlagt brukertest gir et inntrykk av profesjonalitet, som igjen vil øke tilliten til testen for de som deltar, noe som vi anser som viktig på med tanke på at vi rekrutterer fra en sårbar gruppe.

Pilot-testen var todelt der vi hadde en gjennomgang av støtteprogram med en dyslektiker og en oppgavegjennomgang med en medstudent. Årsaken til at oppgavegjennomgangen ble utført på en medstudent var at vi ikke hadde tilgang til personer med dysleksi på dette tidspunktet. Vi prøvde derfor å gjennomføre testen i form av rollespill, der han var deltaker og vi test-personell. Selv om medstudenten ikke hadde dysleksi, hadde han et nært familiemedlem som er dyslektiker og kunne derfor til en viss grad leve seg inn i rollen. Uavhengig av at vi ikke gjennomført pilot-testen med en person med dysleksi opplevde vi det som verdifullt. Vi fikk blant annet øvd oss på rollene som testleder og observatør, i tillegg så vi at gjennomføringen overskred tidsestimatet fra test-planen, noe som førte til at vi kuttet ned på antall oppgaver.

Før brukertesten startet var det viktig for oss å få bekreftet at deltakerne var innforstått med testprosedyren og deres rettigheter. Med tanke på at våre deltakere har lesevansker var det viktig for oss at de var og ble informert muntlig om hva de skulle gjennom og hvorfor, før testen startet, da denne informasjonen kun hadde blitt meddelt skriftlig tidligere. Deltakerne ble også bedt om å signere et samtykkeskjema på at deltakelse var frivillig, og at de var innfunnet med at skjerm- og lydopptak ble gjort. Vi tilbød også å lese opp samtykkeskjemaets



innhold til deltakerne. Anonymiseringstiltak vi gjorde i forbindelse med brukertesten blir gjenfortalt i kapittel 3.4.

Totalt ble det gjennomført fem brukertester der alle testene, foruten en, ble gjennomført på Institutt for Informatikk ved Universitet i Oslo. Den siste testen ble gjennomført via videokonferanse i den nettbaserte videotjenesten Zoom på grunn av problematikken rundt covid-19 som hindret personlig oppmøte for alle parter. Vi vil her begrunne noen av valgene vi har tatt med tanke på gjennomføring, samt hvilke forberedelser vi hadde før testene.

Fordeling av våre roller som testpersonell ble gjort før hver test, der en var testleder, mens den andre var observatør. Testleder er den som kommuniserer med deltakerne under testen og introduserer oppgaver samt svarer på eventuelle spørsmål fra deltakeren. Om deltakerne blir stille, eller stopper opp, er det viktig at testleder oppfordrer til å tenke høyt, slik at årsaken til hvorfor det stoppet opp blir funnet. Observatørens rolle var å følge med på hva deltakerne sa og gjorde, og noterte funn og avvik i tillegg til interessante tanker fra deltakerne.

Alle deltakerne ble tilbudt å bruke eget utstyr og tekniske hjelpemidler, men dette ble valgt bort av samtlige. Fra intervjuene med deltakerne før brukertesten ble vi fortalt at de ikke ville ha benyttet tekniske hjelpemidler i en reell brukskontekst når det gjelder interaksjon med en chatbot. Lese- og skrivestøtteprogram ble i større grad benyttet i utdanningssammenheng eller når man skulle produsere en formell tekst i for eksempel en e-post. Det var derfor mottakeren eller viktigheten av det tekstlige innholdet som avgjorde når de ønsket å bruke støtteprogrammer for å lese og skrive. Hovedargumentet for at de var selektive med å benytte støtteprogrammer var etter vår oppfatning at det opplevdes som kronglete og tidkrevende å bruke ekstern programvare.

For å forenkle etterarbeid og gjennomføring av brukertestene benyttet vi programvare som tok opptak av lyd og skjerminteraksjon. Dette ga oss rom for å være mer til stede under testen i tillegg til at vi i større grad kunne observere og notere på detaljnivå. Opptakene ble deretter lagret på vår datamaskin og ble på samme dag, eller dagen etter, transkribert. Transkribering innebar å skrive ned hva de sa, og hva de gjorde samt dokumenter med skjermbilder fra opptaket.

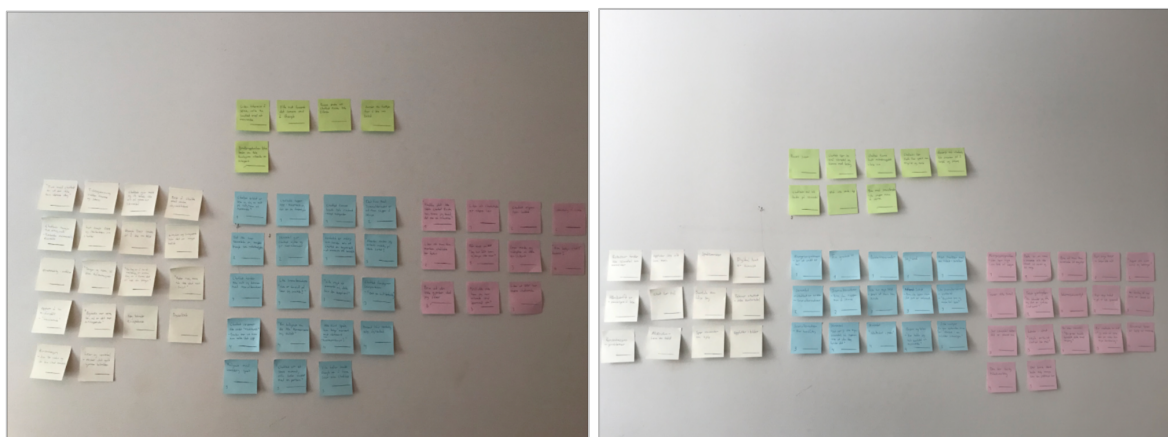
Ved gjennomføring fikk deltakerne opplest hver oppgave av oss, i tillegg til å ha et oppgaveark tilgjengelig. Vi var påpasselig med å minne de på om å tenke høyt under

gjennomføring og var noe tilbakeholden med å gi direkte svar eller hjelp om de møtte på utfordringer underveis, da vi ønsket å se den reelle opplevelsen de hadde med grensesnittene. Dette var noe de ble informert om innledningsvis, før teststart.

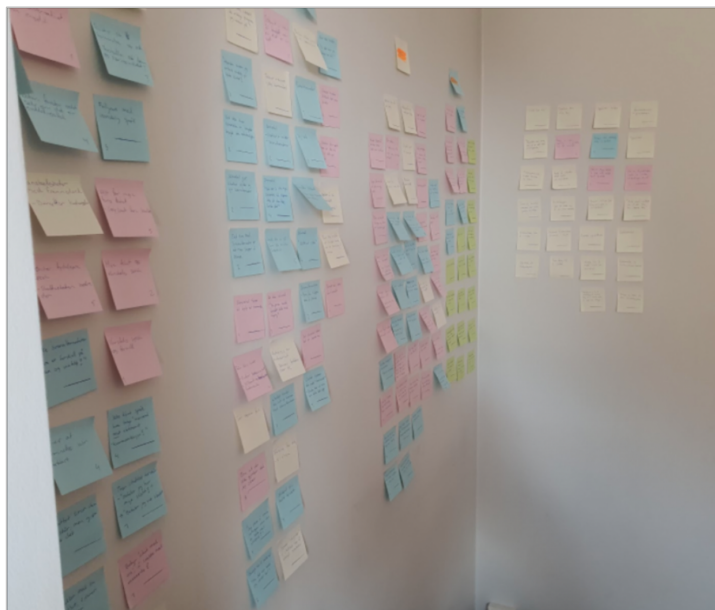
### 3.3.5 Data analyse

Analyse av brukertesting er ifølge Lazar et al. (2017, s. 292) likt som å analysere data i alle andre typer forskning. Mange velger å dokumentere det med store formelle rapporter, andre skriver en rask oppsummering og enkelte lager post-it lapper og presenter dem, men det finnes ingen beste tilnærming for å analysere eller rapportere funn i brukervennlighetsstudier hevder Nielsen (2005). Dataen som blir samlet inn under brukertesting er kvalitativ, og ifølge Sharp et al. (2019, s. 324) analyserer man kvalitativ data enten induktivt eller deduktivt. En induktiv analyse beskrives som utforskende og passende dersom man ønsker å skape konsepter gjennom datamaterialet, og deduktiv benyttes om man allerede har analytiske rammer slik som teorier eller kategorier med tanke på studiets mål. Analyse handler stort sett om hvordan man gir mening til data man samler inn, og vi forstår det som en kontinuerlig prosess der analytiske elementer typisk kan vokse frem underveis i forskningsprosessen. Walsham (2006) sin tilnærming til analyse er både uplanlagt og løs, og han argumenterer for at en forsker sitt beste verktøy for analyse er sin egen kunnskap supplementert med andres kunnskap.

### Gjennomføring



**Figur 3.3:** Beskrivelser og funn fra en enkelt brukertest. Hvor hvite lapper representer intervjuer, Skatteetaten er blå, NAV er rosa og grønne lapper for ønsker/behov en deltaker hadde.



**Figur 3.4.** Organisering og kategorisering av data.



**Figur 3.5.** Dele inn temaer

Analytiske elementer hadde allerede vokst frem i forskningsprosessen, noen temaer var identifisert etter vi hadde blitt kjent med datamaterialet i datainnsamlingen, transkribering, samtaler og oppsummeringen av funnene. Men vi ville få et enda nærmere forhold til dataen fra brukertesting, og valgte derfor å analysere den mer formelt med en induktiv tilnærming. Stegene i analysen handlet om å kategorisere funn inn i temaer og utforske om det var andre temaer vi kanskje ikke hadde oppdaget. Analyseprosessen ble for det meste gjort sammen slik at vi kunne diskutere oss imellom og for å unngå fallgruver slik som å lage temaer basert på personlige erfaringer. Det overordnede målet med analysen var å finne temaer og kategorisere funn fra oppsummeringen, og prosessen beskrives videre under i tabell 1.

Steg	Beskrivelse av analyse stegene
1.	Skrive lapper med funn og beskrivelser fra hver brukertest (se figur 3.3). Dette delte vi opp slik at vi kunne gjøre det individuelt, resten av analyse stegene gjorde vi sammen.
2.	Organiserte og kategoriserte lappene vi mente passet sammen (se figur 3.4) og identifisere tidligere temaer.
4.	Samle funn og beskrivelser inn i temaer (se figur 3.5), dersom lappene ikke passet prøvde vi å identifisere potensielle nye temaer.

- |    |  |
|----|--|
| 5. | Gå igjennom alle lappene i hvert tema for å sikre at vi var enige om valgene, og de endelige temaene vises i kapittel 4.3. |
|----|--|

**Tabell 1.** Viser vår analyseprosess av datamaterialet fra brukertesting og hvordan vi arbeidet med det.

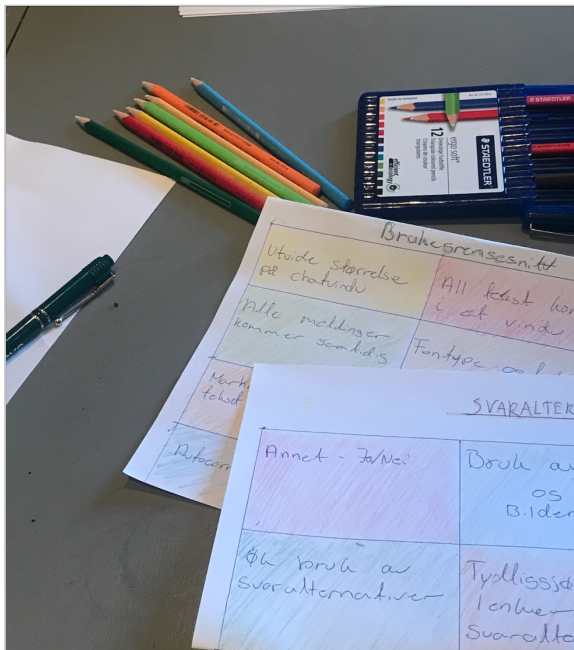
### 3.3.6 Metoder og teknikker for design

I de neste avsnittet presenterer vi metodene vi benyttet for å utvikle designet etter brukertesting, og hvordan de ble benyttet for å komme frem til designforslag. Metoder og designteknikker som har blitt benyttet er brainstorming, skissering, prototyping og en evaluerings-workshop.

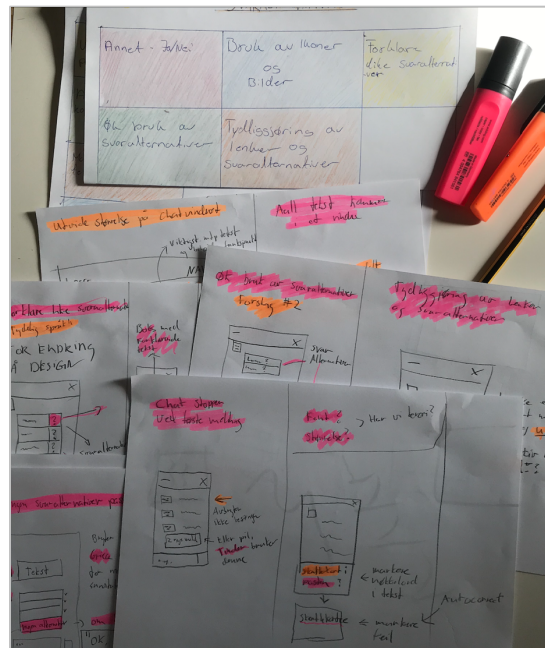
#### 3.3.6.1 Brainstorming og skissering

Med en oversikt over identifiserte temaer fra analysen, ønsket vi å skape ideer for å løse problemstillingene. Det finnes en rekke teknikker for å skape ideer, blant annet brainstorming og skissering, som vi har valgt å bruke. Sandnes (2011) beskriver brainstorming som en kreativ tankeprosess som normalt gjøres sammen med flere i to faser (1) den første fasen er en divergerende prosess der man utforsker mange ideer til problemet, og (2) i neste fase konvergerer man ved at man sammen velger en ide. Skisser lages av kunnskap man allerede besitter, og når man tolker skissene lages ny kunnskap hevder Buxton (2007, s. 114). Det fine med skisser er at man kan lage konsepter raskt, uten å binde mye tid og følelser i forslaget, noe som gjør at det enkelt kan forkastes. Det er også en god metode når man samarbeider med andre for å evaluere hverandres konseptforslag, og muligheten for implementering av ny funksjonalitet (Sandnes, 2011).

## Gjennomføring



Figur 3.5. Brainstorming.



Figur 3.6. Skissering.

Brainstorming som teknikk ble brukt for å utforske ideer og alternativer til problemer som var identifisert i analysen av brukertesting. For hvert problem identifisert hadde vi et papirark som var delt inn i seks deler. På sesjoner som varte i 6 minutter skrev vi ned alle mulige ideer vi kunne komme på. Deretter oppsummerte vi ideene i fellesskap. Det hendte at vi ble inspirert til å lage nye ideer basert på ideene som ble oppsummert. Det fine med å arbeide sammen var at vi begge hadde ulike erfaringer som ga oss en rekke forskjellige ideer til inspirasjon og diskusjon. Denne tankeprosessen var divergerende da vi så mange ideer som potensielt kunne forbedre eller løse problemene funnet i analysen. Den neste prosessen var å konvergere. Det vil si at vi gikk fra mange ideer til bare noen få. Dette ved å velge de vi mente var realistisk og hensiktsmessig å bygge videre på i forhold til bakgrunns litteratur.

### 3.3.6.2 Prototyping

For å visuelt presentere designforslagene ønsket vi å utforme prototyper. Prototyper beskrives av Houde og Hill (1997) som representasjoner av designideer. De har et bredt spekter av formål og kan blant annet brukes i brukertesting og evaluering av ideer for å få svar på spørsmål (Sharp et al., 2019). Hvordan man designer en prototype avhenger i stor grad av hva man ønsker å få svar på og derfor tenker man også på hva formålet med prototypen skal være.

Når prototyper i mindre grad uttrykker hvordan det fremtidige designet kan se ut er de typisk lavoppløselig eller på engelsk «low-fidelity». Begrepet fidelity beskrives av Houde og Hill (1997) som designe sin nærhet til det fremtidige sluttproduktet. En lavoppløselig prototype er ikke nært det fremtidige designet og er typisk begrenset i interaksjon. Lavoppløselige prototyper blir normalt brukt tidlig i designprosessen, for eksempel under konseptualisering fordi de oppmuntrer til utforskning og modifisering (Sharp et al., 2019). De brukes til å skildre alternativer for design i stedet for å modellere brukerens interaksjon med et system. Dette står i motsetning til høyoppløselige prototyper der designet skal føles mer som det endelige produktet med en fullstendig interaksjon (Rudd, Stern & Isensee, 1996),

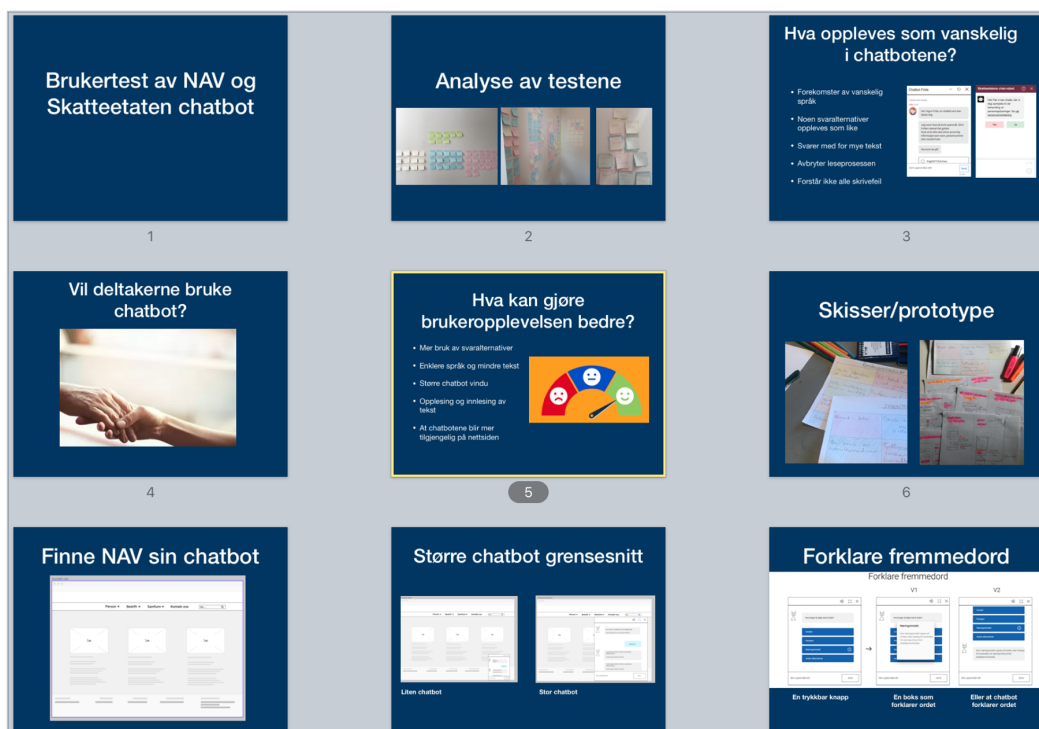
### **Gjennomføring**

Vi utformet prototyper som er informert av skisser, bakgrunns litteratur, eksperter og analyse. Prototypene er en konkretisering av vår forståelse av deltakernes behov og våre ideer til hvordan chatbotfunksjonalitet kan designes for å dekke behovene. Prototypene er lavoppløselige, og er utformet i verktøyet Figma. Prototypene blir vist og forklart i kapittel 5.

#### 3.3.6.3 Evaluerings-workshop

Vi ønsket å evaluere designforslagene med deltakere som hadde vært med på brukertesten. En evaluering er nyttig for det kan hjelpe med å avkrefte eller bekrefte om forslagene har noe for seg, og om det er noe man ønsker å iterere videre på. Vi kaller dette en evaluerings-workshop, der handlingsløpet var først å presentere funnene fra analysen og deretter diskutere prototypene.

## Gjennomføring



**Figur 3.7.** Et utdrag fra presentasjonen vi holdt for to av deltakerne i evaluerings-workshopen.

Hovedpoenget med å evaluere med deltakere var for å spille inn de forslagene vi hadde designet, samt hva vi hadde valgt bort, slik at vi kunne få tilbakemeldinger om hvordan de oppfattet disse endringene. Evalueringen hadde også til hensikt å gi deltakerne en mulighet til å bekrefte eller avkrefte om vi hadde tolket brukervennlighetsbarrierene de opplevde korrekt. Vi ser det som viktig å få tilbakemelding fra brukergruppen på om dette er faktiske tiltak som kan fungere for dem, og i like stor grad er det viktig å diskutere hva som eventuelt ikke fungerte og hva som kunne ha blitt utbedret.

To av fem deltakere som var med i brukertesten deltok i evalueringen som varte i ca. 1 time. På bakgrunn av covid-19 pandemien var det ikke mulig å møte deltakerne personlig. Vi avtalte derfor å møtes på videokonferansetjenesten Zoom. Vi startet med å ha en presentasjon der vi informerte om resultatene fra brukertesten og hvilke funn vi hadde gjort (se Figur 3.7). Videre presenterte vi prototypene av konseptene vi hadde utarbeidet. Etter presentasjonen diskuterte vi tilgjengeligheten til chatbotene og prototypene vi hadde utformet.

### 3.3.7 Rekruttering

Gjennom hele rekrutteringsprosessen, både til intervjuer og brukertester, hadde vi et tett samarbeid med Dysleksi Norge. Innledningsvis fikk vi kontakt med leder for Dysleksi Ungdom som er underorganisasjon av Dysleksi Norge. Lederen har vært en sentral ressurs for oss, både som ekspert i forbindelse med innsiktsarbeid, og kontaktperson ved rekruttering av utvalg. Vi spesifiserte hvilke kriterier vi hadde for ønsket utvalg av deltakere. To deltakere i ble rekruttert gjennom eget nettverk.

### 3.4 Etske betraktninger

Merriam og Tisdell (2015, s. 261) mener at kvalitative forskere må sannsynligvis ta stilling til etiske vurderinger i flere faser av forskningen. Noen etiske vurderinger er forhåndsbestemt, og andre kan oppstå underveis i feltet, men hvordan man håndterer dette kommer an på våre etiske verdier og vurderinger av hva som rett eller galt (Merriam & Tisdell, 2015; Bogdan & Biklen, 1997). Dette kapittelet tar for seg etiske betraktninger vi har gjort oss i forskningsprosessen, og hvordan vi har gått frem for å løse disse.

Når man involverer mennesker i forskning, er det tradisjonelt to retningslinjer for etikk man alltid har fulgt: fritt informert samtykke og beskytte deltaker mot skade (Bogdan & Biklen, 1997). I Norge blir dette regulert når man ønsker å behandle personopplysninger ettersom man er meldepliktig til å søke Norsk Senter For Forskningsdata (NSD). Den 8. november 2019 fikk vi tillatelse til å behandle personopplysninger av NSD for to utvalg: eksperter og personer med funksjonsnedsettelse. Det sistnevnte utvalget har blitt anonymisert som innebærer at de ikke skal kunne identifiseres i datamaterialet. En anonymisering innebar å pseudonymisere utvalget med fiktive navn og kjønn, og andre indirekte bakgrunnsvariabler har ikke direkte blitt koblet til pseudonymiseringen. Lydopptak og videoopptak som ble transkribert har blitt slettet innen 01.06.2020. Datamaterialet vi har samlet inn har kun vært tilgjengelig for forfatterne av studiet, og blitt lagret forsvarlig i henhold til avtalen vi har med NSD.

Deltakere som har deltatt i studie har blitt informert muntlig og skriftlig og presentert et samtykkeskjema for signering før deltakelse i intervju eller brukertest (se vedlegg A og B). I dette skrivet informerer vi blant annet om formål med studien, hva det innebærer å delta,



deres rettigheter og at det er frivillig, vår kontaktinformasjon og at man når som helst kunne trekke seg uten å måtte gi en spesiell grunn for det.

Et etisk og metodisk dilemma som oppsto underveis i studiet var om vi skulle informere deltakerne på forhånd angående hva vi ønsket at de skulle teste. Et av målene i brukertesten var å undersøke tilgjengeligheten til chatbotene ved å se om de ble oppfattet som et menneske eller en maskin. Dersom deltakerne allerede hadde forkunnskap om at det var chatboter som skulle testes, mener vi det var mer naturlig for deltakerne å trekke slutninger til at det var maskiner. Likevel har deltakerne en rett til å vite hva de samtykker til, og etter våre egne vurderinger så vi ikke at formålet kunne gå på bekostning av deres rett til å være godt informert. Vi informerte derfor deltakerne om at det var chatboter som skulle testes i dette studiet.

Vi mener om det oppstår en konflikt eller en bekymring, har vi en forpliktelse til å finne alternative løsninger for å beskytte deltakeren. Det var tre eksperter som uttrykket underveis at de ikke alltid snakket på vegne av deres organisasjon. Noen eksperter valgte også å trekke samtykke for å bli sitert. Eksperter som har samtykket og blitt presentert i form av sitering har også fått tilsendt utkast av hvordan de har blitt sitert i studiet.

Vi ba deltakerne i dette studiet om mye, både i form av tid, innsats og innsikt inn i deres liv. Det har derfor vært viktig for oss å gi noe tilbake. Dette er i tråd med det Bogdan og Biklen (1997) mener om å uttrykke gjensidighet. Det er ifølge dem et prinsipp som ofte har blitt tatt for gitt blant kvalitative forskere: at de tar mye og gir mindre eller ingenting tilbake. Som et motargument at deltakere ikke får tilbake har kvalitative forskere vist til hvordan forskningsbidraget bidrar tilbake til samfunnet (Bogdan & Biklen, 1997). Uansett er vi av den oppfatning at det er viktig å gi tilbake dersom man får.

