

Mathias Hagset

Hvordan blir kunstig intelligens "intelligent"?

En studie av utviklingsprosessen bak en AI-løsning til diagnostikk i helsetjenesten

Masteroppgave i Studier av kunnskap, teknologi og samfunn (STS)

Veileder: Kristine Ask

Medveileder: Stig Kvaal

Juni 2024

Mathias Hagset

Hvordan blir kunstig intelligens "intelligent"?

En studie av utviklingsprosessen bak en AI-løsning til
diagnostikk i helsetjenesten

Masteroppgave i Studier av kunnskap, teknologi og samfunn (STS)
Veileder: Kristine Ask
Medveileder: Stig Kvaal
Juni 2024

Norges teknisk-naturvitenskapelige universitet
Det humanistiske fakultet
Institutt for tverrfaglige kulturstudier



NTNU

Kunnskap for en bedre verden

Studier av kunnskap, teknologi og samfunn (STS)

Læringsutbytte

En student som har fullført programmet, forventes å ha oppnådd følgende læringsutbytte, definert i kunnskap, ferdigheter og generell kompetanse:

Kunnskap

Kandidaten har

- avansert kunnskap om hvordan vitenskap og teknologi utvikles, brukes og implementeres i samfunnet med spesielt fokus på RRI (Responsible Research and Innovation), dvs. samfunnsetisk tenking omkring dette
- kunnskap på et høyt nivå om så vel historiske som samtidige endringsprosesser knyttet til vitenskap, ekspertise, demokrati og teknologi
- inngående kunnskap om hvordan vitenskap og teknologi samproduseres med sosiale, politiske og økonomiske aktiviteter på ulike samfunnsområder

Ferdigheter

Kandidaten kan

- analysere og forholde seg kritisk til problemstillinger knyttet til teknologiutvikling og -bruk, og derigjennom se flere tilnæringsmåter og mulige utfall
- identifisere og arbeide selvstendig med praktiske og teoretiske problemer knyttet til effekter av vitenskap og teknologi i konkrete samfunnsmessige sammenhenger
- utføre avansert kunnskapsmekling i forbindelse med tverrfaglige prosjekter og prosesser

Generell kompetanse

Kandidaten kan

- sette seg inn i og analysere omfattende faglige problemkompleks innenfor en relevant etisk ramme
- anvende sine kunnskaper og ferdigheter på nye områder gjennom tverrfaglige dialoger med eksperter fra andre fagområder
- selvstendig vurdere og bruke ulike framgangsmåter for å bidra til innovasjon og nyskaping på en bevisst og samfunnsetisk måte
- formidle resultater av eget faglig arbeid på en selvstendig måte, både til allmennhet og andre eksperter, muntlig og skriftlig

Sammendrag

Denne masteroppgaven har som hovedfokus å undersøke utviklingsprosessen bak en kunstig intelligens- (AI) løsning for diagnostikk til bruk i helsetjenesten. Norges helsetjeneste sliter med underbemanning, og dette er et problem som kommer til å bli enda større i fremtiden på grunn av eldrebølgen. Den offentlige utredningen «*Tid for handling*» foreslår nye digitale løsninger og teknologi som en løsning på denne utfordringen, og trekker frem AI som en mulig teknologisk løsning (NOU 2023: 4, s. 11, 12, 19, 20 og 258). På grunn av dette, har AI i helsetjenesten blitt et stadig mer aktuelt tema. Denne masteroppgavens problemstilling, er «*Hvordan utvikles AI for diagnostikk i et norsk tverrfaglig forskningsprosjekt?*». I tillegg har masteroppgaven tre underproblemstillinger som er knyttet til hvert av analysekapitlene i oppgaven. Disse er:

«Hvilke fortolkninger og fremtidsvisjoner har prosjektdeltakerne rundt AI?»

«Hvordan er prosjektet organisert?»

«Hvordan forholder prosjektet seg til forsvarlighetskrav til helseteknologi?»