Dyslektikere som har deltatt i dette studiet er en gruppe som potensielt er utsatt for sårbarhet når det gjelder skriving og lesing, og vi har på forhånd, underveis og i ettertid tatt etiske vurderinger med tanke på det. Det ble for eksempel klart for oss at vi skulle tilby deltakere å lese opp samtykkeskjema og oppgavene dersom de ønsket det. Det ble også tydelig nevnt at vi ikke var interessert i å teste deres evner, og at det var chatbotene som skulle testes. Når vi kommuniserte med deltakerne prøvde vi å være påpasselig med hvordan vi skrev til dem,

helst med korte setninger og enkle ord. Vanskeligheter med skriving og lesing kan i enkelte situasjoner føre til at de mister motet og gir opp. En oppmuntring underveis i brukertesten har derfor vært viktig. Som en avslutning på studiet hadde vi en oppsummerende presentasjon for Dysleksi Norge. For oss var dette en god måte å gi tilbake til deltakerne, i motsetning til å avslutte kontakt med dem brått.

## 4. Brukertesting av chatbotene

Dette kapitlet presenterer en oppsummering av brukertestene og intervjuene vi gjennomførte med deltakerne. Intervjuene blir først gjennomgått og deretter blir hver oppgave oppsummert med en hensikt om en gjengivelse av gjennomføringen.

For å skjule identiteten vedrørende sensitiv informasjon har vi satt fiktive navn og kjønn på deltakerne, og kaller dem videre Einar, Mari, Kari, Anne og Bjørn.

### 4.1 Intervju med deltakerne

Før vi hadde brukertesten ble det gjennomført preintervjuer med deltakerne, som omhandlet deres livserfaringer som dyslektiker, teknisk kompetanse og erfaringer med chat og chatbot.

#### 4.1.1 Om deltakerne

I første del av intervjuet ønsket vi å forstå mer om hvem deltakerne er, og hva dysleksi innebærer for dem. Deltakerne er mellom 18 – 30 år, hvor tre er kvinner og to er menn. Samtidig som at de har dysleksi har deltakerne forskjellige grader av dysleksi som varierer mellom mild og kraftig, noen har også andre kognitive nedsettelse. Samtlige deltakere nevner at aktiviteter som innebærer å lese og skrive kan ta lang tid fordi man både leser og staver ord feil. Noen har også opplevd stigmatisering knyttet til dette. Med årene har det imidlertid blitt enklere og mer avslappende med skriving og lesing. Det har for eksempel blitt bedre for noen av deltakerne ved at de har tatt i bruk tekniske lese- og skrivestøtteprogram eller lært seg diverse teknikker for å håndtere utfordringer knyttet til dysleksi.

Fra tidligere intervjuer med Dysleksi Norge hadde vi også blitt informert om at aktiviteter som omhandlet skriving og lesing var noe man helst ville unngå. Dyslektikere vil heller snakke enn å skrive, og de vil heller lytte enn å lese.

#### 4.1.2 Lese- og skrivestøtteprogram

Det varierer i støtteprogram som blir brukt blant deltakerne, noen eksempler på program som de har brukt er CD-ord, Lingwrite og Lingdys. Disse har diverse funksjonaliteter slik som stavekontroll, talesyntese, ordfullføring og ordbok. Andre eksempler er (1) Grammarly som er en utvidelse i nettleser for stavekontroll funksjonalitet, (2) Google søk for stavekontroll og

Google Docs som støtter innlesing av tekst til dokument, og (3) stavekontroll som er innebygd i Apple og Android telefoner. Det varierer når disse blir brukt og hvor ofte disse programmene blir brukt.

Lese- og skrive støtteprogram blir i større grad benyttet i utdanningssammenheng eller når man skulle produsere en formell tekst i for eksempel en e-post. Stavekontroll på telefon blir for to deltakere mer aktivt brukt i daglig sammenheng, og noen nevner de har «vokst fra seg» nytten med støtteprogram. Det har blant annet blitt opparbeidet diverse teknikker, med og uten støtteprogram, for å håndtere vansker med dysleksi bedre. Blant annet ved å lære gjennom bilder, tegne eller tilpasse støtteprogrammene. «En ting er å ha et redskap, en annen ting er å forme det til noe du trenger», sier Mari i denne sammenhengen.

#### 4.1.3 Erfaringer med Chat

Når det kommer til å «chatte» via tekst med andre mennesker på sosiale medier, meldinger og kundeservice eller lignende, blir det sagt av samtlige at de ikke foretrekker det. Generelt blir det heller foretrukket å snakke muntlig. Dersom de først chatter tar dette ofte lengre tid og det kan hende man blir sliten av det. Vi blir fortalt at samtaler ofte kan gå i lås. Noen har også en ide eller en følelse om hva de ønsker å skrive, men de får det ikke ned i tekst. Da hender det typisk at man spør om hjelp fra andre. Flere av deltakerne sier at de har opplevd at skrivefeil har blitt kommentert og påpekt på nettet, noe som oppleves ubehagelig. Mari forteller oss at det er best å chatte med andre mennesker som har dysleksi, for da er man på samme nivå og kjenner hverandres begrensninger.

#### 4.1.4 Erfaringer med Chatbot

Generelt har alle deltakerne i en eller annen form benyttet chatbot innen kundeservice. Kari og Mari har benyttet Skatteetatens chatbot, men ingen har benyttet NAV sin. Andre chatboter som har blitt benyttet tidligere er Lånekassen, DNB, Telenor, MSN, Apple, OneCall, Microsoft og GET. De fleste sier at de foretrekker å snakke med et menneske enn med chatboter. Kari sier hun vanligvis bare pleier å skrive «rådgiver» for å unngå samtalen med chatbot, og Mari sier at chatboter mangler noe emosjonelt som man får med mennesker. Det fine med chatbot er at den ikke kan dømme deg forteller hun videre.

## 4.2 Brukertesting rapport

Dette kapitlet tar for seg en oppsummering av brukertestene med fem deltakere som har dysleksi. Oppgavene blir presentert i lik rekkefølge som de ble gitt til deltakerne. Vi vil påpeke at figurene som vises i de videre oppsummeringene er replikert av bilder som vi tok underveis i brukertesten. Siden chatbotene lærer kontinuerlig kan det hende at figurene ikke viser det nøyaktige bildet.

Før vi presenterer rapporten ønsker vi forklare målene med de forskjellige oppgavene. Målene i brukertesten var:

- Forstå ulike behov ved dysleksi
- Er chatbotene tilgjengelig på nettsidene?
- Hvordan oppleves grensesnittene?
- Hvordan oppleves svaralternativer?
- Klarer chatbotene å predikere brukerintensjonen til deltakerne?
- Forstår deltakeren informasjonen fra chatbotene?
- Får deltakeren svar på spørsmålet sitt?

### **Forstå ulike behov ved dysleksi**

Et av de overordnede målene var å undersøke og bedre forstå deltakernes behov. Et behov, eller noe de har bruk for, kan når som helst uttrykkes i både intervjuet og brukertesten, men det er ikke alltid behovene kommer godt nok frem. For eksempel om en deltaker synes testmiljøet er skummelt kan deltakeren være noe reservert for å åpne seg. Dette var noe vi som testpersonell måtte være bevisste på, slik at vi kunne tilrettelegge og ufarliggjøre brukertestsituasjonen så godt som mulig. Vi hadde spørsmål som «Siste inntrykk» og om de kunne forestille seg «Den perfekte chatboten» for å redusere effekten av at behovene til deltakerne ikke kom frem. Det er også av interesse å få feedback og tanker fra deltakerne uavhengig av hvilket scenario som oppstår.

### **Er chatbotene tilgjengelig på nettsiden?**

Med dette spørsmålet ønsket vi å finne ut om det var en teknisk tilgjengelighet til stede på nettsidene til Skatteetaten og NAV. Det vil si at nettsiden er tilrettelagt slik at deltakerne når frem til chatbotene, og at chatbotene er utformet slik at deltakerne oppfatter at det er en

chatbot de oppretter dialog med. Dette delmålet hadde oppgave 1 og 5 i brukertesten hensikt med å utforske.

### **Hvordan oppleves grensesnittet?**

For å bedre forstå hvordan vi kan designe for deltakernes behov ønsket vi spesifikt å vite mer om deres oppfatning av chatbotgrensesnittene. Innsikt til dette ville komme fra alle oppgaver i brukertesten, men vi ønsket spesifikt å spørre om dette i oppgave 1 og 5. Det vil si når de hadde navigert seg frem til chatbotene.

### **Hvordan oppleves svaralternativer?**

Svaralternativer er noe som blir brukt i begge chatbotene. Vi ønsket å undersøke hvordan en slik alternativ funksjonalitet i en chat opplevdes for deltakerne. Det ble derfor utformet spesifikke testoppgaver som ville føre til at chatbotene svarte med svaralternativer.

Skatteetatens chatbot svarte som regel alltid med svaralternativer dersom man ikke inkluderte alle nøkkelordene som var nødvendig. Derimot i NAVs chatbot måtte vi prøve oss frem for å finne når svaralternativ forekom. I oppgave 3, 4, 7 og 8 antok vi at det var sannsynlig at svaralternativer ville forekomme.

### **Klarer chatbotene å predikere brukerintensjonen til deltakerne?**

Med dette ønsket vi å undersøke om deltakerne opplevde om de ble forstått, og hvordan de opplevde det. Alle oppgavene var tilrettelagt slik at det var mulig å komme frem til rett svar. Det var således opp til chatbotene om de klarte å predikere rett i forhold til hva deltakerne skrev til dem eller valgte i svaralternativene. Avhengig av chatbotenes evne til å predikere brukerintensjonen vil det bli gitt ulik respons som enten er en feilmelding, svaralternativer eller svar. Vi ønsker å se på hvordan deltakeren velger å formulere seg, og om chatboten vil ta hensyn til eventuelle stavefeil og feilformuleringer.

### **Forstår deltakeren informasjonen fra chatbotene?**

I tillegg til det forannevnte delmålet ønsket vi å forstå hvordan deltakerne opplevde informasjon som chatbotene responderte med. To av oppgavene (3 og 6) ville føre til at chatbotene produsert lengre svar eller flere handlinger. Vi ønsket å innhente tanker og refleksjoner fra deltakerne rundt mengden og språket på det tekstlige innholdet. Disse var de mest informasjonsintensive svarene, og var ment som de «vanskeligste» oppgavene.

## Får deltakeren svar på spørsmålet sitt?

Det siste delmålet tok for seg å finne ut om de klarte oppgavene eller ikke, og hvordan dette opplevdes. De videre test oppgavene som er markert med fargekodene, gult, grønn og rød gir en indikasjonen om hvordan det gikk.

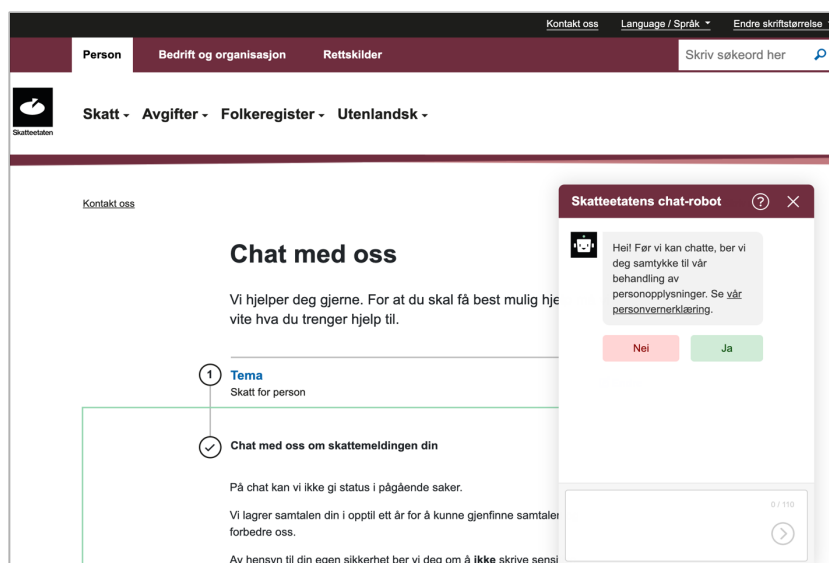
Oppgavene var for øvrig stilt opp slik at de først hadde en enklere vanskelighetsgrad, og så ble gradvis vanskeligere, også litt enklere igjen. Baktanken var at deltakerne skulle føle at de fikk en mestringsfølelse ved start og slutt. Siste oppgave (8) der man skal spørre om å komme i kontakt med et menneske var en oppgave vi antok som «enkel», men det er ikke nødvendigvis slik at det vil opplevdes sånn.

### 4.2.1. Skatteetatens chatbot oppsummering

Dette er en oppsummering av brukertestene fra Skatteetatens chatbot. Som nevnt over indikerer (1) grønn fargekode at deltakerne klarte oppgavene med få problemer underveis, (2) gul indikerer at det oppstod problemer, og (3) rød betyr at oppgaven ikke ble fullført.

#### Oppgave 1: Naviger deg til Skatteetaten sin chat «skatt for person».

Einar	Mari	Kari	Anne	Bjørn



Figur 4.1. Skatteetatens chatbot landingsside (Lenken for å navigere seg til chatboten er plassert på øvre del av nettsiden, under lenken «kontakt oss»).

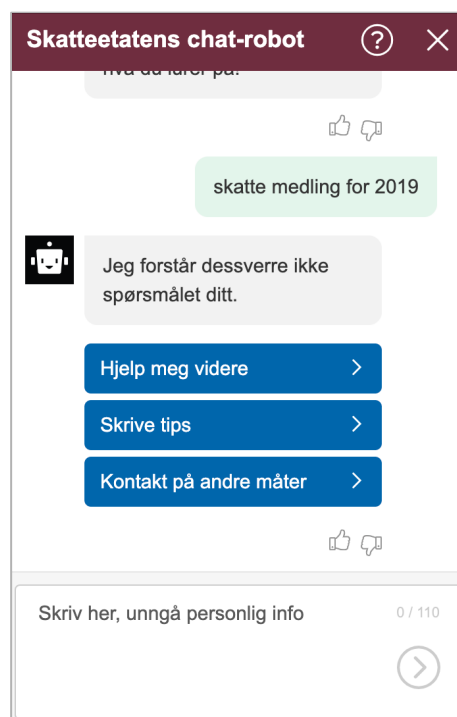
Generelt klarte deltakerne å lokalisere seg frem til chatboten som vist i figuren over. For å komme frem til nettsiden blir det enten brukt Google eller adressefeltet i nettleseren. Einar måtte, etter å ha brukt noe tid, ha hjelp med å rette skrivefeil for å komme til nettsiden, men derfra var det ingen problemer med å finne chatboten. For Anne og Bjørn var det heller ikke uproblematisk å finne lenken som førte til undersiden til der man startet chatten. De påpeker at den burde være lettere tilgjengelig: «Hadde forventet å finne den nede på høyre side ... en pop-up hadde vært veldig kjekt» og «Liker når den er nederst på siden og sier sånn: Hei, snakk med oss». Når deltakerne startet chatboten var førsteinntrykket at de forsto at dette var en maskin og ikke et menneske. Mari, Bjørn og Anne syntes chatbotgrensesnittet var lite i forhold til tilgjengelig plass på nettsiden. Mari sier i denne sammenhengen at hun skulle gjerne hatt mulighet til å forstørre grensesnittet, slik at man kunne se alt av tekst som ble produsert.

**Oppgave 2:** Du ønsker å få svar på når skattemeldingen for 2019 blir tilgjengelig  
Hvordan går du frem for å finne ut av dette ved å bruke chatten?

Einar	Mari	Kari	Anne	Bjørn



**Figur 4.2.** Skatteetatens chatbot predikerer noe av intensjonen rett.



**Figur 4.3.** Skatteetatens chatbot predikerer feil.



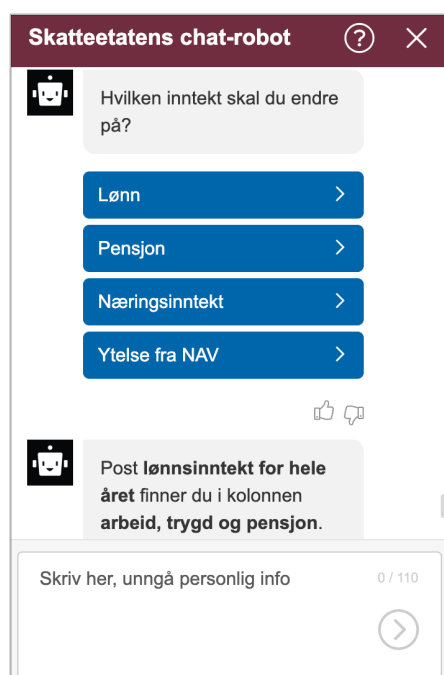
Einar, Mari og Kari hadde vanskeligheter med å formulere seg, og skrivefeil forekom. Mari legger for eksempel selv merke til at ord er skrevet feil, men chatboten forstår at bruker er ute etter skattemelding selv om dette er skrevet feil. Hun er fornøyd med svaralternativene som kommer etter det (se figur 4.2).

Kari fikk en tilbakemelding av chatboten om at den ikke forsto spørsmålet (se figur 4.3), og her oppsto det usikkerhet fra deltakeren om det var en skrivefeil eller om problemet lå i chatbotens evne til å forstå. Kari sier: «Er det fordi jeg skriver feil nå?», hun velger så å kopiere teksten hun skrev inn i Google som kommer med et forslag til riktig stavemåte. Hun får fortsatt ikke riktig svar når hun bruker rettet tekst, og blir oppgitt.

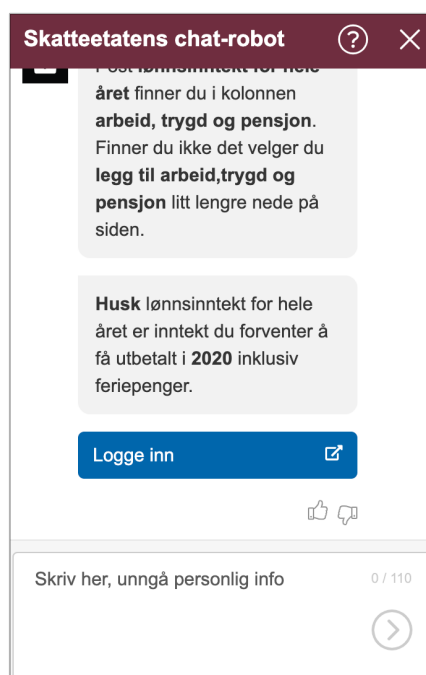
De andre deltakerne benyttet forskjellige fremgangsmåter for å løse oppgaven. Noen formulerer lengre setninger: «Når kommer skattemeldingen for 2019» mens andre fatter seg i korthet «skattemelding 2019». Dette hadde ikke en innvirkning på hvilke resultater deltakerne fikk. Chatboten responderer alle deltakerne med svaralternativer som omhandler skattemelding. Et av alternativene fører til svaret de er ute etter, og alle klarer å navigere via riktig alternativ uten problemer. Deltakerne sier de er positive til svaralternativer. Det fremstår som enkelt og oversiktlig at man får presentert alternativer som omhandler skattemelding.

**Oppgave 3:** Se for deg at du har nettopp fått deg en ny jobb og arbeidsgiver forteller deg at du må oppdatere inntekten på skattekortet. Hvordan går du frem for å få svar på hvordan du endrer inntekten på skattekortet ved å bruke chatten?

Einar	Mari	Kari	Anne	Bjørn



**Figur 4.4.** Skatteetatens chatbot med svaralternativer.



**Figur 4.5.** Skatteetatens chatbot mye tekst.

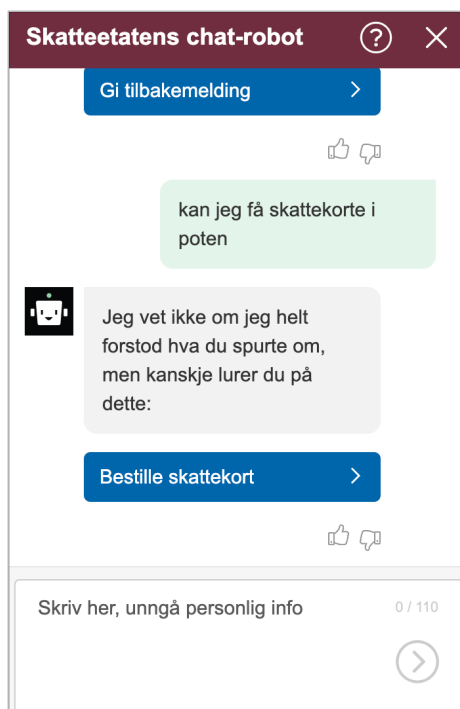
I denne oppgaven er det også noe skrivefeilproblematikk, men oppgaven blir løst av alle deltakerne etter omformuleringer og nye forsøk. Her er det flere utfordringer rundt presenterte svaralternativer. Anne opplever ikke alternativene som relevante ettersom de ikke leder til det rette svaret på spørsmålet hun formulerte. Alternativene hun får presentert av chatboten er generelle emner om skattemelding, der ingen omhandler å endre inntekt / lønn. I dette tilfellet har hun misforstått mellom skattekort og skattemelding. Anne velger å avslutte chatten for så å starte på nytt. Hun tenker at den gamle konteksten fra det forrige spørsmålet har blitt med videre, og sier «... man må finne disse nøkkelordene». Einar, Mari og Anne synes at svaralternativene er like, noe som skaper usikkerhet for hva de skal eller bør velge for å få riktig svar på spørsmålet. Mari klarte oppgaven, men uttrykker at hun: «... ville ringt nå, dette blir for komplisert», når hun måtte velge mellom alternativene (se figur 4.4). Alternativene

som ble opplevd som like var (1) endre inntekt og endre lønn, (2) registrere lønn og endre feil lønn, og (3) lønn og næringsinntekt.

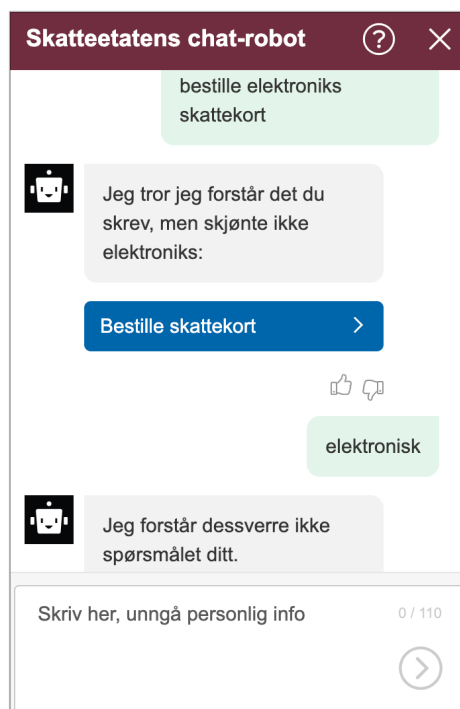
I denne oppgaven responderer chatboten med betydelig mere tekst enn i de foregående oppgavene. Kari og Bjørn synes dette er problematisk. For å lese all teksten må de navigere ved å scrolle seg oppover i chatbotgrensesnittet. «Her tenker jeg alt kunne kommet i ett, og ikke i mange bokser etter hverandre», sier Bjørn (se figur 4.5). Om man starter å lese før chatboten har gjort seg ferdig med å svare blir deltakerne avbrutt mens de leser. Mari sier at «det er helt forferdelig», når lesingen blir avbrutt.

**Oppgave 4:** Du har nå oppdatert lønnen din i forrige oppgave og ønsker å få det oppdaterte skattekortet tilsendt i posten. Hvordan går du fram for å finne ut om dette er en mulighet ved å bruke chatten?

Einar	Mari	Kari	Anne	Bjørn



**Figur 4.6.** Skatteetatens chatbot predikerer noe av intensjonen rett.



**Figur 4.7.** Skatteetatens chatbot blir forstått feil.

I denne oppgaven fortsetter noe av skrivefeilsproblematikken. Som tidligere klarer chatboten å forstå noen skrivefeil. Her svarer chatboten blant annet at den ikke forstår deltakeren, men i noen tilfeller ser man også at den ikke gir relevante svar. Einar har en stavefeil når han skal spørre om skattekort i posten, og skriver teksten: «skattekort i poten» (se figur 4.6). Han sier han ser stavefeilen, men blir tydelig oppgitt over at skrivefeilen ikke blir oppfattet av chatboten. Feilen blir rettet, og meldingen sendes på nytt, men chatboten gir et svar som omhandler lønn, og ikke skattekort i posten. Her gir brukeren opp: «Jeg har jo skrevet alt riktig. Her hadde jeg kontaktet noen for hjelp, for det er umulig for meg å komme videre nå».

Kari opplevde også lignende problematikk, og prøvde å stille spørsmålet på nytt i stikkordsform, noe som resulterte i riktig respons fra chatboten. Hun resonnerer seg frem til at stikkordsform gir bedre resultater enn om man skriver setninger.

Mari lurte på om man kunne bestilles elektronisk skattekort, og ønsket å se om chatboten kunne svare på dette. Hun har en mindre stavefeil i formuleringen og chatboten sier ifra at den ikke forstår ordet. Mari tror at hun kun trenger å rette feilen i ordet, uten å formulere hele spørsmålet på nytt, noe som ikke ser ut til å være tilfellet (se figur 4.7). Hun forteller oss i den sammenheng: «Dette blir jeg frustrert over ... ville heller brukt Google for å prøve å finne svaret».

De andre deltakerne klarte oppgaven uten mye problematikk. Vi ser at orddelingsfeil og feil bruk av dobbeltkonsonanter gjentar seg hos flere i nøkkelord, slik som «skatte melding» og «skatekort». Chatboten klarer som regel å vurdere ordene rett når dette forekommer. Anne sier: «Det er veldig greit for det er noe jeg driver med».

	Siste inntrykk av Skatteetatens chatbot?
Bjørn	Han er overrasket over at chatboten faktisk var ganske bra. Han opplevde det som litt forvirrende når det kom mange meldinger etter hverandre, og skulle ønsket at det heller kom en stor melding. Han hadde også satt pris på at man enkelt kunne få valget om å snakke med et menneske.