For å besvare disse problemstillingene, har jeg gjennomført seks kvalitative intervju med informanter fra tre ulike institusjoner, som har vært med å samarbeide om å utvikle en AI-løsning til diagnostikk. Oppgaven benytter seg av teorier om sosiotechniske forestillinger, innovasjon, SCOT (med fokus på fortolkningsmessig fleksibilitet) og teori rundt tverrfaglighet.

Analysene viser at samtlige informanter ser for seg at AI kommer til å fungere som et støtteverktøy i fremtidens helsetjeneste. Fem av seks informanter virket også å tolke AI som en «*revolusjonerende*» teknologi med et stort potensiale, mens en av informantene virket å tolke AI som bare et lite hjelpemiddel. Utviklingsarbeidet av denne AI-løsningen er en iterativ prosess som baserer seg på et tverrfaglig samarbeid mellom flere aktører, og hvor prototyper spilte en viktig rolle. Når det kommer til hvordan aktørene jobbet med tanke på forsvarlighet, handlet mye om å samle inn nok data med tilstrekkelig variasjon for å unngå bias. Informantene fortalte også om ulike utfordringer når det kommer til forsvarlighet, samt tiltak de har dersom de ikke klarer å samle inn nok data.

Helt til slutt, ser oppgaven på tvers av analysekapitlene, og finner at fremtidsvisjonene, måten aktørene har organisert utviklingsprosessen på, og hvordan de jobber med forsvarlighet, virker å henge sammen og påvirker hverandre.

Abstract

The main focus of this master's thesis is to study the developmental process behind an artificial intelligence- (AI) solution for diagnostics in the health service. The health service in Norway is facing problems in terms of lack of personnel, a challenge which in the future will grow to become even bigger because of the increase in the amount of elderly people. The public inquiry "*Time for action*" is suggesting new digital solutions and technology as a way to solve this problem and highlights AI as a possible technological solution (NOU 2023: 4, s. 11, 12, 19, 20 and 258). Because of this, AI in healthcare has become a topic of increased relevance. The problem formulation of this master's thesis is "*How is AI for diagnostics developed in a Norwegian interdisciplinary research project?*". In addition to this, the master's thesis has three research questions in relation to each of the analysis chapters in this thesis. These are:

"Which interpretations and visions of the future do the project participants have regarding AI?"

"How is the project organized?"

"How are the soundness requirements regarding health technology addressed in this project?"

To answer these problem formulations, I have conducted six qualitative interviews with informants from three different institutions who have collaborated to develop an AI-solution for diagnostics. The thesis makes use of theories regarding sociotechnical imaginaries, innovation, SCOT (focusing on interpretive flexibility) as well as theory regarding interdisciplinarity.

The analysis finds that all the informants envision AI as a tool for support in future healthcare. Five of the six informants seemed to interpret AI as a "revolutionizing" technology with big potential, whilst one of the informants seemed to interpret AI as merely a minor tool for help. The developmental work of this AI-solution is an iterative process which is based on an interdisciplinary collaboration between multiple actors, and where prototypes played an important part. In regard to how the actors worked in relation to soundness, a big part was about collecting enough data with an adequate amount of variation to avoid bias. The informants also talked about different challenges when it comes to soundness, as well as measures they have if they fail to collect enough data.

At the very end, this thesis looks across the analysis chapters and finds that the visions of the future, the way the actors have organized the developmental process, and how they work in relation to soundness, seem to be connected and influence each other.

Forord

Å skrive masteroppgave har vært en spennende og utfordrende reise. Disse to årene har gått veldig fort, og jeg kan trygt si at de har vært de to beste årene av min skolegang. Det var bare en tilfeldighet at jeg fant dette masterstudiet da jeg var på leting etter masterstudier som jeg kunne være interessert i å ta etter at jeg var ferdig med bachelorgraden. Jeg hadde vært innom nettsiden til bachelorstudiet jeg gikk på for å se på hvilke masterstudier jeg var kvalifisert til, og jeg hadde flere ganger sett et studie som het «Studier av kunnskap, teknologi og samfunn». Men jeg klikket meg aldri inn på dette studiet fordi jeg ikke synes at det hørtes ut som noe som var relevant i forhold til min bachelorgrad. Til slutt klikket jeg meg inn og leste om studiet, og fant med en gang ut at dette var et studie jeg ville søke på, og jeg har ikke angret siden!

Det er flere jeg har lyst til å takke for å ha gjort det mulig for meg å skrive denne masteroppgaven. Først og fremst vil jeg si tusen takk til informantene som tok tid fra sin egen arbeidshverdag for å bli med på masterprosjektet mitt. Takk for at dere var så imøtekommende med meg, det setter jeg veldig stor pris på!

Jeg ønsker også å rette en stor takk til veilederne mine, Kristine Ask og Stig Kvaal. Jeg setter uendelig stor pris på veiledningen deres og alle de gode rådene og innspillene dere har kommet med. Det har vært til stor hjelp. Jeg setter pris på engasjementet dere har vist i veiledningene og hvor lett det har vært å ta kontakt med dere når jeg har lurt på noe. Jeg må også nevne Stig og alle dine metaforer og Kristine som ble like oppgitt hver gang det kom en ny metafor. De humoristiske innspillene satt jeg alltid pris på. Dynamikken dere imellom var kjempebra og til veldig god hjelp, så tusen takk!

Jeg vil også si takk til alle studiekameratene som jeg har vært så heldig å fått gått i klasse med i disse to årene, og spesielt dere som jeg har sittet på lesesal med. På grunn av dere, har jeg gledet meg til å dra på skolen hver dag. Det har vært fint å få dele denne reisen med dere!

Jeg vil også takke venner og familie som har vært gode støttespillere i løpet av denne reisen og som har «pushet» meg i tider hvor jeg har stått litt fast. Tusen takk!

Innholdsfortegnelse

STUDIER AV KUNNSKAP, TEKNOLOGI OG SAMFUNN (STS)	1
SAMMENDRAG	2
ABSTRACT	3
FORORD	4
KAPITTEL 1 – INTRODUKSJON – AI SOM SVAR PÅ HELSETJENESTENS UTFORDRINGER?	7
OPPGAVENS PROBLEMSTILLING.....	7
HVORFOR AI I HELSE?.....	8
HVA ER AI?.....	9
TIDLIGERE FORSKNING: AI I HELSESEKTOREN.....	10
BEKYMRINGER RUNDT AI.....	11
OPPGAVENS STRUKTUR.....	15
KAPITTEL 2 - TEORI – STS OG INNOVASJON, SOSIOTEKNISKE FORESTILLINGER, FORTOLKNINGSMESSIG FLEKSIBILITET, OG TVERRFAGLIGHET	16
STS.....	16
INNOVASJON.....	16
SOSIOTEKNISKE FORESTILLINGER.....	17
SCOT OG FORTOLKNINGSMESSIG FLEKSIBILITET.....	18
TVERRFAGLIGHET.....	19
KAPITTEL 3 - METODE	20
KVALITATIV METODE OG INTERVJU.....	20
VALG AV PROSJEKT OG UTVALG AV INFORMANTER.....	20
INTERVJUGUIDE.....	21
INFORMERT SAMTYKKE, KONFIDENSIALITET OG KONSEKVENSER AV Å DELTA I PROSJEKTET.....	22
KODING.....	22
RELIABILITET, VALIDITET OG OVERFØRBARHET.....	23
KAPITTEL 4 - FREMTIDSVISJONER OG FORTOLKNINGER AV AI	24
RETTFERDIG STØTTEVERKTØY.....	24
ERSTATNING.....	27
TRENING OG OPPLÆRING.....	28
INGEN ENDRING.....	29
OPPSUMMERING.....	30
KAPITTEL 5 – EN ITERATIV PROSESS FOR UTVIKLING	32
MØTENE.....	32
TVERRFAGLIG SAMARBEID.....	34
MILJØ FOR Å KOMME MED INNSPILL.....	39
ITERATIV PROSESS.....	40
PROTOTYPER.....	41
OPPSUMMERING.....	44
KAPITTEL 6 – FORSVARLIGHET	46
UTFORDRINGER MED FORSVARLIGHET.....	47
SKAFFE NOK DATA FOR Å UNNGÅ BIAS.....	49
OVERFITTING SOM RISIKO.....	51
BARE EN FORSKNINGSPLATTFORM.....	53
OPPSUMMERING.....	53