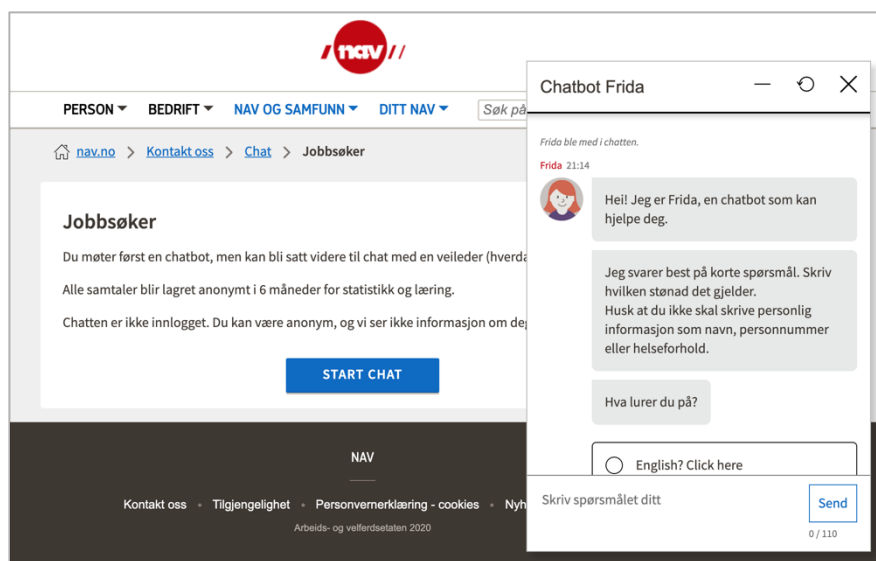
Anne	Anne synes chatboten er god til å gi svar på enkle spørsmål. Chatboten skulle vært mer tilgjengelig på nettsiden, for eksempel et ikon nederst på nettsiden. Hun ønsker at det var svaralternativ som også sier «ingen av svarene», for ellers kan man stoppe opp.
Einar	Han opplever chatboten som grei å bruke, men at det opplevdes som noe stressende.
Mari	Hun synes det kan være en fallgrube med vanskelig terminologi. Hun synes den var grei i bruk og at den virket til å forstå ganske mye. Hun likte at den til en viss grad kunne hjelpe med ord om det var feil stavet. Hun sier at chatboten er et «stressmoment», og at hun heller ville ha snakket med et menneske.
Kari	Kari sier at dette var en bedre opplevelse i forhold til andre chatboter hun tidligere hadde benyttet. Men hun skulle ønske at chatboten var mer tydelig på å si ifra om at stikkord var en mulig måte å snakke med chatboten på.

## 4.2.2. NAVs chatbot oppsummering

Dette er en oppsummering av brukertestene fra NAVs chatbot. Som nevnt over indikerer: (1) grønn fargekode at deltakerne klarte oppgavene med få problemer underveis, (2) gul indikerer at det oppstod problemer, og (3) rød betyr at oppgaven ikke ble fullført.

### Oppgave 5: Naviger deg til NAV sin chat «jobbsøker».

Einar	Mari	Kari	Anne	Bjørn



**Figur 4.8.** NAVs chatbot landingsside (Lenken for å navigere seg til chatboten er plassert på nedre del av nettsiden, under lenken «kontakt oss»).

Mari, Bjørn og Kari går direkte til Google og søker etter «NAV chatbot», og kommer direkte til chatboten som vist i figur 4.8. Det er derimot en større utfordring for de som velger å navigere seg via landingssiden til NAV. Anne og Einar bruker tid på å lete etter chatten og gir til slutt opp. Einar poengterer at den var veldig vanskelig å finne: «Den bør ligge lettere tilgjengelig. For eksempel et ikon synlig i et hjørne». Mari spør: «Hva betyr «chat med oss», betyr det å prate med et menneske?», før hun skal starte chatten. Når chatten starter uttrykker hun seg positivt og sier: «Liker at den er rett på sak, som et menneske». Anne er derimot mer skeptisk og synes at det var litt rart at chatboten hadde navnet «Frida» og at den hadde en kvinnelig avatar og sier «du er ikke et menneske ... jeg vet ikke om jeg ville forstått at dette var en robot med det første».

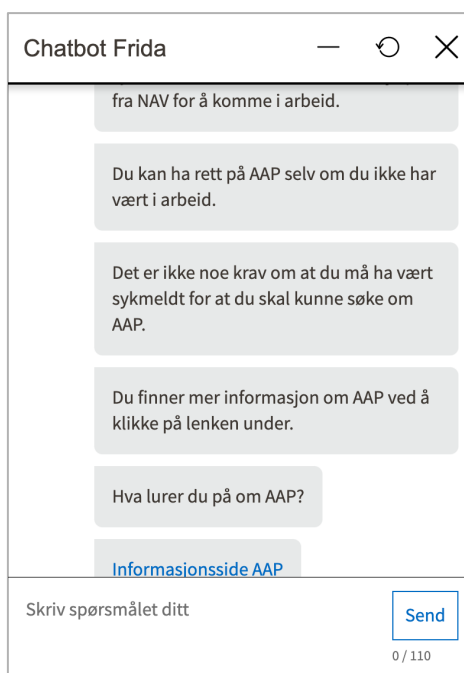
Ved spørsmål om de tror det er et menneske eller en chatbot de prater med, svarer samtlige at de tror det er en chatbot.

**Oppgave 6:** Du ønsker å finne ut hva Arbeidsavklaringspenger (AAP) betyr. Hvordan går du frem for å finne ut hva dette er ved å bruke chatboten?

Einar	Mari	Kari	Anne	Bjørn



**Figur 4.9.** NAVs chatbot predikerer intensjonen rett.



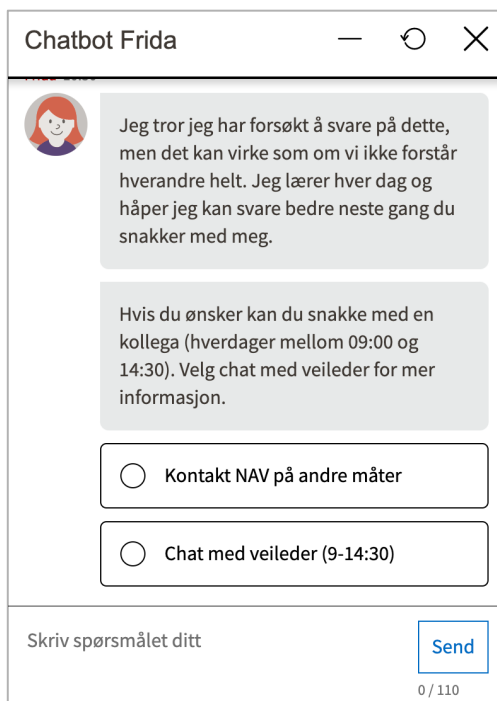
**Figur 4.10.** NAVs chatbot med mange svarmeldinger.

Alle deltakerne klarer å løse oppgaven. Selv om to av deltakerne har skrivefeil i ordet «arbeidsavklaringspenger» predikerer chatboten spørsmålet rett (se figur 4.9). Dette er oppgaven som produserer mest tekst, noe som blir mottatt negativt. Det poengteres at mengden tekst er overveldende, noe som gjør at to av deltakerne velger å ignorere eller hoppe over deler av teksten (se figur 4.10). Einar forteller at det blir slitsomt å lese når teksten blir gjentakende. Han sier det fører til at ord blir blandet, eller at ord og setninger blir hoppet over. Kari velger også å ignorere deler av teksten for så å trykke på en lenke i nederste melding. På spørsmål om hvorfor hun valgte å trykke på lenken sier hun: «Hadde håpet det var mindre tekst: Liker ikke å lese så mye». Anne synes også det er mye unødvendig mye tekst i oppgaven og at språket er vanskelig: «Det er en veldig lang setning, og den trenger ikke å

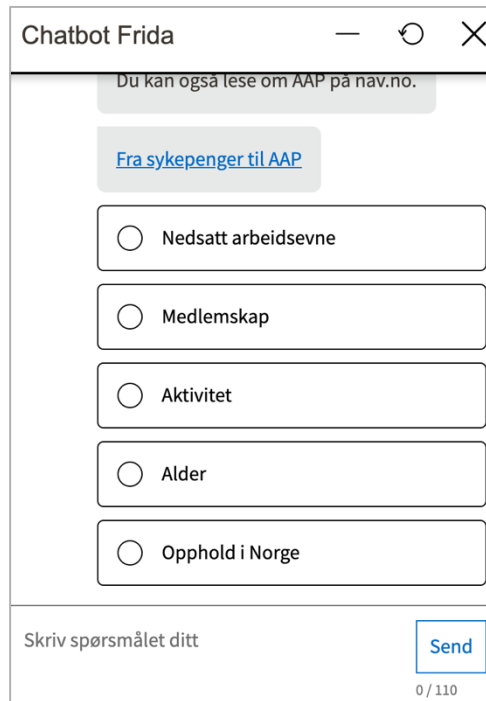
svare med dette for det var ikke noe jeg lurte på». Svaret fra chatboten blir delt opp i flere meldinger som gjør at første del ikke er lesbart før man scroller opp i samtalen. Einar, Mari og Kari hadde problemer med å scrolle opp i chatten da det ikke er en konvensjonell scroll-bar tilgjengelig i chatbotgrensesnittet. Mari sier hun: «... er mer positiv til denne enn den andre. Savner valgmuligheter enn så lenge».

**Oppgave 7:** Se for deg at du har nedsatt arbeidsevne og lurer på om du har krav på AAP. Hvordan går du frem for å finne informasjon om dette ved å bruke chatten?

Einar	Mari	Kari	Anne	Bjørn



**Figur 4.11.** NAVs chatbot vurderer at brukerintensjonen allerede er besvart.



**Figur 4.12.** NAVs chatbot med svaralternativer.

Einar får beskjed fra chatboten at den allerede hadde svart på dette spørsmålet tidligere (se figur 4.11). Dette blir løst når han formulerer seg på nytt. Einar tror årsaken var en skrivefeil og sier «... dette gir ikke mening, det var bare en liten skrivefeil, en person hadde forstått dette med engang». Mengden tekst denne oppgaven produserer virker for flere deltakere som overflødig i forhold til spørsmålet som blir stilt. Bjørn forteller oss at det første svaret fra chatboten forklarte allerede godt nok på spørsmålet som ble stilt. Han kommenterer også at teksten virker noe smått som fører til at han velger å zoome inn for å få teksten større. Anne



forteller at det virker som chatboten prøver å komme henne i forkjøpet ved å svare på alt rundt temaet arbeidsavklaringspenger. Hun sier videre at det ikke er nødvendig å forklare definisjoner da hun synes chat og NAV er komplisert nok fra før. Hun tilføyer at chatboten «spyr ut alt om temaer selv om jeg ikke har bedt om alt». Einar presiserer at den heller kunne vært konkret i svaret istedenfor å overforklare. Kari forteller at det er veldig mye å tekst å lese og selv om det virker oversiktlig er hun usikker på om hun hadde klart å komme seg igjennom all teksten. Kari prøver å rette opp setningen etter at en feil blir funnet, men får ikke tid til å rette opp før chatboten produserer flere svar i chatbotgrensesnittet. Hun forteller at hun heller ville ha tatt kontakt med et menneske for å finne ut om man har krav på arbeidsavklaringspenger med nedsatt arbeidsevne.

Mari og Bjørn synes det er greit å få svaralternativer (se figur 4.12). Anne sier at NAVs chatbot burde ha brukt svaralternativer oftere. Som ved Skatteetatens chatbot ser vi usikkerhet rundt språket i svaralternativene. Bjørn lurte for eksempel på hva svaralternativet «medlemskap» betyr i kontekst av arbeidsavklaringspenger. Mari og Anne opplever at chatboten avbryter leseprosessen da de startet å lese. Anne kommenterer at «... man begynner å lese på det først så avbryter chatboten deg».

**Oppgave 8:** La oss si at det ikke var nok informasjon om AAP og du ønsker å finne ut mer om dette. Du prøver derfor å komme i kontakt med et menneske. Hvordan går du frem for å chatte med et menneske om AAP ved å bruke chatten?

Einar	Mari	Kari	Anne	Bjørn



**Figur 4.13.** NAVs chatbot viser et svaralternativ.



**Figur 4.14.** NAVs chatbot predikerer intensjonen feil, men gir rett valg.

Alle deltakerne skrev til chatboten for å komme i kontakt med et menneske. Anne hadde forstått at man kunne navigere via chatboten for å komme i kontakt med et menneske ettersom at hun hadde lest dette før man startet chatten. Flere brukte enkle stikkord, for eksempel skriver Kari kun «veileder», og sier at dette er hun kjent med. Noen tilfeller av enkle skrivefeil forekommer i denne oppgaven, men chatboten gir svar som oppfattes som tilstrekkelig. Igjen stusser Mari over svaralternativet og begrepet «chat» og spør: «betyr det å snakke med et menneske da?» (se figur 4.13).

Når Bjørn skriver «kontakte behandler» klarer ikke chatboten å vurdere hva han spør om (se figur 4.14). Likevel får han et svaralternativ for å komme i kontakt med et menneske. Han opplever det som irriterende at «... man må være spesifikk, men samtidig ikke for spesifikk når man skriver til chatboten».

	Siste inntrykk av NAVs chatbot?
Bjørn	Han opplever chatboten som litt rotete. Han opplever mye av informasjonen som unødvendig, og mener det burde ha vært kortet ned og mer konkret. Han forventet at chatboten skulle være dum, men hadde håpet at den skulle være smartere.
Anne	Anne ønsker at chatboten hadde vært tydeligere i språket. Hun føler at mye av teksten er «copy paste» fra lovdata. Avataren til chatboten gir henne ingen assosiasjon til en maskin, noe som hun mener kan føre til misforståelser
Einar	Han opplever chatbot som en bra teknologi, men at den må bli bedre på skrivefeil og eventuelt komme med forslag.
Mari	Mari likte at chatboten hadde et litt større brukergrensesnitt. Hun sier hun ikke er interessert i å skrive til den, og ville heller snakke med et menneske eller brukt Google søk.
Kari	Hun synes at denne chatboten var oversiktlig, men forstår at man ikke kan få alle svar fra en chatbot. For eksempel vil et menneske kunne forklare lover og regler enklere.

Om du kan forestille deg den perfekte chatboten, hvordan hadde den vært?

Avslutningsvis forhørte vi oss om hvordan deres «perfekte» chatbot hadde vært, og eventuelt hvilken funksjonalitet de ville ha satt pris på. Innebygd funksjonalitet for innlesing og opplesing av tekst blir sett på som en funksjonalitet som kunne gjøre opplevelsen bedre i møte med en chatbot. Tre deltakere ønsker seg opplesing av tekst, enten at alle eller enkelt meldinger blir lest opp. Anne sier «... hadde vært veldig greit om det ble lest opp, en innebygd funksjon i chatbot hadde vært greit», og uttrykker i den sammenheng at en opplesningsfunksjonalitet med kunstig stemme er noe hun ikke liker. Einar og Mari ønsket seg funksjonalitet for å lese inn teksten. «Hadde vært fint å kunne lese inn og ut».

Det blir igjen nevnt at chatboten bør bruke mer svaralternativer, gi bedre svar, ha flere språk, ha en høyere toleranse for skrivefeil og mindre menneskelig etterligning.

## 4.3 Analyse av brukertesting

Denne kapittelet presenterer temaer vi har kommet frem til med en analyse. Temaene adresserer tilgjengelighet. Mer spesifikt er tilgjengelighet: teknisk tilgjengelighet slik at man har tilgang til og kan oppfatte informasjonen i løsningen og brukervennlighet slik at informasjonen i tjenesten er intuitiv, effektiv og forståelig. Den siste delen vil ta for seg temaet menneskelig kontakt.

### 4.3.1 Chatbot brukergrensesnittet

Generelt blir det uttrykt at de er positive til hvordan chatbotene er utformet, men vi legger merke til noen utfordringer i chatbotenes grensesnitt. Størrelsen i begge grensesnittene, det vil si bredde og høyde, oppleves som liten i størrelse i forhold til plassen som er tilgjengelig på nettsidene. Mari forteller oss at chatbotene avbryter leseprosessen og at hun gjerne skulle hatt mulighet til å forstørre vinduet, slik at man kunne se mer av teksten som ble produsert. Bjørn opplever også teksten som liten og velger å zoome inn. Dette går på bekostning av at han ser mindre av teksten som blir produsert av chatboten. I flere oppgaver blir det registrert at det var nødvendig for deltakerne å scrolle i grensesnittet for å lese alt innhold chatbotene hadde produsert.

### 4.3.2 Problemer i forhold til navigasjon

De første oppgavene handlet om å finne frem til chatbotene. Det vi ønsket å undersøke var om chatbotene hadde en teknisk tilgjengelighet for deltakerne. Dersom de ikke finner chatboen kan de heller ikke bruke den, dermed kan man si at det ikke er en teknisk tilgjengelighet til stede. Generelt fant alle deltakerne frem til Skatteetatens chatbot, men vi registrer at flere ikke umiddelbart fant frem til den. Anne, Bjørn og Einar kommenterer at den burde vært lettere tilgjengelig, og flere forteller at de hadde forventet å finne den nede på høyresiden i en pop-up av noe slag. NAVs chatbot er tilgjengelig i en navigasjonsmeny nederst på NAVs nettside med en lenke kalt «Kontakt oss». Men i motsetning til Skatteetatens chatbot klarer ikke to av deltakerne å navigere seg til NAVs chatbot.

### 4.3.3 Er chatten styrt av et menneske eller en maskin?



**Figur 4.15.** Avataren til Skatteetatens og NAVs chatbot. Fra venstre er Skatteetatens avatar.

Deltakerne opplever at NAVs chatbot prøver å etterligne et menneske. Bjørn opplever dette som unødvendig når man vet at det er en chatbot. Anne synes også det var litt rart at chatboten hadde navnet «Frida», og at den representerte seg selv med en kvinnelig avatar. Hun ville heller ønsket å se en robot avatar med musefletter, fordi det ikke er et menneske. Generelt blir den kvinnelige avataren ikke assosiert som et bilde som representerer en maskin, men heller et menneske blant deltakerne. Mari kommenter at hun kanskje hadde trodd dette var et menneske om det ikke hadde stått «Frida chatbot» øverst i chatbotgrensesnittet. Anne er også i en oppfatning om at dette kan mistolkes: «Du er ikke et menneske [...] jeg vet ikke om jeg ville forstått at dette var en robot med det første». Men ved spørsmål om de tror det er et menneske eller en chatbot de prater med, svarer samtlige at de tror det er en chatbot.

Derimot i møte med Skatteetatens chatbot var det tydelig for deltakerne at de snakket med en maskin og ikke et menneske. Det blir kommentert at dette er tydelig fordi chatboten har en avatar som etterligner en robot.



### 4.3.5 Svaralternativer og lenker

Det er fem funn knyttet til svaralternativer og lenker: (1) svaralternativer oppleves som et kognitivt hjelpemiddel, (2) svaralternativer som er like er vanskelig å velge mellom, (3) svaralternativer er ikke alltid riktig i forhold til brukerintensjonen, (4) lenker oppleves som unødvendig, og (5) lenker i chatbotene ser ikke ut som lenker for alle.

Skatteetatens chatbot benyttet ofte svaralternativer for å veilede deltakerne til svaret de lette etter. Svaralternativene som blir presentert i chatbotene er knapper som er ganske likt utformet, men de skiller seg ut ved at de benytter ulik fargeprofil, ikoner og visuelle hint. Generelt informerte deltakerne om at svaralternativer var en funksjonalitet de likte at chatbotene benyttet. Vi blir her forklart at svaralternativer hjelper med tanke på skrivevanskene deltakerne har. Dersom de kan unngå å skrive er det mindre kognitiv belastning, i tillegg til at de unngår å bli misforstått. «Det fine med svaralternativer er at man slipper å skrive», sier Einar.

En observasjon vi gjorde var at Bjørn valgte å benytte tidligere svaralternativer til å utføre en senere oppgave i brukertesten. Anne forteller oss at hun heller ønsker å velge blant svaralternativer, enn å få svar med mye unødvendig informasjon.

Det var noen brukervennlighetsproblemer som ble registrert med svaralternativer. Vi blir informert om at svaralternativene kan være vanskelig å velge mellom. Det blir beskrevet blant flere at en usikkerhet og forvirring oppstår når svaralternativer oppleves som like. Det kan derfor også knyttes til temaet vanskelig språk som vi forklarer nærmere i kapittel 4.3.7. En konsekvens av usikkerheten er at deltakere ikke nødvendigvis vil ta et valg ettersom det kan være feil. Eksempler med svaralternativer som virker like hverandre er som følger (1) endre inntekt og endre lønn, (2) registrere lønn og endre feil lønn, og (3) lønn og næringsinntekt.

I tilfeller Skatteetatens chatbot hadde predikert feil, hendte det at den svarte med svaralternativer som ikke sto i overensstemmelse med deltakerens forventning. Flere deltakere opplevde at de ble sittende fast når dette skjedde. Anne sier at det burde vært en mulighet å fortelle chatboten at ingen av svaralternativene er riktige. En deltaker spurte om man kunne stille ett nytt spørsmål, eller om man da tar med gammel kontekst inn i samtalen, mens andre

trodde samtalen var helt ferdig. Anne sier hun ville ha avsluttet samtalen her og begynt på nytt.

Bruk av lenker er noe flere av deltakerne synes er greit at chatbotene anvender. Når lenker ble trykket på i brukertesten var det derimot noe kritikk i forhold til hva lenken tilførte av informasjon. Mari sier «Jeg har fått svar på det jeg lurer på, og jeg trenger ikke mere informasjon». Kari mener informasjonssiden som lenkene fører til presenterer alt for mye tekst, noe som hun ser på som negativt.

#### 4.3.6 Skrivefeil og chatbot predikasjon

Det er flere funn i forhold til skrivefeil og forståelse: (1) chatbotene klarer å tolke noe skrivefeil, typisk når det er skrivefeil i nøkkelord, (2) bruk av stikkord fremfor hele setninger ser ut til å øke sannsynligheten for å bli forstått, (3) når chatbotene ikke klarer å predikere rett, leter de etter nøkkelord for å gi generelle oppfølginger vedrørende nøkkelordet, og (4) deltakere har en forventning om at enkelte skrivefeil skal bli forstått.

Begge chatbotene har en toleranse for skrivefeil. Det ser ut til at toleransen er høy når det benyttes nøkkelord, som for eksempel «skattekort» og «arbeidsavklaringspenger». En høy toleranse betyr at ordene blir i større grad rettet. Generelt hadde alle deltakerne en eller annen form for skrivefeil underveis i brukertesten. Det skilte seg derimot i hvor mye skrivefeil som forekom mellom de ulike deltakerne, men det virker som om dette ikke hadde noe spesiell innvirkning på om chatbotene predikerte rett eller feil. Deltakerne oppfatter ofte selv når de har skrivefeil, og det var flere tilfeller der de poengterte egne feil. I disse tilfellene ble det gjort få forsøk på å rette eller finn ut hva som var feil setningen. Dermed ble meldinger sendt selv når de var klar over at den inneholdt en eller flere feil.

Det var en forventning fra deltakerne om at chatbotene skulle forstå ord som hadde mindre skrivefeil. Det skapte derfor frustrasjon når det de tolket som en liten skrivefeil ikke ble forstått. Et eksempel er der Einar skrev ordet «poten» istedenfor «posten». Chatboten responderer med svar om at den ikke forstår. Han opplever dette som frustrerende for han mener feilen var liten og ubetydelig: «Det var jo ikke mye skrivefeil, så skjønner ikke at den ikke forsto det». I tilfeller der chatboten ikke predikerer rett, og de omformulerer spørsmålet, legger vi merke til at deltakerne bruker merkbart med tid for å prøve å rette feilen. I de fleste tilfeller endte det med at hele meldingen ble omformulert fremfor at skrivefeil ble rettet. Det



var også tilfeller der Google ble benyttet for å finne ut hvordan man skrev et ord eller en setning riktig.

Det er utfordrende å få et klart bilde over hva chatbotene klarer å tolke av skrivefeil, men ut ifra denne brukertesten ser det ut til chatbotene i større grad oppfatter nøkkelord som har dobbeltkonsonant- eller orddelingsfeil. Vi tolker det som et brukervennlighetsproblem at chatbotene er inkonsekvent når de oppfatter skrivefeil eller kontekst, og når de ikke gjør det. Det skaper en usikkerhet for deltakerne om hvor problemet ligger, og det uttrykkes tvil blant deltakerne om det er de eller chatboten som er problemet.

Det ser ut til at begge chatbotene bruker to ulike fallback-mekanismer når de ikke klarer å vurdere intensjonen fra ytringene til deltakerne. En fallback er som tidligere forklart en reserve melding. Chatbotene gir enten (1) svaralternativer basert på nøkkelord fra ytringen, eller (2) en melding om at den ikke forsto spørsmålet, og at deltakeren måtte prøve å spørre på en annen måte. Som nevnt i avsnittet om svaralternativer (se kapittel 4.3.5), var det ikke alltid en overenstemmelse mellom intensjonen og svaralternativene. Flere deltakere reagerte negativt på at chatboten prøvde å «sikre» seg med svaralternativer. Eksempelvis i oppgave 3, der deltakerne skulle finne svaret på hvordan man endret lønn på skattekortet, får to deltakere svaralternativer som er generelle temaer rundt skattekort, men ingen av alternativene omhandler det å endre lønnen / inntekten. Det er tydelig at chatboten ikke klare å predikere intensjonen rett, men fanger opp nøkkelordet «skattekort» i brukerutsagnet. Det var flere lignende tilfeller som dette, og opplevelsen med å bli misforstått førte til forvirring og frustrasjon: «Chatboten forstår ikke nok og kommer med svaralternativer», «Chatboten tror den forstår, men gjør ikke det», «Slitsomt å ikke bli forstått, selv om det er en chatbot», «Skrivefeil gjør chatboten usikker så den gir svaralternativer» var noen av tilbakemeldingene fra deltakerne da dette inntraff.