KAPITTEL 7 - KONKLUSJON	55
ALT HENGER SAMMEN.....	56
HVORFOR ET FORSKNINGSPROSJEKT?	58
VEIEN VIDERE - AI SOM SVARET PÅ ALLE UTFORDRINGER?	58
REFERANSELISTE	60

Kapittel 1 – Introduksjon – AI som svar på helsetjenestens utfordringer?

I forordet i Nasjonal Strategi for Kunstig Intelligens fra 2020, står det at kunstig intelligens (AI) kommer til å være viktig for at vi i Norge skal kunne håndtere fremtidens utfordringer som for eksempel eldrebølgen. Videre står det at AI også kommer til å være viktig når det kommer til å klare å opprettholde velferdsnivået i fremtiden (Kommunal- og moderniseringsdepartementet, 2020, s. 2). I en offentlig utredning som heter «*Tid for handling*» fra 2023, står det at den norske helsetjenesten sliter med å ha tilstrekkelig med helsepersonell, og at dette kommer til å bli enda verre i fremtiden. I tillegg til dette, skrives det at norsk helsepersonell har blitt utsatt for et større press den siste tiden. Videre skriver de at det de neste årene ikke bare vil være en fortsatt økning av eldre, men at antallet personer som er i en alder hvor de kan jobbe, vil reduseres. For å kunne minske behovet vi har for helsepersonell, foreslås digitale løsninger samt teknologi som en løsning. Videre skriver de at behovet for helsepersonell kan reduseres og produktiviteten økes, dersom man bruker teknologiske løsninger på riktig måte. I den anledning trekker de frem kunstig intelligens og viser til en rapport fra Teknologirådet om hvordan AI kan brukes i helsetjenesten i fremtiden. Her trekkes det frem områder som medisinsk forskning, anvendelser relatert til folkehelse og kliniske anvendelser, samt administrasjon (NOU 2023: 4, s. 11, 12, 19, 20, 258). Som vi kan se her, har den norske helsetjenesten flere store utfordringer, blant annet at de ikke har nok helsepersonell. I tillegg ser vi at helsetjenesten vil møte på en stor utfordring i form av den «berømte» eldrebølgen, som vil slite enda mer på den altfor tynne bemanningen, som det også kommer til å bli mer strekk i. Videre kan vi se at de lanserer ny teknologi, blant annet i form av AI som en potensiell løsning på disse problemene. Derfor har AI i helsetjenesten blitt mer og mer aktuelt.

Oppgavens problemstilling

AI fremstilles som en teknologi som kommer til å spille en stor rolle i både helsetjenesten og samfunnet generelt i fremtiden. Et relevant spørsmål å stille, er dermed hvordan denne typen teknologi blir til? I denne masteroppgaven, ser jeg derfor nærmere på utviklingen og utviklingsprosessen av en AI-løsning for diagnostikk til helsetjenesten. Utviklingsprosessen bak AI-løsninger er litt «black-boxed», som vil si at det er vanskelig å vite hva som har inngått i selve utviklingen av AI-teknologien, man ser bare resultatet, men ikke hvordan og hvorfor den ble som den ble (Skjølsvold, 2017, s. 21). AI-teknologien er i rask vekst og vil nok få større og større innvirkning på livene våre i tiden fremover. Derfor kan det også være viktig å ha kunnskap om hvordan denne teknologien blir utviklet. På denne måten vil AI som teknologi, kunne bli mer transparent for brukere, og ikke så «mystisk» som den er nå. I oppgaven har jeg fokusert på en forskningsinstitusjon, som jeg kaller for «Forskningsinstitusjonen», som utvikler AI-løsninger til helsetjenesten.