En observasjon vi gjorde var at deltakerne lærte hvordan de burde ordlegge seg for å øke sannsynligheten for at chatbotene predikerte brukerintensjonen rett. Underveis i testen ble flere av deltakerne oppmerksom på at det lønner seg å skrive i stikkordsform fremfor fulle setninger. Kari sier underveis i en av oppgavene «Jeg pleier å skrive setninger, men det er bedre med stikkord» og konkluderer med at man må skrive i stikkord for å bli forstått. Dette var en tilpasning flere av deltakerne tok innover seg. Bruk av korte setninger eller stikkord er noe begge chatbotene henter til ved oppstart. NAVs chatbot skriver at: «Jeg svarer best på korte spørsmål» og Skatteetatens chatbot sier «Gi meg et par stikkord om hva du lurer på».

Basert på hvordan testen utviklet seg, og at ingen av deltakerne kommenterte dette, er det tydelig at disse meldingene ikke ble oppfattet eller tolket dithen at man burde forenkle setninger ned til stikkordsform.

#### 4.3.7 Mengde tekst og Vanskelig språk

Som nevnt var formatet med at mange meldinger ble produsert etter hverandre et brukervennlighetsproblem for leseprosessen til deltakerne. En annen faktor som var med på å gjøre leseprosessen utfordrende var det språklige innholdet og mengden tekst i svarene fra chatbotene.

Dette var spesielt gjeldene for NAV. Vi observerte flere tilfeller av at deltakerne valgte å hoppe over tekst om det var svaralternativer tilgjengelig. Eksempelvis i oppgave 7, som var en av oppgavene som produserte mye tekst i tillegg til svaralternativer, ble det observert at deltakerne ignorerte store deler av teksten for så å gå direkte til svaralternativene. Når vi spør Anne om hvorfor hun valgte å ikke lese teksten forteller hun: «Det er litt overflødig tekst. Trenger ikke å ha med alt som står her. Den øverste boksen forklarer egentlig godt, så man kunne bare hatt med lenke om man vil lese mere». Lignende utsagn var gjentakende i flere av oppgavene rundt NAVs chatbot, Det blir etterlyst presise svar og et ønske om at chatboten ikke detaljert forklarer alt som omhandler et tema. Dette er noe lik problematikk vi har sett tidligere. Chatbotene prøver å «sikre» seg ved å gi mest mulig informasjon om et nøkkelord det fanger opp fra en brukerintensjon den sannsynligvis ikke klarte å vurdere rett. Dette medfører mye tekst som for deltakerne oppleves som irrelevant innhold i forhold til spørsmålet som ble stilt, som igjen førte til frustrasjon og misforståelser.

Det blir også poengtert at mye tekst oppleves masete, slitsomt og overflødig. Det uttrykkes et ønske om at flere av svarene burde ha tilbudt svaralternativer fremfor tekst, slik at de kan navigere til riktig og ønsket innhold selv. Anne sier «Det virker som den prøver å komme meg i forkjøpet og spyr ut temaer. Jeg vil heller ha svaralternativer». Fra de innledende brukerintervjuene hadde vi forstått at flere hadde et anstrengt forhold til det å lese tekst. Vi ble fortalt at desto lengre en tekst er jo vanskeligere blir det å holde konsentrasjonen oppe, og flyten i det de leser.

I tillegg til mye tekst var språk og terminologi noe som også bidro til at innholdet i svarene skapte usikkerhet og frustrasjon. Mari forteller oss at det er en fallgrube med vanskelig språk

og fremmedord for dyslektikere, for det skaper tvil om man har forstått innholdet riktig eller ikke. Vi opplever at deltakerne i større grad reagerer på språk og terminologi hos NAV. Dette blir uttrykt både muntlig fra deltakerne i tillegg til at vi observerer at oppgaveløsningen ble påvirket av dette. Når en deltaker ikke forsto innholdet i teksten, visste de heller ikke om informasjonsbehovet ble dekket av svaret de fikk.

#### 4.3.8 Menneskelig kontakt

Temaet menneskelig kontakt er et aspekt som vi ikke anser som et brukervennlighet- eller tilgjengelighetsproblem, men som kan være av interesse i et bredere perspektiv.

I intervjuene vi hadde med deltakerne før brukertesten fortalte flere av deltakerne at de foretrakk å snakke med et menneske enn en chatbot. Dette ser vi er et tema som også gjentar seg flere ganger i løpet av brukertesten. Samtlige forklarer at de ville ha kontaktet et menneske når (1) svaralternativene opplevdes som like, (2) når chatbot ikke ga svaret deltakeren lette etter, eller (3) når deltakere ble presentert med mye tekst. Vi ser også at chatbotene gjentatte ganger spør deltakere om de vil snakke med et menneske dersom chatbotene med sannsynlighet ikke er i stand til å predikere intensjonene rett.

Mari var blant flere som opplevde at svaralternativene virket like og at hun hadde lyst til å gi opp. Da vi spurte hva hun ville ha gjort her, sier hun: «.. ville ringt et menneske nå, dette blir for komplisert». Hun forklarte videre at hun liker å skrive lengre setninger supplert med kontekst for å frem hennes situasjon. Hun sa at hun vanligvis gjør det når hun skriver til mennesker, men å skrive til chatbot er annerledes fordi chatbot har visse begrensninger. Dette oppleves som et «stressmoment» for Mari.

I flere tilfeller ga ikke chatbotene svaret deltakere søkte etter. Bjørn for eksempel, opplevde det som irriterende når han prøvde flere ganger uten å få riktig svar. Deltakere ga ofte i den sammenheng uttrykk for at de ville ha kontaktet et menneske for å løse problemet.

Kari hadde erfaringer med chatboter fra før, og forteller oss at hun hadde egne teknikker for å komme i kontakt med et menneske. For i de fleste tilfeller ville hun heller snakke med et menneske enn en chatbot, særlig når det kommer til skatt. Dette ble vist da hun bare skrev veileder i NAVs chatbot for å komme i kontakt med et menneske. Hun opplevde også at chatboten produserte mye tekst da hun spurte om hun hadde rett på arbeidsavklaringspenger.

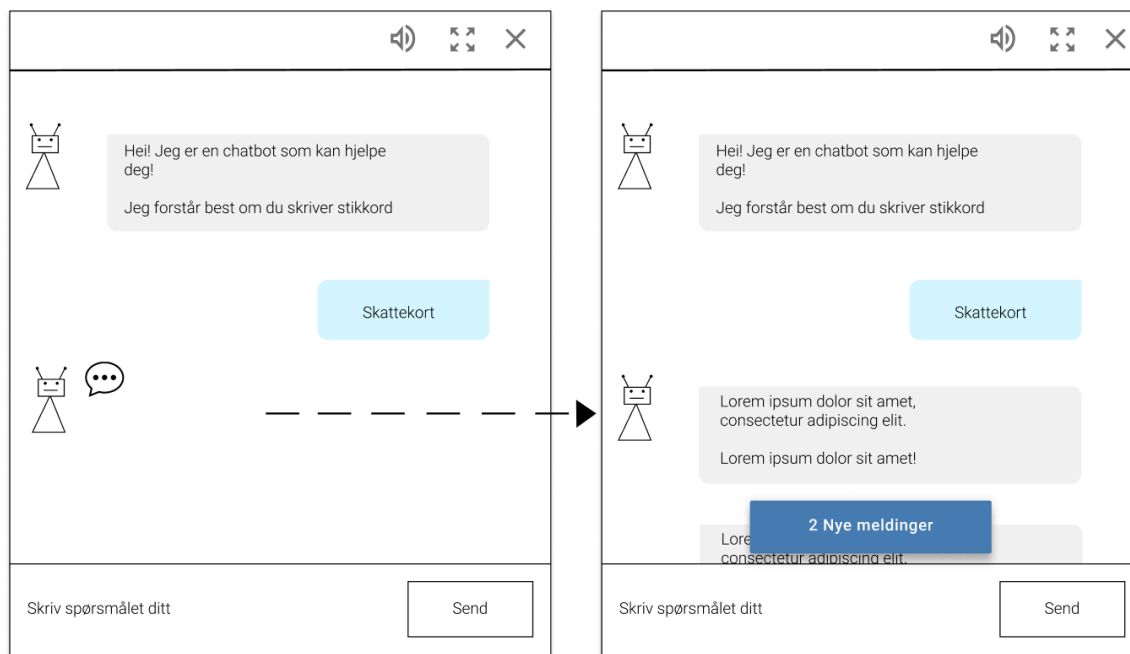
Som et siste inntrykk til NAVs chatbot går hun inn på dette temaet igjen, og forteller at hun «.. ville heller snakket med et menneske som kan snakke enkelt om lover og regler». Det å komme i kontakt med et menneske er et inntrykk som deles blant flere av deltakerne da de fortalte oss hvordan de hadde opplevd å ha en samtale med de ulike chatbotene.

## 5. Designforslag

Dette kapitlet presenterer hvordan vi har kommet frem til designforslag som kan forbedre tilgjengeligheten i chatbotene. Først vil vi presentere prototyper, og deretter noen tekstlige anbefalinger. I denne fasen har vi valgt å konvergere mot noen av konseptene som vi mener er realistisk å implementere teknisk, og som kan øke tilgjengeligheten i chatbotene for dyslektikere, men også andre brukergrupper. Som beskrevet i metodekapitlet informeres forslagene av skisser, bakgrunns litteratur, innsikt av eksperter og analysekapitlet. Analysen identifiserte en rekke problemer til forbedring for deltakerne, men noen av funnene er mer eller mindre knyttet til utfordringer dyslektikere har. Funn som vi mener kan virke utfordrende på bakgrunn av deres funksjonsnedsettelse er for eksempel avbrytelse i leseprosessen, mye tekst eller at skrivefeil ikke alltid blir forstått av chatbotene.

### 5.1 Prototyper

#### 5.1.1 Forhindre avbrytelser i lesingen



**Figur 5.1.** Prototype av en chatbot som forhindrer avbrytelse i leseprosessen.

Lesing er en tidkrevende prosess og en avbrytelse i leseprosessen kan virke til å være en belastning for dyslektikere.

Generelt kan dyslektikere trenge å lese på nytt for å gi mening til det de har lest, og antas enten til å være raske unøyaktige lesere eller trege nøyaktige lesere (Shaywitz & Shaywitz, 2005). Ifølge modellen «Simple view» er lesing som nevnt avkodning pluss forståelse, der dyslektikere har problemer med avkodingsferdighetene. Disse problemene karakteriseres ofte som langsomt, anstrengt eller unøyaktig ordavkodning. Det er derfor viktig å la dyslektikere få nok tid å lese og ikke bli avbrutt. Nok tid til å lese blir blant annet adressert i WCAG 2.0 retningslinjene 2.2 under prinsippet opererbar.

Hendelsesforløpet til en avbrytelse er som følger: (1) chatboten sender et svar til brukeren, (2) deltakeren begynner å lese chatbotens svar, (3) chatboten sender et nytt svar til deltakeren, og (4) det første svaret som deltakeren leste på har nå blitt dyttet ett hakk opp i chatbotgrensesnittet og brukeren blir dermed avbrutt (se figur 4.16).

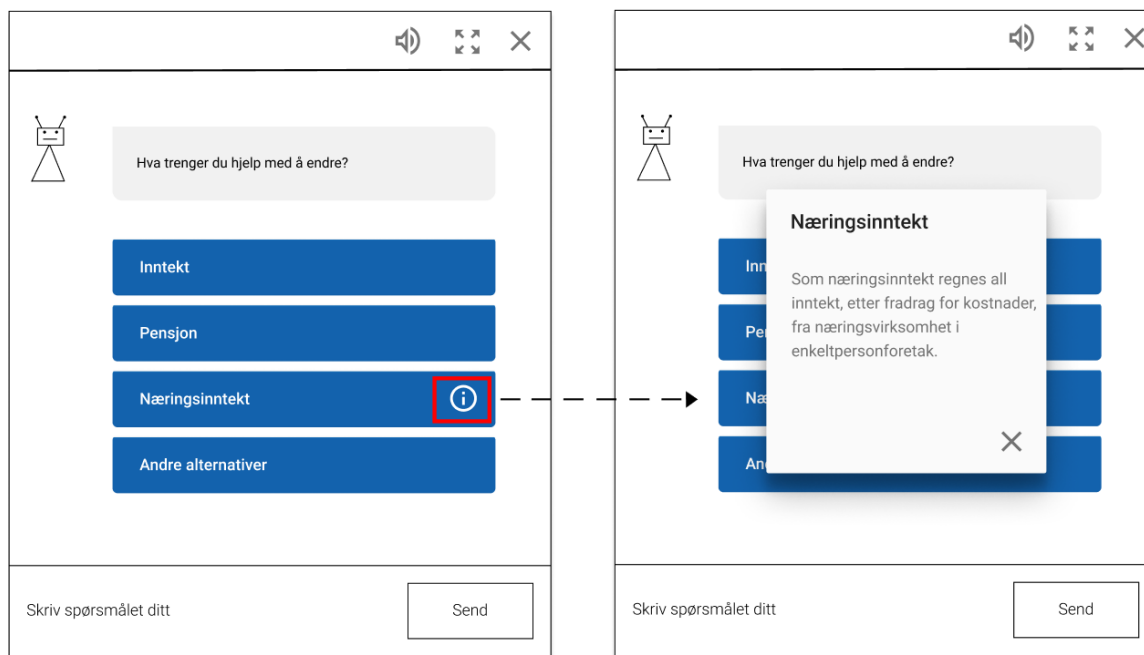
Det er flere endringer som kan vurderes for å redusere avbrytelse i leseprosessen. Kristian Simonsen i Funka sier til oss at avbrytelse i leseprosessen: «Ligger i chatbot sin natur. Det kommer flere meldinger når chatbot predikerer. Man kan vurdere en lang sammenhengende melding, men det kan bli veldig mye». En avbrytelse ville ikke vært til stede dersom chatbotene hadde forholdt seg til å sende kun én melding, noe som vi kommer tilbake til når vi snakker om mengde tekst chatbotene produserer. En sammenhengende melding kan bli mye som Simonsen sier, og det er anbefalt å forholde seg til å dele opp informasjon og presentere det stykkevis (Sandnes, 2011), slik som responsene fra chatbotene presenteres i dag med chatmeldinger.

En annen endring som kan redusere avbrytelse er å justere hastigheten på sekvensen av svar fra chatbotene. NAVs chatbot er vesentlig raskere til å svare deltakeren. I denne chatboten gis det lite rom for å begynne å lese et svar før den er ferdig med å svare deltakeren, og det kan være årsaken til at vi så færre tilfeller av avbrytelser her. En sekvens med raske svar kan imidlertid påvirke brukeropplevelsen negativt. Deltakere opplever responsen til chatboten som hakkete, forvirrende og overveldende.

Det tredje forslaget er å «stoppe» samtalen, og dermed forhindre problemet med hastighet og avbrytelse i begge chatbotene (se figur 5.1). Med å stoppe samtalen mener vi at resten av

chatbotenes svar dyttes ned. I dag er oppførselen at meldingene dyttes opp når de svarer. Designforslaget vil føre til at samtalen stopper ved første produserte melding, og en bruker kan lese uten at chatboten avbryter. Videre kan en indikasjon slik som en knapp med tekst fortelle at chatboten har produsert flere meldinger.

### 5.1.2 Forklar begreper og vanskelige ord



**Figur 5.2:** Prototype av en chatbot som forklarer begreper og vanskelige ord.

Det er noen problemområder i bruken av svaralternativer slik de forekommer i chatbotene. Det oppsto situasjoner der deltakerne var usikre rundt hvilket svaralternativ som passet deres intensjon. Som nevnt i analysen ble det kommentert at alternativene fremstår like, samtidig som noen alternativer virket fremmed og uforståelig. Begreper eller vanskelige ord kan være problematisk for dyslektikere. I et eksperiment med 23 dyslektikere og en kontrollgruppe sammenlignet Rello, Baeza-Yates, Dempere-Marco og Saggion (2013) tekster der innholdet i teksten ble erstattet av synonymer som forekommer hyppigere i daglig tekst og tale, samtidig som de kortet ned ordene. Ved bruk av ord som forekommer hyppigere ble lesbarheten for dyslektikere forbedret, mens kortere ord økte forståelsen av det som ble lest.

Når det omhandler begreper eller vanskelige ord ser vi det som fordelaktig om enklere synonymer til ordene blir benyttet. Med enklere, mener vi ord som forekommer hyppigere i

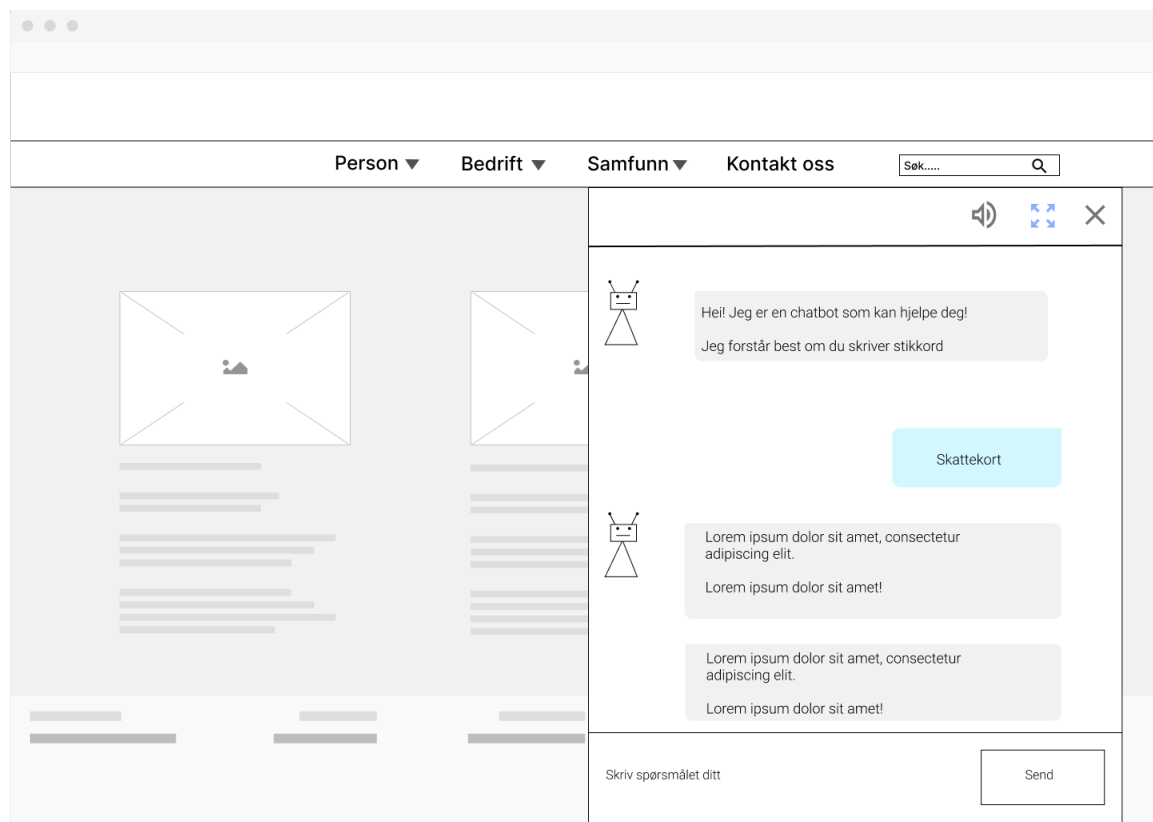
daglig tale og tekst. WCAG 2.0 sin suksesskriterie 3.1.5 under prinsippet forståelig virker til å være relevant her, og handler om at nettinhold som krever leseferdigheter over grunnskolenivå skal ha en alternativ versjon tilpasset forventet lesenivå etter endt grunnskole.

Det språklige innholdet i etater som NAV og Skatteetaten ofte må forholde seg til kan være vanskelig å endre på grunn av trygdefaglige og juridiske årsaker. Men vi mener det kan være mulig å forenkle språket uten at det går på bekostning av dette. For eksempel gjennomførte Norsk regnesentral i 2018 en eksperimentell utprøving for å forbedre NAVs kommunikasjon, ved å gjøre informasjonen som er på nett lettere å lese og forstå. De testet med opprinnelig tekst fra NAV, og en endret variant som hadde et vesentlig enklere språk, uten at det gikk på bekostning av kvaliteten på informasjonen. Resultatene fra testen viste at deltakerne deres synes de endrede versjonene var klart bedre når det gjaldt leselighet, innhold og forståelse. Deres deltakere satte også pris på at lenker i teksten som kunne forklare vanskelige ord eller spesifikke «NAV-begreper», som eksempelvis «uttaksgrad» og «pensjonsgrad» (Solheim & Halbach, 2018).

Der det ikke er mulig for etatene å bytte ut vanskelig ord eller begreper med forenklede synonymer, ser vi det som hensiktsmessig at ordet eller begrepet blir forklart. Vårt forslag er å benytte et klikkbart informasjonsikon som vil gi en forklaring av ordet i en pop-up melding i grensesnittet (se figur 5.2).



### 5.1.3 Større grensesnitt, mindre scrolling



**Figur 5.3.** Prototype av en forstørret versjon av chatbotgrensesnittet.

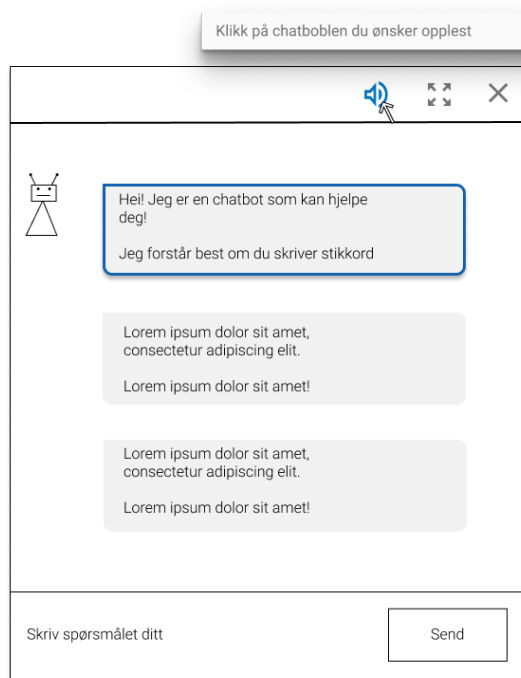
Det forekommer ofte at chatbotene produserer mye tekst. Mye tekst oppleves til tider overveldende og unødvendig blant deltakerne. Lesing er en bred sammensatt prosess av kognitive ferdigheter som dyslektikere har vansker med. Lengre segmenter av tekst er ikke kun utfordrende å lese for dyslektikere, de er vanskelig å lese for folk flest (Sandnes, 2011). Mye tekst kan i tillegg være vanskelig å håndtere dersom man har redusert kapasitet i arbeidsminne. Redusert arbeidsminne er sett på som en av de vanligste og fremtredende funksjonsnedsettelsen hos dyslektiker, og er sett på som en viktig komponent for lesing, dekoding av ord og oppfattelse av lengre tekster (Smith-Spark & Fisk, 2007). Mye tekst kan derfor være vanskelig å håndtere for dyslektikere, og flere deltakere benytter støtteprogram for lesing når de vet at de skal lese mye tekst.

Når mye tekst blir produsert av chatbotene, det vil si rundt 3 - 4 meldinger, må deltakerne scrolle opp i chatbotgrensesnittet for å se teksten som ble produsert. Et lignende funn blir også rapportert om i Funkas brukertest av chatboter: «Jeg syntes det kom mye informasjon på en gang som gjør at jeg må scrolle og lese meg opp på lite relevante meldinger ofte» (Simonsen,

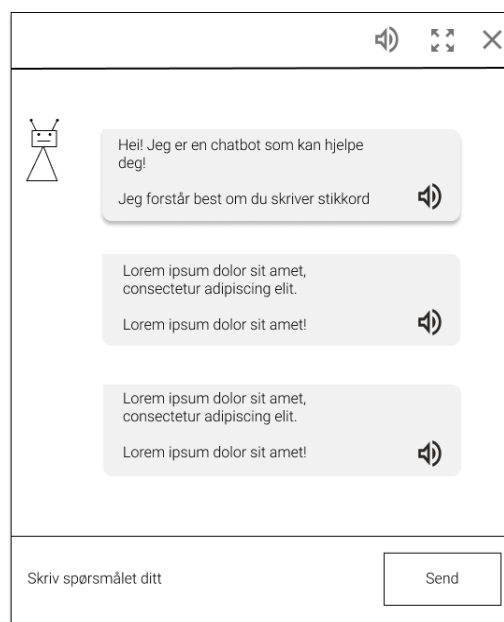
Jacobsen, Kling, Henstad & Laurin, 2020, s. 11). Et forslag, og et ønske, blant deltakerne i vår studie er å lage brukergrensesnittet større for å forhindre dette problemet. Det blir uttrykt at dette kan hjelpe i forhold til at man slipper å scrolle opp «hele tiden» i begge grensesnittene.

Med et større grensesnitt kan man potensielt se flere av meldingene som chatbotene produserer uten at man må scrolle opp i grensesnittet. Et større grensesnitt kan også gagne flere som har vansker med syn eller motorikk (Sandnes, 2011). I et designforslag som omhandler grensesnittets størrelse, må man være påpasselig med hvordan dette utformes. Sandnes (2011) hevder at mennesker generelt leser kolonner som er smale mer effektivt enn brede. Når man ikke har begrensninger i bredden kan det derfor bli utfordrende å lese teksten produsert av chatbotene. Vi mener derfor at høyden er det man bør fokusere på å gjøre større, og ikke bredden. Forslaget til et større grensesnitt vises i Figur 5.3.

#### 5.1.4 Les teksten opp for brukeren



**Figur 5.4.** Prototype av tekst-til-tale funksjonalitet i øvre del av chatbot.



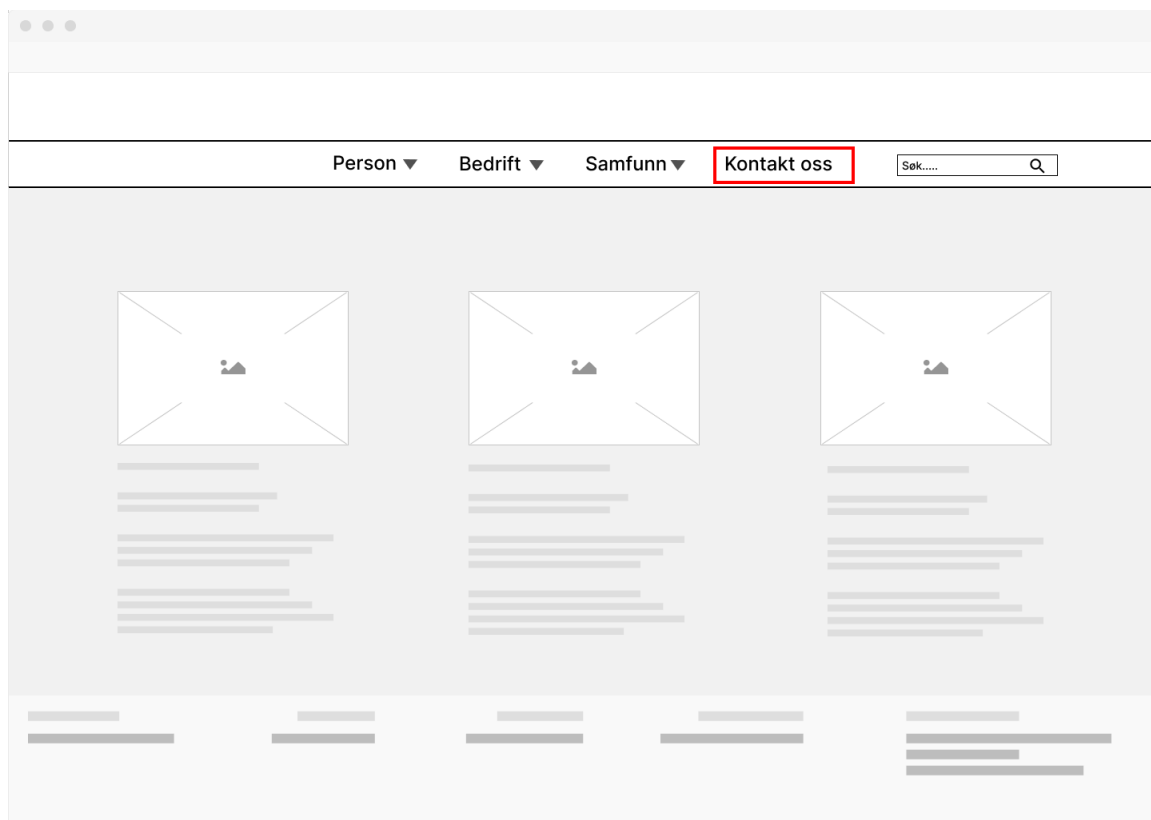
**Figur 5.5.** Prototype av tekst-til-tale funksjonalitet i chatbot meldingene.

Et uttrykt ønske er å ha innebygd funksjonalitet for opplesning av meldinger eller spesifikke ord. Deltakerne har tilgang til støtteprogrammer for lesing, men disse programmene ble i større grad benyttet i sammenheng med utdanning eller når man produserte en formell tekst. Vi opplever terskelen for at de vil benytte støtteprogrammer, for eksempel for lesing i en

chatbot som stor. Dersom en funksjon for opplesning hadde vært til stede sier de fleste deltakerne at de ville brukt den.

Fra brukertesten blir opplesning av tekst ønsket av samtlige deltakere. Vi anser at slik funksjonalitet kan ha stor nytteverdi for dyslektikere, spesielt når chatbotene produserer mye tekst. Det uttrykkes et ønske om å ha valgmuligheter i opplesningsfunksjonaliteten da de ikke nødvendigvis vil ha all tekst opplest. Eksempelvis var muligheten for å selv velge hvilke meldinger de ville ha opplest foretrukket. Av dette er vårt forslag å implementere funksjonalitet for opplesning av tekst, som presenteres i to forskjellige prototyper (se figur 5.4 og 5.5). Sandnes (2011) anbefaler at man benytter flere modaliteter i brukergrensesnitt for å redusere avhengigheten av det tekstlige innholdet for dyslektikere. Funksjonalitet for opplesning av lyd har en mulighet til å ikke ha en syntetisk stemme, som er vanlig i støtteprogrammer. En menneskestemme ser vi som teknisk mulig ettersom svarene blir manuelt lagt inn i chatbot programvaren av chatbot-trenere i Skatteetaten og NAV, og svarene kan derfor være forhåndsinnspilt. Det er derimot en vurdering om man ønsker å overføre menneskelige egenskaper til chatbotene (antropomorfisme).

### 5.1.5 En mer tilgjengelig navigasjon



**Figur 5.6.** Prototype av en klarer navigasjon på NAVs nettside.

En av utfordringene deltakerne hadde var å navigere seg til NAVs chatbot, der to deltakere ikke klarte å finne frem til den. Når chatboten ikke er tilgjengelig, er det ikke mulig å betjene den. Vi mener dette er et teknisk tilgjengelighetsproblem. Navigering på nett er ikke en ukjent utfordring for dyslektikere. En brukervennlighetstest utført av The Disability Rights Commission (2004) fant ut at dyslektikere mislyktes med 17% av de 20 oppgavene gitt til dem. Hvor to av hovedproblemene som ble identifisert omhandlet at navigering på nett kunne oppleves som forvirrende og desorienterende. Et designforslag som anbefales er å lage en konsistent og klar navigasjonsstruktur slik at dyslektikere kan navigere seg enklere gjennom nettinholdet (The British Dyslexia Association, 2018).

Deltakerne forteller av erfaring at de forventet et chatbot-ikon nederst til høyre på begge nettsidene. Dette er imidlertid problematisk sett i et større tilgjengelighetsperspektiv sier Kristian Simonsen fra Funka til oss. Bruk av chatbot-ikon for å starte en chat på nettsiden kan føre til at det blir vanskeligere å finne frem til chatbotene for mennesker som benytter seg av skjermlesere. Det kan bli vanskeligere fordi det kan bli mange elementer skjermleserbrukere

må tabbe seg igjennom på tastatur for å navigere seg til en chatbot. Simonsen sier samtidig at det også ofte er dårlig kodet, og da er det utfordrende å finne frem til chatboten i leseliste.

Våre resultater viser at deltakerne klarte å navigere seg til Skatteetatens chatbot. Denne chatboten er navigerbar øverst i navigasjonsmenyen «Kontakt oss» på nettsiden til Skatteetaten. Det viste seg at de fleste forstod at chatbotene var under navigasjonslenken «Kontakt oss» dersom den var øverst på nettsiden. Denne plassering mener vi kan oppleves som et tidlig blikkfang når man besøker nettsiden. Et designforslag er derfor å flytte navigasjonslenken som leder til chatboten til den øverste navigasjonsmenyen på NAVs nettside, som vist i figur 5.6.

## 5.2 Anbefalinger

### 5.2.1 Mer bruk av svaralternativer

Selv om det var noen problemer med svaralternativer, ble de i stor grad satt pris på av deltakerne. Årsaken til at svaralternativer ble foretrukket kan ha bakgrunn i at dyslektikere har nedsatt kapasitet når det kommer til arbeidsminne. Svaralternativer kan derfor ha en positiv effekt med tanke på dette. Blant annet fordi svaralternativer reduserer mengden tekst. Når brukeren får svaralternativer må man ikke nødvendigvis formulere et spørsmål i respons til chatboten, noe som igjen kan redusere problematikken som omhandler skrivning og skrivefeil. Vi mener derfor at tilbudet om svaralternativer kan redusere noe av den kognitive belastningen og effektivisere gjennomføringen, og dermed øke brukervennligheten for dyslektikere. Av den grunn ser vi det som fordelaktig å øke forekomstene av svaralternativer i chatbotene. Et forslag er å tilby en mulighet for å velge flere alternativer dersom de som ble presentert ikke er korrekt i kontekst av intensjonen til brukeren. Ved å få flere runder med alternativer kan sannsynligheten øke for at alternativ som omhandler intensjonen blir funnet. Det må også nevnes at fallback-mekanismer som hindrer en «loop» av alternativer bør implementeres. Det kan for eksempel være etter at et antall forslag med svaralternativer vil man bli bedt om å omformulere spørsmålet på nytt.

### 5.2.2 Vurder en annen toleranse for stavefeil

Chatbotene sine interne metoder for å rette stavefeil til riktig stavemåte lykkes ikke alltid. På bakgrunn av at rettskrivningsalgoritmene ikke klarer å rette til riktig stavemåte hender det at brukerintensjoner ikke blir forstått, og i verste fall blir intensjonene misforstått. Skrivning oppleves noe utfordrende blant deltakerne. De har typisk en fornemmelse om at det de har skrevet inneholder en eller flere feilstavelser, men hvor eller hva stavefeilen er, finner de som regel ikke klarhet i. Etersom de ikke vanligvis benytter seg av lese- og skrivestøtteprogram i lignende situasjoner, blir det ønsket at chatbotene hadde dette innebygget for å motvirke den negative effekten av feilstavelser. Det blir direkte og indirekte foreslått blant deltakerne at chatbotene bør ha en høyere toleranse for stavefeil, en synlig stavekontroll og mulighet for innlesing av tekst.

Et system som har en toleranse for stavefeil betyr at den aksepterer at brukere skriver feil og prøver ved hjelp av diverse pre-prosesseringsmetoder (slik som en stavekontroll) å predikere hva brukeren egentlig spør etter. Et søkesystem som antas til å ha en høy toleranse for stavefeil er Google sin søkemotor, fordi den i stor grad aksepterer stavefeil, og i tillegg klarer å hente relevante dokumenter i forhold til søkestrengen. For å gjøre søkesystemer mer tilgjengelig anbefaler Berget og Sandnes (2016) å ha en høy toleranse for stavefeil som både gagnar dyslektikere og personer uten dysleksi. Studien til Berget og Sandnes handler om søkemotorer og ikke chatboter. Deres funn er nødvendigvis ikke helt overførbart til chatboter siden det antageligvis er mer rom for stavefeil i søkesystemer, ettersom de leverer mange flere dokumenter i responsen enn en chatbot, som må forholde seg til å respondere med et par meldinger. Likevel, våre funn gir en indikasjon på at feiltoleransen i Skatteetatens chatbot kunne vært høyere. For eksempel blir ord med stavefeil slik som «elektroniks», «poten» og «tiljengelig» ikke rettet av chatboten (se tidligere figur 4.7). I motsetning kan funn fra L. Simonsen (2019, s. 62) gi en indikasjon på at feiltoleransen i NAVs chatbot kunne vært lavere. Hun undersøkte chat-logger i chatboten til NAV, og fant for eksempel at stavekontrollen «retter» ordet «foresatt» til «fortsatt» som fører til at brukeren blir misforstått. For å gjøre chatboter mer tilgjengelig med tanke på toleranse for feilstavelser trenger man mer innsikt for å vurdere om en høy eller en annen toleranse er bedre egnet til å rette stavefeil.

Man kan også vurdere andre funksjoner enn en toleranse for stavefeil for å motvirke den negative effekten av stavefeil. Berget og Sandnes (2016) studie viser for eksempel at en rød-

strek under feilstavelser eller en autocomplete til å fullføre setninger kan hjelpe skriveprosessen for dyslektikere.

### 5.2.3 La en bruker korrigere stavefeil

I tillegg til å justere toleranse for stavefeil kan man også implementere spesifikke funksjoner som kan redusere krav til riktig stavemåte for dyslektikere. Berget og Sandnes (2015, s. 23) foreslår for eksempel å gi en tilbakemelding på hvilken del av spørringen som ikke er i samsvar, for å redusere krav til riktig stavemåte for dyslektikere. Dette forslaget er igjen ment for søkesystemer, likevel ser vi det ble benyttet i en av chatbotene vi undersøkte når det forekom feilstavelser.

Eksempelvis hendte det at Skatteetatens chatboten svarte med at den ikke forstod brukerintensjonen og hvilket ord den ikke klarte å vurdere. Vi så at denne svarmekanismen, som i utgangspunktet kan redusere belastningen av stavefeil, virket mot sin hensikt. Dette ble vi oppmerksomme på da Mari skrev til chatboten: «bestille elektronisk skattekort». I dette eksempelet er et ord feilstavet og chatboten svarer deltakeren tilbake: «Jeg tror jeg forstår det du skrev, men skjønnte ikke elektronisk». Mari tror videre at det er hennes tur til å rette stavefeilen, og svarer kun med «elektronisk», men nå forstår ikke chatboten hennes intensjon lenger. Problemet som oppstår, er at hun tror at hun kan rette stavefeilen i det ene ordet, noe som ikke er tilfellet. Dette er en ytring fra chatboten som «foreslår». En ytring som foreslår, inngår i talehandlingskategorien «assertive» til Searle (1976). Dersom chatboten blir forstått til å foreslå, men faktisk ikke foreslår, er det noe man bør endre på for å forbedre den naturlige samtale i chatboten. Dette kan endres enten ved at chatbot-trenere ikke lager predefinerte svar som foreslår, eller ved at det blir mulig at brukeren kan korrigere stavefeilen. Det sistnevnte er en teknisk implementasjon som vi ser på som naturlig og konvensjonell i en chat mellom mennesker. Det går også under prinsippet om felles enighet. For å forhindre at chatbotsamtaler ikke ender med misforståelser er det viktig at chatboten anerkjenner brukeren ved å repetere det som blir sagt.

### 5.2.4 Redusere unødvendige chatbot meldinger

Deltakere opplever flere av meldingene produsert av chatbotene som unødvendig. I oppgave 6 i brukertesten opplevde flere deltakere at det første svaret fra NAVs chatbot allerede dekker deres informasjonsbehov. De resterende fire meldingene ble sett på som unødvendig i forhold

til deres informasjonsbehov. Som nevnt i forrige avsnitt, når man designer samtaler i en chatbot kan det være et godt utgangspunkt å se på hvordan mennesker kommuniserer med hverandre. At informasjon oppleves som unødvendig for deltakerne bryter med samarbeidsprinsippet til Grice (1975) om å gi så mye informasjon som nødvendig, men ikke mer. Når prinsippet brytes kan det hende at mennesker blir frustrert eller forvirret, for man antar at en samtale vil ha rett mengde med informasjon. En tidligere nevnt undersøkelse viser i sammenheng at redusering av unødvendig informasjon og innhold på NAVs nettside gir en positiv effekt for deres brukere (Solheim & Halbach, 2018).

Vi anbefaler å undersøke om man kan redusere unødvendige meldinger som chatbotene produserer. Dette innebærer at systemkomponentene i chatbotene blir bedre trent til å predikere de rette brukerintensjonene, men å trene NAVs chatbot til å predikere rett kan være en utfordring. NAV ivaretar hele befolkningen, og må derfor ta høyde for å snakke med mennesker i individuelle og ulike situasjoner. Dette fører til at chatboten kan ha vanskeligheter med å predikere, og sikrer seg derfor med mye unødvendig tekst. Skatteetaten sier at de allerede vet hva brukerne lurte på ved det første spørsmålet. Dette fordi de vet at konteksten er skatt, og her er det klare lover og regler som skal følges. De har derfor et bedre utgangspunkt enn NAV til å predikere rett, og kan svare brukeren med nødvendig innhold.



## 6. Evaluering

Dette kapitlet presenterer evalueringen vi hadde med to dyslektikere, som også var med i brukertesten. Hovedpoenget med å evaluere med deltakere var for få spille inn de prototypene vi hadde designet, samt hva vi hadde valgt bort, slik at vi kunne få tilbakemeldinger om hvordan de oppfattet disse endringene i chatboten.

### 6.1 En mer tilgjengelig navigasjon

Vi valgte å flytte «Kontakt oss» til en mer synlig i navigasjonsmenyen på nettstedet for å skape en tydeligere og klarere navigasjonsstruktur (se figur 5.6). Begge deltakerne kunne ikke huske at det var et stort problem å finne chatboten, men anser det som en fornuftig implementasjon for å øke tilgjengeligheten. Det ble også ytret at selve lenke-teksten kunne være enda mer beskrivende, for eksempel «chat med oss» eller lignende, da «kontakt oss» ikke trenger å være åpenbart med tanke på å starte en chat. Et annet poeng som ble diskutert, og for øvrig ble nevnt i brukertestene og utvikling av prototypene, er å benytte et chatbot-ikon som skal representere chatboten i et av hjørnene på nettsiden. Fra denne ideen diskuterte vi videre problematikken rundt å benytte slike ikoner, da det kunne føre til tilgjengelighetsproblemer for blant annet blinde som er avhengig av skjermlesere. Når det kommer til andre tilgjengelighetsproblemer poengterte en deltaker «... når man først har startet chatten er det også viktig for meg at den følger etter når man går til en ny side». Dette er noe chatbotene ikke gjør i dag, og om man følger en lenke fra chatbotgrensesnittet eller på nettstedet, vil ikke chatboten «følge» etter. Dette ble sett på som problematisk da deltakeren mente at man da kan bli forvirret over hvor chatten ble av, noe som kan forårsake at man må starte samtalen på nytt.

### 6.2 Større grensesnitt, mindre scrolling

Når vi evaluerte muligheten for å forstørre chatbotgrensesnittet ble dette i all hovedsak sett på som positivt (se figur 5.3). Begge deltakerne synes det er forståelig og konvensjonelt med å bruke et ikon som forstørrer og minimerer grensesnittet. De så på det som en god måte å få en bedre oversikt over en større del av samtalen, i tillegg så de denne endringen som en fordel om det er mye tekst, og at et stort grensesnitt forhindrer eller reduserer at de må scrolle i samtalen for å få med alt innholdet.

### 6.3 Forhindre avbrytelser i lesingen

Ved evaluering av hvordan man kunne forhindre avbrytelser i leseprosessen ble prototypen som omhandlet dette diskutert (se figur 5.1). De så det som fordelaktig at samtalen «stopper» når chatboten velger å skrive flere svar, men de var noe usikker på bruk av knapper til å fortsette samtalen. En av deltakerne synes knappene dekket over noe tekst, og ville ha flyttet slik funksjonalitet bort fra teksten så det ikke hindret lesing. Igjen kom forslaget om å få hele svaret inn i en og samme melding, men dette kan være problematisk, noe som vi har diskutert tidligere i kapittel 5.

### 6.4 Forklar begreper og vanskelige ord

Problematikken rundt vanskelig språk og begreper var noe begge deltakerne husket godt fra brukertesten. Et av designforslagene var å forklare vanskelig ord og begreper etatene benytter som ikke lar seg forenkle med synonymer (fig 3). Dette forslaget ble godt mottatt, men det ble poengtert viktigheten om at forklaringene burde være enkle og konsise, da en forklaring av ord fører til mere tekst for de å lese.

### 6.5 Mer bruk av svaralternativer

Fra brukertestene var det tydelig at svaralternativer forbedret brukeropplevelsen til deltakerne. Designforslaget med å øke forekomster av svaralternativer samt ha muligheten for å få flere slike alternativer om de som ble presentert ikke passet, ble godt mottatt av evalueringdeltakerne. En av deltakerne sier i den forbindelse «.. svaralternativer er noe folk kjenner til, og man trenger ikke opplæring». Det ble også nevnt at det er typisk for dyslektikere å prøve på å forklare hva man ønsker med lange setninger, og ikke i stikkordsform, noe som i stor grad virker mot sin hensikt når man skal kommunisere med en chatbot. De trodde derfor at de ville ha større utbytte av chatbotene dersom man kunne redusere innskrivning av tekst, og at å øke forekomster av svaralternativer var en god måte for å løse dette.

### 6.6 Les teksten opp for brukeren

Fra analysen så vi et gjentagende ønske om innebygd funksjonalitet for opplesning av tekst (se figur 5.4 og 5.5). I evalueringen viste vi frem flere prototyper til hvordan slik

funksjonalitet kunne fungere. For eksempel så de det som fordelaktig om man selv kunne avmerke hva som ble lest opp. Det kunne være enkeltord, eller deler av chatbotsvaret, da de mente at det kan være enkeltord som er problematiske, og at man derfor ikke trenger å få opplest alt innhold. Som tidligere får vi bekreftet at eksterne støtteprogrammer ikke er noe de vanligvis bruker når de navigerer på nettet eller skriver i en chat. Var det derimot innebygd i chatboten, var ikke terskelen for å bruke slik funksjonalitet stor. De likte også ideen om chatbotsvaret kunne bli lest inn av et menneske, da de ikke var spesielt begeistret for syntetisk-stemme i programmene de var med.

## 7. Diskusjon

I dette kapitlet diskuterer vi studiets forskningsprosess, svakheter og styrker. Kapitlet er delt inn i tre deler. Den første delen omhandler diskusjon av metodevalg, og da spesielt tanker rundt brukertesting med dyslektikere. Den andre temaet handler om designforslagene med tanke på universell utforming. Til slutt ser vi på litteratur som har blitt brukt og på chatbot teknologien i et bredere perspektiv med tanke på menneskelig kontakt.

### 7.1 Brukertesting som metode

Innledningsvis i studiet ser vi at «chatbot og universell utforming» er et uberørt og mangelfullt tema innen forskning, samtidig som man vet lite om hvilke utfordringer dyslektikere har med nettbaserte grensesnitt (se kapittel 2.4.2). Vi så det derfor som hensiktsmessig å benytte en metode som ga oss rike beskrivelser fra personer med dysleksi, slik at vi kunne få en forståelse over hva som allerede fungerte bra og hva som kunne forbedres når det gjelder et tilgjengelighetsperspektiv for dyslektikere.

Brukertesting som metode er ofte benyttet iterativt i en designprosess for å eliminere problemer tidlig, underveis og i slutten av prosessen. Chatbotene som ble testet er allerede utviklede grensesnitt som har gjennomgått flere iterasjoner med forskjellige former for testing før de ble implementert på etatenes nettsider. Dette gjør at vår brukertest i større grad slipper å ta hensyn til systemfeil og andre «bugs» som man typisk opplever i en tidlig design og utviklingsfase. Vi mener valget med å benytte brukertesting som formativ evalueringsmetode har muliggjort å se på hvilke aspekter ved grensesnittene i nåværende tilstand oppleves som bra eller dårlig, og hvorfor det var slik for våre deltakere.

### 7.2 Bruk av støtteprogram

Lese og skrive støtteprogrammer blir ofte diskutert i relasjon til tilgjengelighet (Berget, 2016, s. 122). Slike programmer har diverse funksjonaliteter slik som stavekontroll, talesyntese, ordfullføring og ordbok som kan være nyttig for personer med dysleksi. Støtteprogrammer ble vanligvis ikke brukt av deltakerne når det kom til aktiviteten av å chatte på datamaskin. At støtteprogrammer ikke blir brukt av dyslektikere er noe vi også finner igjen i andre studier (Berget, 2016; Simonsen, Jacobsen et al., 2020). Flere av deltakerne i vår studie var kjent med diverse støtteprogrammer, og brukte det ofte i formelle sammenhenger. To deltakere

informerte om at de benyttet stavekontroll på mobil. Dette studiet har imidlertid ikke utforsket tilgjengelighet på andre skjermstørrelser enn en datamaskin. Resultatene i denne studien ville trolig vært annerledes dersom de hadde benyttet talesyntese, stavekontroll og andre støtteprogrammer.

Vi valgte å tilrettelegge for bruk av eget utstyr og støtteprogram ved gjennomføring av brukertesten, og vi ser i ettertid at dette kunne ha blitt problematisk med tanke på antall deltakere som var med. For eksempel, om tre av fem deltakere hadde valgt å bruke støtteprogram, kunne dataen vært noe mer sprikende i forhold til tilgjengelighetsproblemene som ble funnet. De med støtteprogram kunne for eksempel redusert store deler av skrivefeilsproblematikken i motsetning til de som ikke brukte støtteprogram. Nå var det slik at alle deltakerne selv valgte bort støtteprogram, og vi fikk også bekreftet at de ikke ville ha benyttet det i en reell brukssituasjon med en chatbot.

Fuglerud (2011) argumenterer for at man skal teste opp mot et så reelt bruksscenario som mulig, og må for eksempel la personer med funksjonsnedsettelse bruke de tekniske hjelpemidlene som de vanligvis har tilgang til. Vi er stort sett enige i dette poenget, men vi ser også viktigheten av å være konsise i en test med få brukere. Dette er for å ikke få mangelfull data på grunn av ulike forutsetninger fra de som deltar i studie. Dette mener vi spesielt gjelder når man ser på en spesifikk og sårbar gruppe mennesker. Vi mener det derfor er viktig å avklare faktorer som bruk av hjelpemidler på forhånd.

Som vurdert og diskutert i kapittelet om etiske betraktninger ble deltakerne informert i rekrutteringsprosessen om at vi ønsket å teste chatboter. Vi var derimot tilbakeholdne om hvilke chatbotløsninger som skulle testes, og vi mener at forhåndsinformasjonen vi ga ikke hadde stor innvirkning på resultatene av testen. Det er uansett verdt å nevne da det kan være en mulighet for at deltakerne har trukket slutninger eller «øvd» på lignende chatbotløsninger før testen. Det er derimot naturlig å tenke at resultatene rundt deres oppfattelse om chatten var menneske- eller maskinstyrt ble påvirket av forhåndsinformasjonen.

### 7.3 Har vi tilstrekkelig antall deltakere?

Som nevnt er det uenighet i hvor mange brukere man bør benytte i en brukertest. Nielsen (2000) sin tommelfingerregel om at fem brukere vil finne 85% av brukervennlighetsproblemer i et grensesnitt møter noe motstand i forskningsmiljøet. Spool og

Schroeder (2001) testet nettsider med 49 brukere, og sier de fem første bare fant 35% av brukervennlighetsproblemene. I utgangspunktet hadde vi ønsket å rekruttere flere deltakere til testing, men vi opplevde at dataen vi genererte fra brukertestene inneholdt problemer som var gjentakende og det var derfor få unike brukervennlighetsproblemer som oppsto. Denne erfaringen støtter også teorien om at dyslektikere blir rapport som velegnede testere til å identifisere problemer i et brukergrensesnitt (Dixon, 2007). På bakgrunn av dette, og våre erfaringer mener vi at det har vært tilstrekkelig med fem deltakere til brukertesten i dette studiet.

## 7.4 Rekruttere deltakere fra en sårbar gruppe

Selv med rekrutteringshjelp fra Dysleksi Norge opplevde vi det som utfordrende å rekruttere deltakere til å delta på brukertesting. Denne problemstillingen belyste vi til organisasjonen som informerte oss om at det kan være mye traumer i forbindelse med det å ha skrive- og lesevansker, og en brukertestsituasjon kan være en belastning for enkelte. Dette er også en problemstilling Fuglerud (2014) skriver om i sin doktoravhandling: I et prosjekt som omhandlet nettbasert stemmegivningstjeneste, opplevde hun det som en større utfordring å rekruttere personer med kognitive nedsettelse som dysleksi, fremfor andre funksjonsnedsettelse. Hun peker på at en grunn til dette kan være at det å ha kognitive utfordringer, ofte i motsetning til psykiske og motoriske nedsettelse, er forbundet med stigma. I den sammenheng, var en viktig del av vår rekrutteringsprosess å bruke mye tid på å informere om hva vi ønsket å gjøre, hvorfor de var viktige bidragsytere og at det var grensesnittet vi testet og ikke deres evner.

Det har med stor sannsynlighet ikke vært mulig for oss å rekruttere nok deltakere til brukertest og intervjuer uten hjelpen fra Dysleksi Norge, og de har vært en viktig bidragsyter i dette studiet. Til tross for dette må det nevnes at det å rekruttere via en organisasjon gjør at man mister noe av kontrollen i rekrutteringsarbeidet.

Dysleksi Norge som organisasjon når ut til mange, og har personlig kjennskap til enkeltindivider som kunne ha lyst og interesse i å delta i forskningsprosjektet vårt. Ulempen er at vi blir avhengig av Dysleksi Norges interesse og mulighet til å følge opp i rekrutteringsprosessen med interessentene som melder seg. Vi har noen eksempler på, og full forståelse for, at deltakere har trukket seg i prosessen før testingen startet. Dysleksi Norge sier

at testing av chatboter kan oppleves som vanskelig og skummelt for dyslektikere, noe som har vært en utfordring for de i arbeidet med å rekruttere for oss.

## 7.5 Arbeidet med designforslag og universell utforming

Hvilke funn og ideer vi har valgt å gå videre med har vært en balansegang mellom hva vi antar er teknisk mulig og hva som fjerner eller reduserer de største problemene som ble oppdaget. Vi har vært noe tilbakeholden med å designe prototyper med mye kompleks funksjonalitet. Dette fordi brukergrensesnittet til et chatbotgrensesnitt har sine begrensninger, spesielt med tanke på grensesnittets størrelse. Vi mener også at løsningen ikke nødvendigvis behøver mye ny funksjonalitet, men heller designløsninger som kan implementeres uten at det går på bekostning av det gjenkjennbare i et grensesnitt for chatting. En interaksjon i chat mellom menneske-menneske er etter vår oppfattelse konvensjonelt og intuitivt, og mye ny funksjonalitet kan forringe dette. Ifølge Sandnes (2011) utnytter et effektivt grensesnitt brukernes hukommelse med hjelp av konvensjoner. Vi har derfor tatt bevisste valg om å være tilbakeholden på å utforme designforslag med mye, og kompleks funksjonalitet.

Noen av designforslagene ble konseptualisert i form av lavoppløselige prototyper som vi benyttet for å evaluere med deltakerne. Selv om det ble poengtert under evaluering at dette var enkle prototyper og ikke en ferdig utviklet løsning, kan den visuelle fremstillingen ha påvirket deres oppfatning av forslagene. Vi mener likevel at slike visuelle fremstillinger gir en bedre forståelse av designforslagene, og noe som bidro til en konstruktiv diskusjon underveis når vi evaluerte med deltakerne.

Et designforslag, tekst-til-tale, er en funksjonalitet deltakerne ser et behov for. Først og fremst fordi de ikke liker å lese tekst, men er den også innebygd i chatboten er ikke terskelen for å bruke slik funksjonalitet like stor. Ideen om at teksten ble lest opp av et menneske istedenfor en maskin var også ønskelig. Tekst-til-tale er til å finne i støtteprogram, og man bør derfor vurdere hvorvidt vårt forslag er bedre for tilgjengeligheten enn tilgjengelige støtteprogrammer. Tekst-til-tale i NAVs chatbot er ikke urealistisk å anta bli implementert i fremtiden. Trender i teknologi frem mot 2030 som NAV selv anser som betydningsfulle for deres digitale videreutvikling, blir tekst-til-tale nevnt som en mulig teknologi som kan gagne flere av deres brukere (NAVs omverdensanalyse, 2019).

Tekst-til-tale er imidlertid ikke til nytte for personer som er døve, og kan dermed ha en motsatt effekt i form av opplevd tilgjengelighet dersom dette blir implementert i en hovedløsning. En tilleggsfunksjon som ikke inkluderer alle, møter motstand i universell utforming miljøet. For om man skal benytte universell utforming som et begrep, forutsetter man at løsningens design gjør at alle kan benytte den (Fuglerud et al., 2015). Selv om målet og motivasjonen med universell utforming virker fra et utilitaristisk perspektiv forståelig, er det tvetydig for oss hvordan man skal oppnå det i praksis. På en side blir det sagt at universell utforming er et mulig mål siden alle problemer i et grensesnitt kan enten løses eller forbedres (Lazar, 2007). På en annen side blir dette sett på som umulig. Harper (2006, s. 112) hevder blant annet om man prøver å imøtekomme alle sine behov så er det alltid noen som stiller motstridende behov som igjen fører til ingen løsning.

Fuglerud (2014, s. 157) argumenterer at det ikke nødvendigvis må være en konflikt i inkluderende design å tilby noen alternative løsninger så lenge likeverd og like muligheter tas i betraktning, og at det gis alternativer for alle brukere. Ettersom alternativet for flere døve er tekst, antar vi at deres muligheter til en viss grad er tatt i betraktning i chatbotene. Man kan anta at dyslektikere har «like» muligheter som andre med sine støtteprogram, men når støtteprogram ikke blir brukt i kontekst av å chatte stiller de ikke med de samme mulighetene. Tekst-til-tale vil vi også argumentere for at flere brukergrupper kan ha nytte av. Personer med dyskalkuli og sensoriske funksjonsnedsettelse slik som blinde og svaksynte blant annet. Lesing er ikke unikt utfordrende for dyslektikere, for alle mennesker kan oppleve konsentrasjonsproblemer som gjør det vanskelig å avkode og forstå tekst. Fuglerud et al. (2015) hevder at en spesiell tilrettelegging ikke kan kalles universell utforming, men siden funksjonaliteten vi her foreslår kan gagne flere ser vi ikke på det som en spesiell tilrettelegging for én enkelt gruppe.

I arbeidet med prototypene og anbefalingene har vi støttet oss på litteratur om dysleksi og tilgjengelighet. En slik fremgangsmåte kan kritiseres for å ikke være i samsvar med tankegangen om universell utforming, da noe av teorien er rettet mot en homogen gruppe. Selv om dette kan være problematisk med tanke på universell utforming, så mener vi at flere av våre designforslag er i samsvar med å øke tilgjengelighet for flere brukergrupper. Det inkluderer: mer tilgjengelig navigasjon, et større grensesnitt, redusere unødvendige meldinger og forklare begreper og vanskelige ord. Berget et al. (2016, s. 455) hevder at lignende forslag ikke bare kan øke tilgjengeligheten for dyslektikere, men mest sannsynligvis for alle brukere.



Påstanden er forankret i WCAG 2.0 sine retningslinjer og i Dixons (2007, s. 10) argumentering om at symptomer som er assosiert med dysleksi kan også forekomme for mennesker flest. For eksempel kan stress føre til skrivefeil og dette kan videre antyde at dyslektikere er egnet til å identifisere problemer som kan være til gode for et bredt mangfold.

Universell utforming innebærer teknisk tilgjengelighet og brukervennlighet. Den tekniske tilgjengeligheten kan til en viss grad testes med automatiske testverktøy og gjennom ekspertevalueringer for å kontrollere at chatbotene er i samsvar med standarder og retningslinjer slik som WCAG 2.0. I dette studiet finner vi noen relevante retningslinjer til å samsvare med designforslag som blir nevnt over, samt andre forslag slik som forhindre avbrytelser i lesingen, forhindre stavefeil og la en bruker korrigere stavefeil:

- 2.2. Gi brukerne nok tid til å lese og bruke innhold.
- 3.3. Hjelp brukere med å unngå feil og å rette opp feil.
- 2.4. Gjør det mulig for brukerne å navigere, finne innhold og vite hvor de befinner seg.
- 3.1. Gjør innholdet leselig og forståelig.

Selv om dette er relevante retningslinjer er det også kjent at flere av kriteriene under disse er på et høyt nivå og bør revideres for å bedre imøtekomme dyslektikere (Berget et al., 2016). Som nevnt er det forventet at utgivelsen av WCAG 2.1 i nærmeste fremtid blir å følge og vi identifiserer blant annet at suksesskriteriet 1.4.12 («Tekst avstand») i dette dokumentet kan hjelpe dyslektikere å lese tekst.

En tidligere nevnt studie viser at automatiske evalueringer ikke er tilstrekkelig for å finne avvik i suksesskriterier til WCAG 2.0 (Vigo et al, 2013). Som tidligere nevnt hadde vi en automatisk evaluering for å bedre sikre at tilgang til bruk var mulig (se kapittel 3.3.3). Rapporten som ble generert av testverktøyet indikerer at tid, navigering og farge kunne ha potensielle avvik i chatbotene i forhold til suksesskriterier i WCAG (se vedlegg C og D). Hva som var relevante retningslinjer for dyslektikere i den genererte testrapporten så vi i forhold Berget et al. (2016) sin studie som omhandlet relevante WCAG suksesskriterier for dyslektikere. De fleste avvikene fra testverktøyet var kun indikasjoner som vi selv kontrollert manuelt. For eksempel kontrollerte vi om den var opererbar med tastaturnavigering. Med en manuell kontroll av chatbotene fant vi ingen avvik som vi mente ville forhindre tilgang til bruk. Når vi betrakter testrapporten, og vår manuelle gjennomgang av denne i ettertid, ser vi at det ikke ble funnet noen av avvikene som dyslektikere fant i brukertesting. På

bakgrunnen av denne erfaringen mener vi det er viktig å involvere brukere, og at en automatisk evaluering og en gjennomgang av avvikene ikke er tilstrekkelig for å finne ut hvor skoen trykker.

Vi velger å belyse noe av WCAG i denne studien da flere forbinder dette som relevante for universell utforming av IKT-løsninger i Norge. Dette kommer til uttrykk i statusmålinger, tilgjengelighetsrapporter, forskning og lovgivning (Digitaliseringsdirektoratet, 2018; Simonsen, Jacobsen et al., 2020; Forskrift om universell utforming av IKT-løsninger, 2013).

Når det er sagt har ikke dette studiet hatt fokus på standarder og retningslinjer innenfor universell utforming. Vårt fokus har derimot vært på brukere i en brukersentrert designprosess for å utforske dyslektikere sine opplevelser for å øke tilgjengeligheten i chatbotene. Begnum (2016) argumenterer også for at det synes å være to forskjellige kulturelle holdninger i universell utforming. Der man på den ene siden er opptatt av testbare standarder og retningslinjer som brukes i automatiske tester og ekspert evalueringer. Og på den andre siden er man opptatt av brukere og brukeropplevelse. Det sistnevnte ble undersøkt i vår studie og anses å være forskningsmessig i samsvar med fortolkende kvalitative paradigme problemstillinger.

## 7.6 Samtaledesign som teori

Dersom målet er å utforme en naturlig samtale mellom menneske-maskin sier flere forskere at et godt utgangspunkt til design er å se på hvordan mennesker snakker med hverandre (Jurafsky & Martin, 2019; Moore & Arar, 2018; Følstad & Taylor, 2020). Grice (1975) sine samarbeidsprinsipper og Searle (1976) sine kategorier for talehandlingsteori så vi på som nyttig teori fra samtaledesign til å forklare problemer i chatbotene der samtalen opplevdes unaturlig eller ikke-menneskelig.

En gjennomføring av brukertesting i en simulert kontekst mener vi imidlertid begrenser problemene man finner i samtaler i forhold til en naturlig setting. Til sammenligning, finner L. Simonsens (2019) analyse av chatbot-logger i NAV flere eksempler der samtalen ikke fungerte i NAVs chatbot. Noe som kan indikere at Simonsens setting med data fra naturlige samtaler kan være mer passende dersom man vil benytte samtaledesign som grunnlag for teori.

Samtalene som folder seg i chatbotene i dette studiet ser ut til å følge en del av teorien vi har presentert i forhold til å utforme en naturlig samtale i chatbotene. Det vil si at chatbotene følger flere samarbeidsprinsipper og konsepter som initiativtaking, turtaking og felles enighet. For eksempel tar chatbotene det første initiativet til å starte samtalen, i hovedsak venter de på riktig tur og tidspunkt til å ta ordet, de erkjenner når de ikke predikerer rett og de formidler når prediksjonen blir en sannsynlig antagelse. At samtalene for oss virker til å være noe naturlig kan ha med å gjøre at chatbotene er sluttprodukter som har gått gjennom flere sykluser av evalueringer. At chatbotene følger flere konvensjoner og prinsipper innen samtaledesign ser vi som en styrke i chatbotene, og kan være en årsak til at flere deltakere også ble positivt overrasket over hva noe av denne teknologien var i stand til å formidle.

## 7.7 Fra menneskelig kontakt til chatbot

Begrepet brukervennlighet kommer til uttrykk for å være et viktig prinsipp i universell utforming for flere forskere (Sandnes, 2011; Fuglerud, 2014; Berget, 2016). Det vil si at løsningen skal være forståelig, intuitiv, effektiv og enkel å bruke. Brukervennlighet er på mange måter viktig for en vellykket opplevelse, og de fleste designforslagene våre adresserer problemer med brukervennlighet for dyslektikere. Som nevnt argumenterer Nielsen (2012) at brukervennlighet er en forutsetning for å overleve på nettet, og at brukere vil forlate nettsiden dersom de opplever den som vanskelig å bruke. Nielsens argument er vi delvis enig med, men tenker at påstanden gjelder i større grad for kommersielle enn offentlige tjenester.

Kommersielle nettsider har ofte konkurrenter, og samtidig er man typisk bruker av egne interesser. I det offentlige er det annerledes. Man har ikke et valg når det gjelder etater som NAV og Skatteetaten, og man kan ikke velge en «konkurrent». En bruker kan ikke bare å forlate NAVs nettside dersom brukeren er uføretrygd, for i ytterste konsekvens får ikke brukeren trygd. For å ta Nielsen (2012) sitt argument videre, tenker vi heller at mangel på brukervennlighet fører til en redusert tillit til de offentlige etatene.

En videreutvikling av den digitale forvaltningen som ble skrevet om innledningsvis, har skapt noen bekymringer både hos brukerne og fagfolk (Westin, 2017). I sin kronikk ser Westin med bekymring på risikoen for at tilliten mellom partene blir svekket ved at «en dør til velferden» blir så til ingen dør. NAV henviser til at Stortinget har vedtatt at digital kommunikasjon skal være hovedregelen, men Westin setter spørsmålsteget ved intensjonen: «Kan det være at NAV

nå tar grundig feil?». Etatene effektiviserer, det blir stadig redusert i skattekontor, publikumsmottak og personlig oppmøte ved disse kontorene. Det er deres digitale tjenester som har blitt hovedregelen for kommunikasjon med brukere. Chatbot blir presentert som det første en bruker som søker informasjon og hjelp vil møte i det offentlige. Det er derfor et paradoks at dette er kontaktformen deltakere i dette studiet helst vil unngå å benytte. Dysleksi er ikke unormalt i befolkningen, det vil si opptil 10 prosent, og vi ser det som viktig at chatbot som kommunikasjonstjeneste ikke hindrer dem i å skaffe seg informasjon og hjelp fra offentlig tjenester. Menneskelig kontakt er fortsatt et alternativ til chatbot i etatene, som vi mener per nå er nødvendig å opprettholde. Fra vår analyse kommer det frem at menneskelig kontakt i ulike scenarier er viktig for deltakerne. Mari blant flere opplevde chatbotsvar som vanskelig å forstå, og hun forteller oss at hun hadde lyst til å gi opp «.. ville ringt et menneske nå, dette blir for komplisert». Simonsen, Steinstrøm, Verne og Bratteteig (2020, s. 3) viser til flere studier der veiledere på telefonsentre i offentlig tjenester arbeider både kunnskapsrikt og emosjonelt når de svarer innringere. Det emosjonelle man kan få hos mennesker er noe deltakere opplever som manglende i chatbotene. Samtidig med det emosjonelle har også veiledere en helt annen kapasitet til å forstå deres brukere enn chatbotene gjør i dag. «En person hadde forstått dette med en gang», forteller Einar i det øyeblikket en av chatbotene ikke klarte å vurdere skrivefeil han anså som liten. Selv skjønnte vi hva han spurte om, og en veileder ville trolig kunne svart mer kunnskapsrikt enn «Beklager, men jeg forstod ikke helt hva du spurte om».

Den menneskelige kontaktformen reduseres på grunn av digitalisering. I kronikken «Digitaliseringen og NAV - hva er brukernes erfaringer?» blir det konkludert med at menneskelig kontakt må ivaretas selv om NAV digitaliseres. De mener at det er et behov for mer kunnskap om hvordan teknologien oppfattes for brukerne. Data fra deres brukerundersøkelser viser at alder, utdanning og stønadssituasjon spiller en rolle ved bruk av ny teknologi. For eksempel var det et større behov for menneskelig kontakt ved psykisk sykdom for å bli veiledet på nettstedet eller ved utfylling av skjema (Hansen, Syltevik & Lundberg, 2017). Brandtzæg (Mathisen, 2018) mener det er et paradoks at man designer primært for de som allerede kan og har mulighet til å benytte seg av disse tjenestene, mens man glemmer de som står utenfor. Brandtzæg sier at dette problemet er spesielt stort i Norge da vi er ledende i Europa på digitalisering av offentlige tjenester. Han poengterer at NAV må tilpasse seg brukerne de i utgangspunktet skal gi tjenester til, de som faller utenfor. Vi har

ikke et sammenligningsgrunnlag, men kan det også tenkes at dyslektikere ikke har blitt tilstrekkelig inkludert i utviklingsfasen av disse chatbotene?

Til tross for dette var ikke chatbotteknologien uinteressant for deltakerne. For eksempel synes noen at Skatteetatens chatbot var god på å gi enkle svar og NAVs løsning fikk også skryt for at den virket ryddig. En maskin trenger ikke søvn som mennesker, og kan veilede flere personer parallelt så lenge regnekraft og lagringskapasitet tillater det. Blant deltakerne ble det nevnt at de likte at teknologien var tilgjengelig når som helst på døgnet, at man slapp kø og i noen tilfeller anså de det også som positivt at de ikke måtte forholde seg til et menneske. Det sistnevnte kan omhandle traumatiske opplevelser eller stigma i forbindelse med å lese og skrive som forskningen viser at dyslektikere ofte opplever (McNulty, 2003; Mortimore & Crozier, 2006). Vi mener at chatboter som er kunstig intelligente har et potensiale til å fjerne denne kommunikasjonsbarrieren for dyslektikere, og kan derfor avlaste noe av presset knyttet til skriving og lesing. En chatbot er ikke et menneske, den kan bare beregne, men man skal ikke se bort ifra at de kan bli den foretrukne kommunikasjonsformen for dyslektikere.

## 8. Konklusjon

I denne studien har vi tatt utgangspunkt i dyslektikere sine opplevelser med bruk av to chatboter innen forvaltningen. Med en involvering av et utvalg dyslektikere i en brukersentrert design tilnærming har vi identifisert tilgjengelighetsproblemer med chatbotene slik de er utformet i dag. Vi vil videre svare på forskningsspørsmålene vi har stilt og som har ledet denne forskningen.

### **Hvordan oppleves chatbotene til Skatteetaten og NAV for dyslektikere?**

Dyslektikere sine opplevelser ved bruk av chatbotene til NAV og Skatteetaten danner grunnlag for å mene at chatbotene ikke er brukervennlig, og heller ikke tilstrekkelig møter deres behov og forventninger. Som videre fører til at menneskelig kontakt forblir deres foretrukne kommunikasjonsform. Teknologien blir likevel sett på som nyttig. Den videre oppsummering tar sikte på å svare på hvorfor vi mener det.

Analysen indikerer at det forekommer tilgjengelighetsproblemer ved navigering til begge chatbotene. Det første problemet er at chatbotene oppleves som vanskelig å finne frem til på nettsidene til Skatteetaten og NAV. To funn viser også at deltakere ikke finner frem og gir opp. Dersom man ikke finner frem til chatbotene har man heller ikke tilgang til å bruke kommunikasjonskanalen, som betyr at det foreligger et teknisk tilgjengelighetsproblem. Generelt blir det foreslått av deltakerne at chatbotene bør være mer tilgjengelig. Brukergrensesnittet blir også betraktet som for liten i størrelse, og blant en av chatbotene blir det trukket assosiasjoner til antropomorfisme. Det er både positive og skeptiske tolkninger til menneskelig egenskaper blant deltakerne. Begge chatbotene blir likevel forstått som et dataprogram.

Analysen viser at brukervennlighetsproblemer som har blitt funnet har opphav i lesing og skriving. Funnene medfører en opplevelse som ikke er vellykket og indikerer at chatbotene ikke er intuitive, effektive eller enkle å bruke for deltakerne med dysleksi. Våre funn viser at systemsvar fra chatbotene oppleves som mye eller vanskelig å lese blant deltakerne. Det er på bakgrunn av chatbotenes evne til å predikere intensjoner at mye tekst blir produsert. Mye tekst sliter blant annet på konsentrasjonen, og all tekst produsert blir derfor heller ikke lest. Flere deltakere opplever det som unødvendig at chatbotene produserer mye tekst fordi svarene ikke er

i samsvar med deres forventninger. Mye tekst tilsvarer ofte at flere sekvenser av meldinger som blir produsert etter hverandre. Det vil si at teksten blir delt inn i flere meldinger enn en stor melding. Slike forekomster medfører at deltakere avbrytes når de leser tekst, og dette oppleves som irriterende og frustrerende. Det er en utfordring å avkode og forstå tekst for dyslektikere. Det blir ytterligere utfordrende når vanskelige og like begreper blir brukt i systemsvar av chatbotene. Når svaralternativer er like, eller vanskelig begreper blir brukt i alternativene, medfører dette opplevelser som usikkerhet og frustrasjon når de skal velge mellom alternativene. I verste tilfelle gir deltakere opp.

Forekomster av feilstavelser hos deltakere fører i visse tilfeller til at chatbotene ikke predikerer brukerintensjonen til deltakere riktig. Konsekvensen er at chatbotene ikke svarer, gir feil svar eller foreslår en videre handling som er feil. Slike tilfeller samsvarer ikke med forventningene til deltakere. De blir frustrert når de ikke blir forstått. Det blir blant annet forventet at mindre skrivefeil blir forstått, og det uttrykkes også tvil om det er chatbotene eller de som er problemet.

Det er flere tilfeller i bruk av chatbotene som viser eksempler der menneskelig kontakt er en nødvendighet for deltakerne. Generelt er kontakt med et menneske deres foretrukne kontaktform. Våre funn indikerer opplevelser som er negativ ladet i bruk og interaksjon med chatbotene. Likevel er det også funn som oppleves som positivt, blant annet at bruk av svaralternativer reduserer skriving, og noen svar oppleves som presise og ryddige på deres spørsmål. En chatbot blir sett på som nyttig blant deltakere fordi man slipper å vente i kø, og i tillegg til at den alltid er tilgjengelig til bruk. I noen tilfeller er det også positivt at de ikke må forholde seg til et menneske som kan ha bakgrunn i traumatiske opplevelser eller stigma i forbindelse med å lese og skrive. Dette mener vi svarer på hvordan personer med dysleksi i dette studiet opplevde NAVs og Skatteetatens chatbot.

Designprosessen etter brukertesting tok for seg å se på alternativer til forberedelser i en chatbot, med en hensikt om å imøtekomme deltakernes behov fra brukertesten. Vi har argumentert for at de fleste forslagene kan imøtekomme og øke tilgjengeligheten for dyslektikere, men også ikke-dyslektikere. Med det neste forskningsspørsmålet tar vi utgangspunkt i det vi har funnet ut om deres opplevelser ved bruk av chatbotene.

**Hvordan designe en chatbot for å imøtekomme dyslektikere sine behov?**

Det første problemet vi adresserer er å gjøre det mer tilgjengelig å komme frem til chatboten i NAV med en klarere navigasjon enn i dag. Da vil man muligens ikke oppleve et teknisk tilgjengelighetsproblem. Når man designer en chatbot som skal oppleves som brukervennlig for dyslektikere kan man vurdere en rekke implementasjoner av funksjonalitet: tekst-til-tale, større grensesnitt, forklaring av vanskelig ord og begreper, forhindre avbrytelser i lesingen. Disse har vi laget prototyper til, og dyslektikere som har deltatt i evalueringsprosessen er flere fordeler med prototypene vi har kommet frem til. Andre anbefalinger som også kan imøtekomme dyslektikers problemer er en annen toleranse i stavekontrollen, mer bruk av svaralternativer, redusere unødvendig tekst og la brukeren korrigere stavefeil som blir markert.

## 8.1 Videre forskning

Med dette forskningsarbeidet til grunn foreslår vi flere mulige retninger man videre kan studere. Det vil være særlig interessant å se nærmere på designforslagene i et større tilgjengelighetsperspektiv. Et utgangspunkt er derfor å forhøre seg med eksperter innen universell utforming av IKT-løsninger. Dersom de kjenner til problemområder hos et bredt mangfold kan dette være nyttig innsikt angående presenterte forslag. Deretter er det videre mulig å evaluere hvordan forslagene oppleves for dyslektikere, og ikke-dyslektikere. Dette kan for eksempel innebære å lage høyoppløselige interaktive prototyper som man tester på et utvalg. Det er en mangel på forskning av chatbot og tilgjengelighet. Dette studiet har sett på et utvalg av unge voksne med dysleksi. Utover dette kunne det vært interessant å se på andre funksjonsnedsettelse slik som motoriske og sensoriske nedsettelse. Det er spesielt spennende å se om forslagene vi presenterer gagnar noen av deres behov.

En kontrollert brukertest av chatboter har vært et viktig første steg for å etablere en innsikt om hvordan teknologien oppleves for personer med dysleksi. Imidlertid simulerer dette ikke de naturlige samtalene som man kan finne i chatbot-logger. Tekst er noe dyslektikere ønsker å unngå, og et mer skriftlig samfunn stiller økte krav til deres lese- og skriveferdigheter. For at dyslektikere skal ha en reel sjanse for å henge med i et mer digitalt skriftlig samfunn hadde det vært interessant og viktig å se hvordan samtalen hadde utbredt seg, og hvilke opplevelser de har når denne teknologien blir brukt naturlig i deres hverdag.



Man kan i tillegg tenke seg at det er i større grad mulig å utforske formatet chatboter er basert på. Vi mener chatbotene representerer konvensjoner innenfor chat, der tekst output og input er normen. I vår studie vi i stor grad forholdt oss til dette formatet da vi mener det var et naturlig første steg i et redesign, og spesielt med tanke på å ikke forringe det kjente og konvensjonelle med en chat. Dyslektikerne i vår studie var i utgangspunktet ikke glad i å lese og skrive tekst, og et design som inneholder andre formater kan utforskes videre. Eksempelvis er det interessant å se på designløsninger som inneholder visuelle formater som video og bilde fremfor tekst, hvordan det kan avlaste kognitivt, og eventuelt forbedre opplevelsen hos dyslektikere og andre grupper.

At det mangler noe emosjonelt blir brukt som et argument for å heller snakke med et menneske blant deltakere i dette studiet. Som tidligere nevnt i kapittel 2.4.9 er Microsofts Xiaoice chatbot ment for nettprat, men den kan også bestemme humøret til brukeren og tilpasse sin systemrespons deretter. Kan dette være en teknologi som gjør chatbotene til et mer foretrukket alternativ enn et menneske?

## 9. Litteratur

- Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. Bainbridge, W. Encyclopedia of Human-Computer Interaction.: Sage Publications, 37(4), 445-456.
- Ahissar, M. (2007). Dyslexia and the anchoring-deficit hypothesis. *Trends in Cognitive Sciences*, 11(11), 458-465
- Andersen K. G., Enger A., Knutsen, H. (2019). Skatteetatens Analysenytt. Hentet 20. mai 2020, fra <https://www.skatteetaten.no/om-skatteetaten/analyse-og-rapporter/analysenytt/>
- Andersen, G. (2019). korpus - språkvitenskap. I Store norske leksikon. Hentet 25. april 2020 fra [https://snl.no/korpus\\_-\\_språkvitenskap](https://snl.no/korpus_-_språkvitenskap)
- Arbeids- og sosialdepartementet. (2016). Meld. St. 33 (2015–2016) [Stortingsmelding]. Hentet 20. mai 2020 fra <https://www.regjeringen.no/no/dokumenter/meld.-st.-33-20152016/id2501017/>
- Arbeids- og sosialdepartementet. (u.d.). *Arbeids- og velferdsetaten (NAV)*. Hentet fra: [https://www.regjeringen.no/no/dep/asd/om-arbeids--og-sosialdepartementet/etatstyring/underliggende-etater/arbeids\\_og\\_velferdsetaten/id1511/](https://www.regjeringen.no/no/dep/asd/om-arbeids--og-sosialdepartementet/etatstyring/underliggende-etater/arbeids_og_velferdsetaten/id1511/)
- Begnum, M. E. N. (2016). Methodology for universal design of ITs; epistemologies among norwegian experts. In International Conference on Computers Helping People with Special Needs (s. 121-128). Springer, Cham.
- Berget, G. (2016). *Search and find? An accessibility study of dyslexia and information retrieval* (Doktoravhandling). Det matematisk-naturvitenskapelige fakultet, Universitetet i Oslo, Oslo
- Berget, G., & Sandnes, F. E. (2015). Searching databases without query-building aids: implications for dyslexic users.
- Berget, G., & Sandnes, F. E. (2016). Do autocomplete functions reduce the impact of dyslexia on information-searching behavior? The case of Google. *Journal of the Association for Information Science and Technology*, 67(10), 2320-2328.
- Berget, G., Herstad, J., & Sandnes, F. E. (2016). Search, read and write: An inquiry into web accessibility for people with dyslexia. In *Universal Design 2016: Learning from the Past, Designing for the Future: Proceedings of the 3rd International Conference on*

- Universal Design (UD 2016), York, United Kingdom, August 21–24, 2016 (Vol. 229, p. 450). IOS Press
- Bogdan, R., & Biklen, S. K. (1997). *Qualitative research for education*. Boston, MA: Allyn & Bacon.
- Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: changing user needs and motivations. *Interactions*, 25(5), 38-43.
- Bratteteig, T., & Verne, G. (2018). Does AI make PD obsolete? exploring challenges from artificial intelligence to participatory design. In *Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial-Volume 2* (s. 1-5).
- British Dyslexia Association. (2018). *Dyslexia Style Guide*. Hentet 20. mai, fra [https://cdn.bdadyslexia.org.uk/documents/Advice/style-guide/Dyslexia\\_Style\\_Guide\\_2018-final-1.pdf?mtime=20190409173949](https://cdn.bdadyslexia.org.uk/documents/Advice/style-guide/Dyslexia_Style_Guide_2018-final-1.pdf?mtime=20190409173949)
- Buxton, B. (2007). *Sketching user experiences: Getting the design right and the right design*. Amsterdam: Elsevier.
- Caldwell, B., Cooper, M., Reid, L. G., & Vanderheiden, G. (2008). *Web content accessibility guidelines (WCAG) 2.0*. WWW Consortium (W3C).
- Carawan, L. W., Nalavany, B. A., & Jenkins, C. (2014). Emotional experience with dyslexia and self-esteem: the protective role of perceived family support in late adulthood, *Aging & Mental Health*. *Aging & Mental Health*, 284-294.
- Dalrymple, J. (2009). App Store tally: 20,000 apps in seven months. *Macworld*. Hentet 10. februar, fra <http://www.macworld.com/article/1138753/appstore.html>
- Dent, S. (2016). Facebook opens up Messenger to ad bots. *Engadget*. Hentet 10. februar, fra: <https://www.engadget.com/2016/08/17/facebook-opens-up-messenger-to-ad-bots/>
- Digitaliseringsdirektoratet. (2018). *Statusmåling 2018 - Oppsummering*. Hentet 20. mai 2020, fra <https://uu.difi.no/tilsyn/statistikk-og-undersokelser/statusmaling-2018-oppsummering>
- Disability Rights Commission. (2004). *The Web: Access and inclusion for disabled people: A formal investigation conducted by the Disability Rights Commission*. London: TSO.
- Diskriminerings- og tilgjengelighetsloven. (2008). *Lov om forbud mot diskriminering på grunn av nedsatt funksjonsevne (LOV-2008-06-20-42)*. Hentet Mars 17, 2020, fra <https://lovdata.no/dokument/LTI/lov/2008-06-20-42>

- Dixon, M. (2007). Comparative study of disabled vs. non-disabled evaluators in user-testing: Dyslexia and first year students learning computer programming. In C. Stephanidis (Ed.), *Universal Access in Human Computer Interaction: Coping With Diversity*, 4th International Conference on Universal Access in Human-Computer Interaction, UAHCI 2007, Held as Part of HCI International 2007, Beijing, China, Juli 22-27, 2007, Proceedings, Part I (s. 647-656). Berlin Heidelberg: Springer.
- Dumas, J. S., & Redish, J. C. (1999). What is usability testing. I *A Practical Guide to Usability Testing* (s. 22). Intellect Books.
- Dvergsdal, H. (2019). Digitalisering i Store norske leksikon på snl.no. Hentet 20. mai 2020, fra <https://snl.no/digitalisering>
- Dysleksi Norge. (2017). Faglige retningslinjer for kartlegging, utredning og oppfølging av elever med dysleksi. Dysleksi Norge.
- Dysleksi Norge. (u.d.). Fagstoff. Hentet 20. mai 2020, fra <https://dysleksinorge.no/fagstoff/>
- Falkum, I. L., & Kjøll, G. (2014). The tricky part is to design for when things go wrong.i. *Norsk Lingvistisk Tidsskrift*, 32(2).
- Fialho, P., Coheur, L., Curto, S., Cl'audio, P., Angela Costa, Abad, A., Meinedo, H., & Trancoso, I. (2013). Meet Edgar, a tutoring agent at Monserrate. In Proceedings of the 51st Annual Meeting of the ACL: System Demonstrations, pages 61–66.
- Finansdepartementet. (u.d.). Skatteetaten [Eksternvirksomhet]. Hentet fra: [https://www.regjeringen.no/no/dep/fin/org/underliggende\\_etater/skatteetaten/id270405](https://www.regjeringen.no/no/dep/fin/org/underliggende_etater/skatteetaten/id270405)
- Forente Nasjoner. (2006). *FN konvensjon om personers rettigheter med nedsatt funksjonsevne* (United Nations Convention on the Rights of Persons with Disabilities). Hentet 20. mai 2020, fra <https://www.un.org/disabilities/documents/convention/convoptprot-e.pdf>
- Forskrift om universell utforming av IKT-løsninger. (2013). Forskrift om universell utforming av informasjons- og kommunikasjonsteknologiske (IKT)-løsninger (FOR-2013-06-21-732). Hentet fra <https://lovdata.no/dokument/SF/forskrift/2013-06-21-732>
- Fuglerud, K. S., Skotkjerra, E., & Halbach, T. (2015). Virtuell hjelpemiddellab. Hentet fra Mars 13, 2020, from <http://iktforalle.no/virtuell-hjelpemiddellab/veileder.html#forfattere>
- Fuglerud, K. S. (2005). IKT for et inkluderende arbeidsliv: Resultater fra test og evaluering av to IKT-verkøy.
- Fuglerud, K. S. (2011). An evaluation of web-based voting usability and accessibility. *Springer-Verlag*.

- Fuglerud, K. S. (2011). The barriers to and benefits of use of ICT for people with visual impairment. HCI International, Universal Access in Human–Computer Interaction. Design for All and eInclusion, Orlando, Florida, USA. Springer Berlin/Heidelberg. s. 452– 462.
- Fuglerud, K. S. (2014). *Inclusive design of ICT: The challenge of diversity* (Doktoravhandling). Det humanistiske fakultet, Universitetet i Oslo, Oslo
- Fuglerud, K. S., & Røssvoll, T. H. (2012). An evaluation of web-based voting usability and accessibility. *Universal Access in the Information Society*, 11(4), 359-373.
- Fuglerud, K. S., Reinertsen, A., Fritsch, L., & Dale, Ø. (2009). Universell utforming av IKT-baserte løsninger for registrering og autentisering.
- Følstad A., Taylor C. (2020). Conversational Repair in Chatbots for Customer Service: The Effect of Expressing Uncertainty and Suggesting Alternatives. In: Følstad A. et al. (eds) Chatbot Research and Design. CONVERSATIONS 2019. Lecture Notes in Computer Science, vol 11970. Springer, Cham
- Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. interactions, 24(4), 38-42.
- Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018a). What makes users trust a chatbot for customer service? An exploratory interview study. In International Conference on Internet Science (s. 194-208). Springer, Cham.
- Følstad, A., Skjuve, M., & Brandtzæg, P. B. (2018b). Different Chatbots for Different Purposes: Towards a Typology of Chatbots to Understand Interaction Design. In International Conference on Internet Science (s. 145-156). Springer, Cham.
- Germanó, E., Gagliano, A. & Curatolo, P. (2010). Comorbidity of ADHD and dyslexia. *Developmental Neuropsychology*, 35(5), 475-493.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press. Hentet 25. april 2020 <http://www.deeplearningbook.org>
- Gough, P. & Tunmer, W. E. (1986). Decoding, reading and reading disability. *Remedial and Special Education*, 7, 6–10.
- Gould, J. D., & Lewis, C. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM*, 28(3), 300-311.
- Grice, H. P. (1975). Logic and conversation. In *Speech acts* (s. 41-58). Brill.

- Hansen, H. T., Lundberg, K. G., Syltevik, L. J. (2017). Kronikk: Digitaliseringen og NAV - hva er brukernes erfaringer? Hentet 20. mai, 2020, fra <https://velferd.no/meninger/2017/digitaliseringen-og-nav-hva-er-brukernes-erfaringer>
- Harper, S. (2006). Is there design-for-all? Universal Access in the Information Society, 6 (1): 111-113.
- Hatch, J. A. (2002). Doing qualitative research in education settings. Suny Press.
- Holmquist, E., H. (2017). Intelligence on Tap: Artificial Intelligence as a New Design Material. Hentet 25. april 2020 <https://interactions.acm.org/archive/view/july-august-2017/intelligence-on-tap>
- Houde, S., & Hill. C. (1997). What do prototypes prototype?. In Handbook of human-computer interaction (s. 367-381). North-Holland.
- International Organization for Standardization. (2018). Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts (ISO Standard No. 9241-11:2018) Hentet 8. juni 2020, fra <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260. Artificial intelligence, machine learning and deep learning. In 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE) (s. 1-6). IEEE.
- Joyce, A. (2019). *Nielsen Norman Group*. Hentet fra Nielsen Norman Group: <https://www.nngroup.com/articles/formative-vs-summative-evaluations/>
- Jurafsky, D., & Martin, H. J. (2019). Speech and Language Processing (3rd ed. draft) Draft chapters in progress. Hentet 20. mai 2020, fra <https://web.stanford.edu/~jurafsky/slp3/>
- Kahrs, C., & Flølo, K. (2019). Dyslektikere sliter på nett—Tema—Klar Tale. <https://www.klartale.no/tema/dyslektikere-sliter-pa-nett-1.1534777>
- Klinkenberg, J. E. (2017). Lesevansker. Oppsummering av ny forskning. Tidsskrift for Norsk psykologforening, Vol 55, nummer 9, ss. 834 - 843.
- Kommunal- og moderniseringsdepartementet. (2016). Meld. St. 27 (2015–2016) [Stortingsmelding]. Hentet 20. mai 2020, fra <https://www.regjeringen.no/no/dokumenter/meld.-st.-27-20152016/id2483795/>
- Kommunal- og moderniseringsdepartementet. (2020). Nasjonal strategi for kunstig intelligens [Plan]. Hentet 20. mai, fra <https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/>

- Kommunal- og moderniseringsdepartementet. (2014). Digital kommunikasjon hovedregel— Viktig informasjon om endringer i forvaltningsloven og eForvaltningsforskriften [Brev]. Hentet 20. mai, fra <https://www.regjeringen.no/no/dokumenter/digital-kommunikasjon-hovedregel/id751013/>
- Kommunal- og moderniseringsdepartementet. (2019). Én digital offentlig sektor [Plan]. Hentet 20. mai, fra <https://www.regjeringen.no/no/dokumenter/en-digital-offentlig-sektor/id2653874/>
- Laurin, S. (2019). Universell utforming av chatbots. Hentet 20. mai, 2020, fra <https://www.funka.com/no/vare-opdrag/forskning-og-innovasjon/arkiv---forskningsoppdrag/universell-utforming-av-chatbots/>
- Lazar, J. (2007). Introduction to universal usability. *Universal usability: Designing computer interfaces for diverse user populations*, 1-12.
- Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research Methods in Human-Computer Interaction*. Todd Green.
- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in psychology*, 6, 731.
- Lindgaard, G. & Chattratchart, J. (2007). Usability testing: what have we overlooked? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1415–1424.
- Löwgren, J., & Stolterman, E. (2004). *Thoughtful interaction design: a design perspective on information technology*. Cambridge, Mass: MIT Press.
- Mathisen, G. (2018). Mener digitaliseringen av Nav rammer dem med størst behov for hjelp. Hentet 20. mai, 2020, fra <https://www.handikapnytt.no/mener-digitaliseringen-av-nav-rammer-dem-med-storst-behov-for-hjelp/>
- McCarthy, J. (1998). What is artificial intelligence?
- McCarthy, J. E., & Swierenga, S. J. (2010). What we know about dyslexia and web accessibility: a research review. *Universal Access in the Information Society*, 9(2), 147-152.
- McNulty, M. A. (2003). Dyslexia and the life course. *Journal of learning disabilities*, 36(4), 363-381
- Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation*. John Wiley & Sons (261)

- Moore, R. J., & Arar. R. (2018). Conversational UX design: an introduction. In *Studies in Conversational UX Design* (s. 1-16). Springer, Cham.
- Mortimore, T. & Crozier, W. R. (2006). Dyslexia and difficulties with study skills in higher education. *Studies in Higher Education* 31, 235–251
- Myers, M. (1997). Qualitative Research in Information Systems. *MIS Quarterly*. 21. 10.2307/249422.
- NAV i tall og fakta (2018). NAVs årsrapport for 2018. Hentet 20. mai 2020, fra [https://www.nav.no/no/nav-og-samfunn/om-nav/årsrapport/nav-i-tall-og-fakta/\\_/attachment/download/7cb77433-2164-46ea-999d-7b68c4f776c5:42667e090282205f3df3c83285e8b665e68a2c7c/nav-i-tall-og-fakta-2018.pdf](https://www.nav.no/no/nav-og-samfunn/om-nav/årsrapport/nav-i-tall-og-fakta/_/attachment/download/7cb77433-2164-46ea-999d-7b68c4f776c5:42667e090282205f3df3c83285e8b665e68a2c7c/nav-i-tall-og-fakta-2018.pdf)
- NAVs omverdensanalyse. (2019). NAVs omverdensanalyse 2019 Utvikling, trender og konsekvenser fram mot 2030. Hentet 20. mai 2020, fra <https://www.nav.no/no/nav-og-samfunn/kunnskap/analyser-fra-nav/analyser/navs-omverdensanalyse-2019>
- Nielsen, J. (1994). 10 Heuristics for User Interface Design: Article by Jakob Nielsen. Nielsen Norman Group. Hentet 20. mai 2020, fra <https://www.nngroup.com/articles/ten-usability-heuristics/>
- Nielsen, J. (2012). Usability 101: Introduction to Usability. Nielsen Norman Group. Hentet 20. Mai 2020, fra <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>
- Nielsen, J. (1993). Usability testing. I J. Nielsen, *Usability Engineering* (s. 165). Academic Press.
- Nielsen, J. (2000). *Nielsen Norman Group*. Hentet fra Nielsen Norman Group: <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>
- Nielsen, J. (2005). Formal Usability Reports vs. Quick Findings. Nielsen Norman Group. Hentet 21. mai 2020, fra <https://www.nngroup.com/articles/formal-vs-quick-usability-reports/>
- Nordheim, C. B., Følstad, A., & Bjørkli, C. A. (2019). An Initial Model of Trust in Chatbots for Customer Service—Findings from a Questionnaire Study. *Interacting with Computers*, 31(3), 317-335.
- Norman, D. A. (1988). *The design of everyday things*. Basic books, 1988.
- NOU 2005:8. (2005). Universell utforming: Fra ideal til rettsnorm. Hentet 20 mai. 2020 fra <https://www.regjeringen.no/no/dokumenter/nou-2005-8/id390520/>



- NOU 2019: 5 (2019). Ny forvaltningslov — Lov om saksbehandlingen i offentlig forvaltning (forvaltningsloven) <https://www.regjeringen.no/no/dokumenter/nou-2019-5/id2632006/>
- Paddison, C. & Englefield, P. (2003). Applying heuristics to perform a rigorous accessibility inspection in a commercial context. Proceedings of the 2003 conference on universal usability, Vancouver, British Columbia, Canada: ACM. doi: 10.1145/957205.957228.
- Persson, H., Åhman, H., Yngling, A. A., & Gulliksen, J. (2015). Universal design, inclusive design, accessible design, design for all: different concepts—one goal? On the concept of accessibility—historical, methodological and philosophical aspects. *Universal Access in the Information Society*, 14(4), 505-526.
- Power, C., Freire, A. P., Petrie, H. & Swallow, D. (2012). Guidelines are only half of the story: accessibility problems encountered by blind users on the web. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, Texas, USA, s. 433- 442. 2207736: ACM.
- Rello, L., Baeza-Yates, R., Dempere-Marco, L., & Saggion, H. (2013). Frequent Words Improve Readability and Short Words Improve Understandability for People with Dyslexia. 14th International Conference on Human-Computer Interaction . Cape Town.
- Rongved, E. (2015). Hva er egentlig dysleksi? Utdanningsforskning.no. <https://utdanningsforskning.no/artikler/hva-er-egentlig-dysleksi/>
- Rudd, J., Stern, K., & Isensee, S. (1996). Low vs. high-fidelity prototyping debate. *interactions*, 3(1), 76-85.
- Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach.
- Sanders, E. B. N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *Co-design*, 4(1), 5-18.
- Sandnes, F. E. (2011). Universell utforming av IKT-systemer. Oslo: Universitetsforlaget.
- Schartum, D. W., Tranvik, T., & Jansen, A. (2017). Digital forvaltning: en innføring: juridiske, informatiske og organisatoriske aspekter. Fagbokforlaget
- Searle, J. R. (1976). A classification of illocutionary acts. *Language in society*, 5(1), 1-23.
- Serban, I. V., Lowe, R., Henderson, P., Charlin, L., & Pineau, J. (2018). A survey of available corpora for building data-driven dialogue systems: The journal version. *Dialogue & Discourse*, 9(1), 1-49.

- Sharp, H., Preece, J., & Rogers, Y. (2019). Interaction Design-beyond human-computer interaction 5th, 636. isbn: 978-1-119-54725-9.
- Shaywitz, S. E. & Shaywitz, B. A. (2005). Dyslexia (specific reading disability). *Biological Psychiatry*, 57(11), 1301–1309
- Shum, H. Y., He, X. D., & Li, D. (2018). From Eliza to XiaoIce: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, 19(1), 10-26
- Simonsen, K. M., Jacobsen, A., Kling, J., Henstad, J., & Laurin, S. (2020). Universell utforming av chatbots. Hentet 12. mai, 2020: <https://www.funka.com/no/vare-oppdrag/forskning-og-innovasjon/arkiv---forskningsoppdrag/universell-utforming-av-chatbots/chatbottar-for-alle/>
- Simonsen, L. (2019). Når brukerdialogen automatiseres — hva blir vanskelig? (Masteroppgave). Det matematisk-naturvitenskapelige fakultet, Universitetet i Oslo, Oslo
- Simonsen, L., Steinstø, T., Verne, G., Bratteteig, G. (2020). “I’m disabled and married to a foreign single mother”. Public service chatbot’s advice on citizens’ complex lives, accepted to eGOV 2020, Linköping
- Simonsen, T. (2019). Digitalisering i NAV. Hentet 20. mai 2020, fra <https://www.fylkesmannen.no/globalassets/fm-nordland/dokument-fmno/kurs-og-konferanser/kommunekonferansen/2019/digitalisering-i-nav---simonsen---nav.ppsx>
- Smith-Spark, J. H., & Fisk, J. E. (2007). Working memory functioning in developmental dyslexia. ss. 34-36.
- Solem, C. (2017). Språkløyper. Hentet 20. mai, 2020, fra <https://sprakloyper.uis.no/snarveier/sprakloyper-bloggen/hvorfor-er-hjelpemidler-viktig-article113173-17610.html>
- Solheim, I., & Halbach, T. (2018). *Nytte av lettlest informasjon for NAV-brukere*. Norsk Regnesentral.
- Sosial- og helsedepartementet. (1997). St.meld. nr. 34 (1996-97) [Stortingsmelding]. Hentet 20. mai 2020 fra: <https://www.stortinget.no/no/Saker-og-publikasjoner/Stortingsforhandlinger/Lesevisning/?p=1996-97&paid=3&wid=c&psid=DIVL281>
- Sosialdepartementet. (1977). St.meld. nr. 23 (1977-78) [Stortingsmelding]. Hentet 20. mai 2020 fra: <https://www.stortinget.no/no/Saker-og->

[publikasjoner/Stortingsforhandlinger/Lesevisning/?p=1977-78&paid=3&wid=b&psid=DIVL1556](https://publikasjoner/Stortingsforhandlinger/Lesevisning/?p=1977-78&paid=3&wid=b&psid=DIVL1556)

- Spool, J., & Schroeder, W. (2001). Testing Web Sites: Five Users Is Nowhere Near Enough. *Association for Computing Machinery*.
- Språkrådet. (u.d.). Brukbarhet, brukbarheit og usability. Hentet 20. mai 2020, fra <http://www.sprakradet.no/svardatabase/sporsmal-og-svar/brukbarhet-brukbarheit-og-usability/>
- Statistisk sentralbyrå. (2017). Ni av ti surfer på nettet hver dag. Hentet 20. mai 2020, fra <https://www.ssb.no/teknologi-og-innovasjon/artikler-og-publikasjoner/ni-av-ti-surfer-pa-nettet-hver-dag>
- Statistisk sentralbyrå. (2019). 11004: Bruk av offentlige myndigheters nettjenester de siste 12 måneder (prosent), etter kjønn, statistikkvariabel og år. Hentet 20. mai 2020, fra <https://www.ssb.no/statbank/table/11004/tableViewLayout1/>
- Story, M. F., Mueller, J. L. & Mace, R. L. (1998). The universal design file: Designing for people of all ages & abilities. Raleigh: Center for Universal Design.
- Svendsen, L., F., H. (2018). talehandling i Store norske leksikon på snl.no. Hentet 28. april 2020 fra <https://snl.no/talehandling>
- Swanson, H.L., & Hsieh, C.-J. (2009). Reading disabilities in adults: A selective meta-analysis of the literature. *Review of Educational Research*, 79(4), 1362–1390.
- Szymanski, M. H., & Moore, R. J. (2018). Adapting to Customer Initiative: Insights from Human Service Encounters. In *Studies in Conversational UX Design* (s. 19-32). Springer, Cham.
- Thakur, A. (2018). How virtual agents work and why you should care. Hentet 12. juni 2020, fra <https://www.boost.ai/articles/how-chatbots-work-and-why-you-should-care>
- Torres, C., Franklin, W., & Martins, L. (2018). Accessibility in Chatbots: The State of the Art in Favor of Users with Visual Impairment. In *International Conference on Applied Human Factors and Ergonomics* (s. 623-635). Springer, Cham.
- Turing, A. M. (2009). Computing machinery and intelligence. In *Parsing the Turing Test* (s. 23-65). Springer, Dordrecht.
- Tønnessen, F. E. (2015). *Can We Read Letters?* Sense Publishers.
- Vavik, T., & Gheerawo, R. (2009). The challenges in universal design. *Inclusive buildings, products & services: challenges in universal design*.

- Vigo, M., Brown, J. & Conway, V. (2013). Benchmarking web accessibility evaluation tools: Measuring the harm of sole reliance on automated tests. Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility, Rio de Janeiro, Brazil, p. 1–10. 2461124: ACM
- Walsham, G. (2006). Doing interpretive research. *European journal of information systems*, 15(3), 320-330.
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.
- Westin, S. (2017). Nav må møte mennesker. *Aftenposten* Hentet fra <https://www.aftenposten.no/meninger/kronikk/i/0RbdE/nav-maa-moete-mennesker-steinar-westin?>
- Willcutt, E. G., Pennington, B. F., Duncan, L., Smith, S. D., Keenan, J. M., Wadsworth, S., ... & Olson, R. K. (2010). Understanding the complex etiologies of developmental disorders: behavioral and molecular genetic approaches. *Journal of developmental and behavioral pediatrics: JDBP*, 31(7), 533.
- Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (s. 3506-3510).

## 10. Vedlegg

### Vedlegg A – Samtykkeskjema brukertesting

Vil du delta i forskningsprosjektet

«Universell utforming av chatbotløsninger»?

#### **Bakgrunn og formål:**

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet med studien er å se på hvordan loven om universell utforming i IKT-løsningen «chat-robot» fungerer i praksis. Studiet vil fokusere på interaksjon, bruk og design av chat-robot, samt hvordan virksomheter og leverandører arbeider med å kvalitetssikre en universell utformet chat-robot. Dette studiet gjøres i forbindelse med en masteroppgave ved Institutt for Informatikk ved Universitetet i Oslo.

#### **Hvem er ansvarlig for forskningsprosjektet?**

Institutt for Informatikk ved Universitetet i Oslo er ansvarlig for prosjektet.

#### **Hvorfor får du spørsmål om å delta?**

Universell utforming handler om å utforme løsninger som tar hensyn til alle, uansett funksjonsevne. Vi ønsker i denne studien å se på chatbot-løsninger med mennesker som har forskjellige utfordringer i forbindelse med nedsatt funksjonsevne og derfor er du invitert til å delta. Din og andres innsikt vil være av stor verdi for å se om nye teknologiske løsninger tar hensyn til Universell utforming.

#### **Hva innebærer det for deg å delta?**

Hvis du velger å delta i prosjektet, innebærer det at du deltar i to korte intervjuer og et web-eksperiment. Dette innebærer:

- Et intervju blir gjort før Web-eksperimentet der vi spør om erfaringer, forventinger med chat-robotene.
- Et web-eksperiment som vil ta for seg din skjerminteraksjon med chat-robot. Vi er ikke interessert i å teste dine ferdigheter, men heller selve løsningen.
- Et intervju blir gjort etter web-eksperimentet der vi spør om dine opplevelser ved bruk av chat-robotene.

Vi tar lydopptak og notater i intervjuet, samt opptak av lyd og skjerminteraksjon i web-eksperimentet. Dette gjøres for at vi skal kunne analysere det som har skjedd i etterkant og for å sikre oss at vi har forstått utsagn og handlinger riktig.

Det vil ta deg ca. 45 minutter å delta i dette prosjektet.

#### **Det er frivillig å delta**

Det er frivillig å delta i dette prosjektet- Hvis du velger å delta, kan du når som helst trekke samtykke tilbake uten å oppgi noen grunn. Alle opplysninger om deg vil da bli anonymisert. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

### **Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger**

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrevet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- Datamaterialet er adgangsbegrenset.
- I masteroppgaven vil din identitet i denne undersøkelsen forbli anonym.
- Vi publiserer ikke ditt navn eller opplysninger som vurderes til at de kan identifisere deg.

### **Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?**

Prosjektet skal etter planen avsluttes 1 juni. 2020. Signerte samtykkeskjema, notater, opptak av lyd og skjerminteraksjon vil bli beholdt i personvernet til Anton Lilleby og Steffen Marstein, som er de eneste som har tilgang til denne rådataen til prosjekt avsluttes hvor disse vil bli slettet. Etter publikasjon av rapporten 1. juni 2020, vil det ikke lenger være mulig å trekke ditt samtykke.

### **Dine rettigheter**

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke personopplysninger som er registrert om deg,
- å få rettet personopplysninger om deg,
- få slettet personopplysninger om deg,
- få utlevert en kopi av dine personopplysninger (dataportabilitet), og
- å sende klage til personvernombudet eller Datatilsynet om behandlingen av dine personopplysninger.

### **Hva gir oss rett til å behandle personopplysninger om deg?**

Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra Institutt for Informatikk ved Universitetet i Oslo har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

### **Hvor kan jeg finne ut mer?**

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

<i>Prosjektansvarlig:</i> Guri. B. Verne Universitetslektor Institutt for Informatikk Universitetet i Oslo Epost: <a href="mailto:guribv@ifi.uio.no">guribv@ifi.uio.no</a>	NSD – Norsk senter for forskningsdata AS Epost: ( <a href="mailto:personverntjenester@nsd.no">personverntjenester@nsd.no</a> ) Telefon: 55 58 21 17.
<i>Student:</i> Anton Lilleby Informatikk: design, bruk, interaksjon (Master). Institutt for Informatikk Universitetet i Oslo	<i>Student:</i> Steffen Marstein Informatikk: design, bruk, interaksjon (Master). Institutt for Informatikk Universitetet i Oslo

Epost: [antonli@ifi.uio.no](mailto:antonli@ifi.uio.no)

Epost: [steffeam@ifi.uio.no](mailto:steffeam@ifi.uio.no)

Med vennlig hilsen

Prosjektansvarlig

Student

Student

---

### Samtykkeerklæring

*Jeg har mottatt og forstått informasjon om prosjektet «Universell utforming av chatbotløsninger», og har fått anledning til å stille spørsmål. Jeg samtykker til:*

- å delta i et Web-basert eksperiment
- å delta i to korte intervjuer

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet, ca. 1. juni 2020

---

(Signert av prosjektdeltaker, dato)

## Vedlegg B – Samtykkeskjema intervju

Vil du delta i forskningsprosjektet

*«Universell utforming av chatbotløsninger»?*

### Bakgrunn og formål:

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet med studien er å se på hvordan loven om universell utforming i IKT-løsningen «chat-robot» fungerer i praksis. Studiet vil fokusere på interaksjon, bruk og design av chat-robot, samt hvordan virksomheter og leverandører arbeider med å kvalitetssikre en universell utformet chat-robot. Dette studiet gjøres i forbindelse med en masteroppgave ved Institutt for Informatikk ved Universitetet i Oslo.

### Hvem er ansvarlig for forskningsprosjektet?

Institutt for Informatikk ved Universitetet i Oslo er ansvarlig for prosjektet.

### Hvorfor får du spørsmål om å delta?

Universell utforming handler om å utforme løsninger som tar hensyn til alle, uansett funksjonsevne. Din og andres innsikt vil være av stor verdi for å se om nye teknologiske løsninger, i dette tilfelle chat-robots, tar hensyn til Universell utforming.

### **Hva innebærer det for deg å delta?**

Hvis du velger å delta i prosjektet, innebærer det at du deltar i et intervju. Dette innebærer:

Vi tar lydopptak og notater i intervjuet. Dette gjøres for at vi skal kunne analysere det som har skjedd i etterkant og for å sikre oss at vi har forstått utsagn og handlinger riktig.

Det vil ta deg ca. 30 minutter å delta i dette prosjektet.

### **Det er frivillig å delta**

Det er frivillig å delta i dette prosjektet- Hvis du velger å delta, kan du når som helst trekke samtykke tilbake uten å oppgi noen grunn. Alle opplysninger om deg vil da bli anonymisert. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

### **Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger**

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrevet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- Datamaterialet er adgangsbegrenset.
- Du vil kunne gjenkjennes i publikasjonen, ved at navn, arbeidstitel og arbeidssted kan brukes til å sitere deg i utdrag fra transkribert lydopptak.

### **Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?**

Prosjektet skal etter planen avsluttes 1 juni. 2020. Signerte samtykkeskjema, notater, opptak av lyd og skjerminteraksjon vil bli beholdt i personvernet til Anton Lilleby og Steffen Marstein, som er de eneste som har tilgang til denne rådataen til prosjekt avsluttes hvor disse vil bli slettet. Etter publikasjon av rapporten 1. juni 2020, vil det ikke lenger være mulig å trekke ditt samtykke.

### **Dine rettigheter**

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke personopplysninger som er registrert om deg,
- å få rettet personopplysninger om deg,
- få slettet personopplysninger om deg,
- få utlevert en kopi av dine personopplysninger (dataportabilitet), og
- å sende klage til personvernombudet eller Datatilsynet om behandlingen av dine personopplysninger.

### **Hva gir oss rett til å behandle personopplysninger om deg?**

Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra Institutt for Informatikk ved Universitetet i Oslo har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

### **Hvor kan jeg finne ut mer?**

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

<i>Prosjektansvarlig:</i> Guri. B. Verne Universitetslektor	NSD – Norsk senter for forskningsdata AS Epost: ( <a href="mailto:personverntjenester@nsd.no">personverntjenester@nsd.no</a> ) Telefon: 55 58 21 17.
---	--



Institutt for Informatikk Universitetet i Oslo Epost: <a href="mailto:guribv@ifi.uio.no">guribv@ifi.uio.no</a>	
<i>Student:</i> Anton Lilleby Informatikk: design, bruk, interaksjon (Master). Institutt for Informatikk Universitetet i Oslo Epost: <a href="mailto:antonli@ifi.uio.no">antonli@ifi.uio.no</a>	<i>Student:</i> Steffen Marstein Informatikk: design, bruk, interaksjon (Master). Institutt for Informatikk Universitetet i Oslo Epost: <a href="mailto:steffeam@ifi.uio.no">steffeam@ifi.uio.no</a>

Med vennlig hilsen

Prosjektansvarlig

*Student*

*Student*

---

### Samtykkeerklæring

*Jeg har mottatt og forstått informasjon om prosjektet «Universell utforming av chatbotløsninger», og har fått anledning til å stille spørsmål. Jeg samtykker til:*

- å delta i intervju
- at opplysninger om meg publiseres slik at jeg kan gjenkjennes med navn, arbeidssted og arbeidstittel

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet, ca. 1. juni 2020

---

(Signert av prosjektdeltaker, dato)

## Vedlegg C – Automatisk evaluering Skatteetaten

The screenshot shows the Siteimprove accessibility audit interface for Skatteetaten. On the left, a list of violations is categorized into Text Alternatives, Adaptable, Distinguishable, Enough Time, Navigable, and Compatible. On the right, a chatbot window titled 'Skatteetaten chat-robot' is displayed, showing a conversation where the chatbot asks for consent to process personal information and then asks for a name.

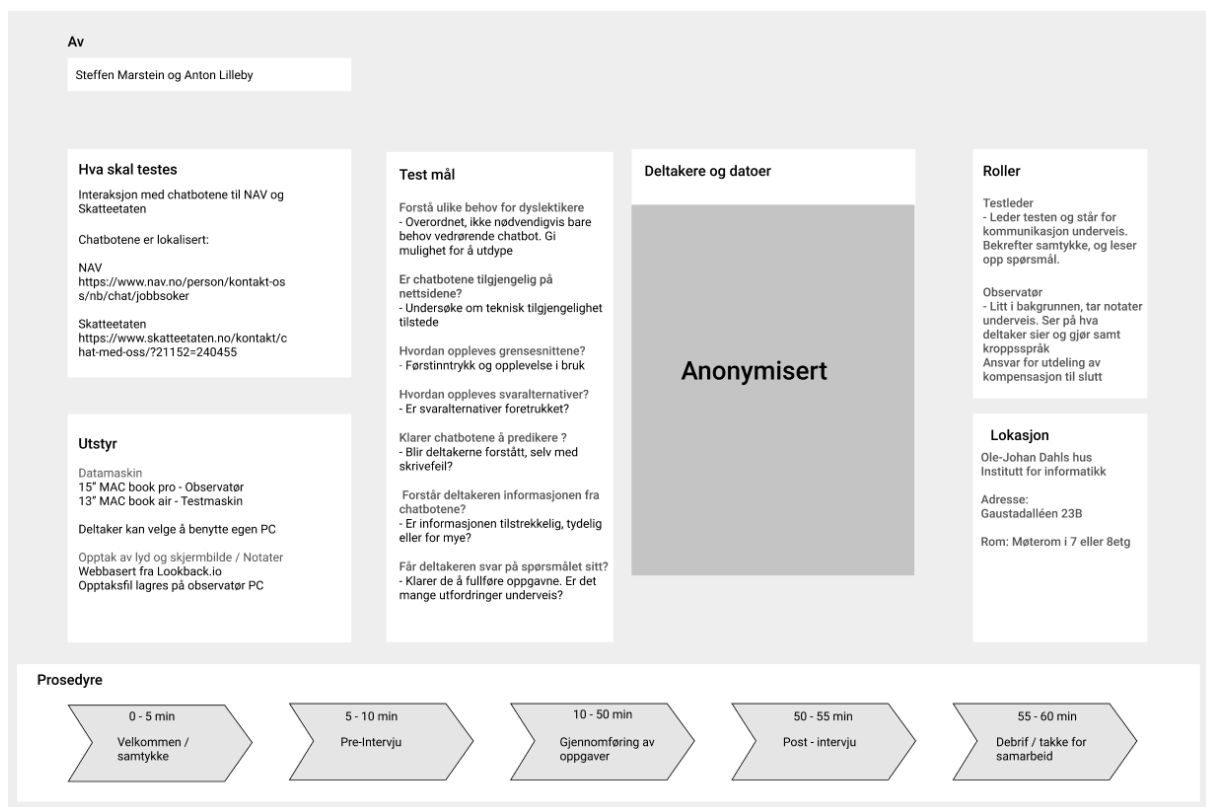
Category	Rule	Occurrences
Text Alternatives	Non-text Content (1.1.1)	2
	Info and Relationships (1.3.1)	1
Distinguishable	Contrast (Minimum) (1.4.3)	1
	Contrast (Enhanced) (1.4.6)	6
Enough Time	Timing Adjustable (2.2.1)	1
Navigable	Bypass Blocks (2.4.1)	1
	Page Titled (2.4.2)	1
	Focus Order (2.4.3)	2
	Focus Visible (2.4.7)	5
	Section Headings (2.4.10)	2
Compatible	Name, Role, Value (4.1.2)	1

## Vedlegg D – Automatisk evaluering NAV

The screenshot shows the Siteimprove accessibility audit interface for NAV. On the left, a list of violations is categorized into Text Alternatives, Adaptable, Distinguishable, Enough Time, Navigable, and Compatible. On the right, a chatbot window titled 'Chatbot Frida' is displayed, showing a conversation where the chatbot introduces itself and asks for a question.

Category	Rule	Occurrences
Text Alternatives	Non-text Content (1.1.1)	28
	Info and Relationships (1.3.1)	6
Distinguishable	Contrast (Enhanced) (1.4.6)	11
Enough Time	Timing Adjustable (2.2.1)	1
Navigable	Bypass Blocks (2.4.1)	1
	Focus Order (2.4.3)	35
	Section Headings (2.4.10)	1
Compatible	Name, Role, Value (4.1.2)	1

## Vedlegg E – Testplan



## Vedlegg F – Funka

### Intervjuguide - Funka

#### Om oss

Hva vi har gjort?

Hva ønsker vi å gjøre.

Fortelle om resultater fra brukertesten.

#### Innledende spørsmål om prosjektet

Kan du fortelle oss kort hvem du er, og hvilken rolle du har i FUNKA?

Hva var årsaken til at dere ville se nærmere på chatbots?

Var det tilfeldig at dere i hovedsak undersøkte chatbots i private bedrifter (Et unntak er Bergen Kommune)

Dere hadde et samarbeid med skatteetaten, på hvilken måte bisto de i prosjektet?

#### Underleverandør

Dere nevner at flere typer feil gikk igjen over flere løsninger og at dette har sammenheng med at grunnmotoren er lik. Kan du komme med flere eksempler på slike typer feil som ble oppdaget utover det som ble nevnt i rapporten?

Har dere testet andre leverandører enn Boost.ai?

Forskjeller?

### **Tilgjengelighet**

Dere nevner blant annet om problematisk plasseringer av chatbot på nettsidene

- Var det noen av nettsidene som hadde «god» plassering?
- Evt eksempler på hvordan man kan gjøre plasseringen mer tilgjengelig for flere

### **Teknisk-test**

- Dere nevner at dere har 37 egne «Funka krav» som er blitt brukt under testen. Kan du fortelle litt om disse kravene og da spesielt krav som innbefatter kognisjon / dyslektikere?
- Hvilke WCAG krav er benyttet og da spesielt krav som innbefatter kognisjon / dyslektikere?
- Hvilke feil av de som ble funnet anser dere som alvorlige?
- Rent praktisk, hvordan gjennomførte dere testen? Benytter dere et rammeverk?
- Hvilke verktøy og hjelpemidler tok dere i bruk, eller anbefaler å bruke ved teknisk test? Og hva må eventuelt sjekkes manuelt?

Hva innebærer ordet ekspert i en teknisk-test?

Hva legger dere i ordet brukervennlig?

### **Brukertesting**

- Kan du fortelle oss litt om hvordan dere gjennomførte brukertestene og analyserte resultatene? (Om lite tid igjen, bare ta for oss testen med dyslektiker)
- Brukte dyslektiker hjelpemidler?
- Var det noen utfordringer de med kognitive funksjonsnedsettelse hadde til felles?
- Hvilke endringer ser du/dere på som nødvendig for at løsningen skal ha god tilgjengelighet?

### **Annet**

- Dere nevner at flere hadde problemer med at et flertall meldinger som blir sendt etter hverandre var et problem, men at dette er vanskelig å endre pga. chatbotens natur. Har du noen tanker rundt hva som kan hjelpe disse brukerne?
- DNBs chatbot scoret dårligst på de tekniske testene, men var best likt av 321 brukerne. Tanker rundt hvorfor?

## Vedlegg G – Dysleksi Norge

### Intervjuguide – Dysleksi Norge

#### Introduksjon

- Hva studerer vi?
- Hva skriver vi masteroppgave om?
- Hvorfor skriver vi om det?
- Det er årsaken til at du er her i dag.

#### Fortell litt om deg selv og din rolle i dysleksi ungdom

- Aldersgruppe?
- Tilrettelegging?
- Digitale-løsninger?
- Medlemstilbud?
- Handlingsplan?

#### Hvordan er det å bruke mobiltelefon og datamaskin for en med dysleksi?

- Hvordan er det å lese på nettsider i dag?
  - Hva tenker du når du ser mye tekst?
  - Hva tenker du om naturlig språk?
- Hvordan er det å skrive tekst på nettsider?
  - Hvordan er det å fylle ut nettskjemaer?
- Verktøy for dyslektikere

#### Hvordan er det å bruke chatløsninger som SMS og Messenger for en med dysleksi?

- De med lese- og skrivevansker oppsøker kundeservice på telefon fremfor e-post og chat?

Chat-robotter begynner nå å bli brukt i kunderservice. Du snakker altså med en robot og ikke et menneske. Hvilke fordeler og ulemper tenker du dette har for de med lese- og skrivevansker?

- Offentlige tjenester
- Hvordan tror du de med skrivevansker blir forstått av chatbot?
- Hvordan tror du med lesevansker forstår tekst med chatbot?
- Hvordan tror du overgangen fra menneskelig kontakt til chatrobott blir for de med lese- og skrivevansker?
- Hva tenker du om scenarioet dersom ikke menneskelig kontakt var tilgjengelig?
- Hva tenker du om at kundeservice kontakt blir først på chatrobott og deretter med menneske?

Vi ønsker å teste chatbot til for eksempel NAV og Skatteetaten med personer med dysleksi.

## Vedlegg H – Intervju NAV

### Intervjuguide NAV - UU-Avdeling

- Hva studerer vi?
- Hva skriver vi masteroppgave om?
- Hvorfor skriver vi om det?

#### Tilgjengelighet

Hvor mange brukere får svar uten å få kontakt med et menneske?

- Har dere noe statistikk tilgjengelig?
- Hvordan tar "IKT-løsning" hensyn til lese og skrivefeil

Hva tenker du om påstanden:

- En brukervennlig chatbot for alle er en universell utformet IKT-løsning?

### **User interface**

Slik vi forstår det har chatboten deres en egen front-end til sammenligning med skatteetaten.

- Hvilke fordeler og ulemper kommer dette med?
- Hvorfor har dere gjort dette?
- Hva er grunnen til personifisering dere har valgt?
- Hvorfor har dere laget den i eget vindu, og ikke mindre slik som skatteetaten?

### **Samarbeid på tvers av etater**

- Samarbeider dere med andre etater ved utvikling av IKT-løsning?
- Burde dere dele erfaringer med f.eks. skatteetaten i IKT-løsninger?

### **Leverandør og krav**

Vet du om noen spesielle krav til chatbots?

- Hva tenker dere resultatet ville vært dersom DIFI testet chatboten deres på kravene?
- DIFI sier til oss at de ikke har testet chatboten mot kravene.

Har dere vurdert andre selskaper, med tanke på hva de kan levere?

- Hvordan arbeider dere med leverandøren for å kvalitetssikre brukbarhet i chatboten?

### **Digitalisering**

Et mål for mange er at chatboter tar over kommunikasjonskanaler med mennesker

- Hva tenker dere om at IKT-løsninger reduserer menneskelig kontakt
- Har dere noen tanker om hvordan sosialhjelp og foreldrepenger chatboten er utviklet i forhold til brukernes situasjon?
- Har dere brukere?

### **Praksis med UU**

Dere skal favne mange ulike mennesker i ulike situasjoner.

- Hvordan sikrer dere en tilgjengelig IKT-løsning?
- Hvordan tester dere IKT-løsning deres?
- Hva med Frida?

## **Vedlegg I – Skatteetaten**

### **Intervjuguide - Skatteetaten**

#### **Formål**

Målsettingen med dette intervjuet er å få en bredere forståelse av Web-baserte chatbotløsninger, investigere utfordringer knyttet til utforming og bruk av disse, samt dokumentere hvilke forbedringstiltak som må iverksettes i forhold til Universell Utforming. Deres forståelse av tilgjengelig informasjon på dette vil være nyttig for oss.

#### **Intervjutype**

Dette intervjuet er semi-strukturert og inneholder tre deler:

- Oppvarmingsspørsmål for å få samtalen i gang
- Hoveddel som inneholder åpne spørsmål rundt tematikk
- Avkjølingsspørsmål som gir mulighet for intervjudeltaker å tilføye noe om ønskelig

Spørsmål deltaker vil hjelpe med å svare:

- Dere skal favne mange ulike mennesker, hvordan tilrettelegger dere for å lage løsninger som fungerer for alle?
  - o Hvem er det som er typiske brukere av chatbotten deres?
  - o Stiller dere design krav til leverandør om hvordan grensesnittet skal være?
  - o Vet dere hvordan personer med nedsatt funksjonsevne bruker chatbotten deres?
- Hvordan arbeider dere med universell utforming av chatbotten?
- Hvordan skiller deres chatbot grensesnitt seg ut i forhold til andre eksisterende løsninger?
- Hvordan måler dere brukeropplevelser i en chatbot?
  - o Blir tommel opp og ned benyttet?
- Fra sluttrapport mener veileder at det er viktig med trygghet og språk og tone i chatbotten. Hvordan måler dere og implementerer dere trygghet, samt språk og tone i chatbotten?
  - o F.eks. forståelse, ta bruker på alvor, selvsikkerhet
  - o Speile brukeren i uttrykk og stil og ikke bruke svake ord
- Forklarer chatbotten i dag hvorfor ting er som der? (spørsmål ifh. sluttrapport)
- Chatbotten deres har en personlighet, hvorfor har dere valget akkurat denne?
  - o Forstår deres brukere hvilke begrensninger chatbotten har?
- Har stemmebasert chatbot vært diskutert brukt i deres chatbot?
- Hvilke typiske tilbakemeldinger av brukere av chatbotten får dere?
- Chatbot De spørsmålene som blir hyppig stilt
- Hva er neste skritt i utviklingen av chatbot for dere?
  - o Integrasjon med deres systemer? Innlogget chatbot en mulighet?
- Hvilke design elementer tilbyr chatbotten deres, utenom normal tekst?
  - o Lenker, fet skrift og emoji f.eks.