Masteroppgavens problemstilling er: «*Hvordan utvikles AI for diagnostikk i et norsk tverrfaglig forskningsprosjekt?*». Problemstillingen er utviklet med et formål om å kunne gi innsikt i hvordan akkurat denne AI-teknologien, som jeg undersøker her, som er ment for diagnostisk bruk i helsetjenesten blir, eller kan bli, utviklet. I tillegg til denne, generelle problemstillingen, har jeg også tre underproblemstillinger som baserer seg på de tre analysekapitelene. Disse er:

«Hvilke fortolkninger og fremtidsvisjoner har prosjektdeltakerne rundt AI?»

«Hvordan er prosjektet organisert?»

«Hvordan forholder prosjektet seg til forsvarlighetskrav til helseteknologi?»

For å undersøke dette, har jeg gjennomført seks semistrukturerte kvalitative intervju med informanter som har vært med på utviklingen av denne AI-løsningen, hvor jeg har spurt dem om idé- og arbeidsprosess, samarbeid, beslutningsprosess og hvordan de jobber med å gjøre AI-løsningen så forsvarlig som mulig. I tillegg til dette, spurte jeg informantene om hva de tror AI i helse er om 10-20 år. Informantene har jeg kategorisert som «Forskere», «Klinikere» og «Designer».

Hvorfor AI i helse?

Det er knyttet optimisme til AI i helsetjenesten. Blant annet i media kan man se store overskrifter om AI i helsevesenet, for eksempel «*Kunstig intelligens: Kan ha svaret på vår tids største helsetrusler*» (Nordanger og Ludvigsen, 2023). Som vi kan se av denne overskriften, fremstår det som at AI vil kunne løse alle problemer innen helsetjenesten. I den samme artikkelen, kan man også se at Ishita Barua, som er lege og som i tillegg har doktorgrad i AI innen medisin, mener at AI holder på å revolusjonere helsetjenesten (Nordanger og Ludvigsen, 2023). Hun har også skrevet en bok som heter «*Kunstig intelligens redder liv – AI er legens nye superkrefter*» (Barua, 2023). Dette gir et eksempel på hvordan noen som faktisk jobber i helsetjenesten, og som i tillegg har god kompetanse når det kommer til AI i medisin, ser på AI i helsetjenesten. Det er også flere som bruker ordet «revolusjon» når de snakker om AI i helsekontekst. I en kronikk på Dagens Medisin sine nettsider, stiller forfatterne et spørsmål i kronikkens overskrift om AI i helsetjenesten er «en stille revolusjon» (Linstad et al., 2021).

Som vi kan se av avsnittet over, så blir ordet «revolusjon» brukt flere ganger når det snakkes om AI i helsevesenet. Dette sier noe om hvor stor rolle mange mener at AI kan ha i helsetjenesten. Det er mange forventninger til AI i helse, og det virker som at mange ser på AI som en teknologi som vil kunne løse alle problemene man har i helsetjenesten. Men er dette tilfellet? Vi har et ferskt eksempel på en ny teknologi som holder på å bli innført i helsevesenet som ikke akkurat kan sies å ha vært en suksess, nemlig Helseplattformen. Selv om AI sitt potensiale er stort og tanken er god, er det ikke sikkert at det er like rett frem som man håper. Mindre komplekse teknologiske løsninger enn AI, som Helseplattformen, har vist seg å være svært problematisk å innføre i helsetjenesten, noe som kan tyde på at det ikke nødvendigvis vil være noe enklere å skulle implementere AI-løsninger og systemer i helsevesenet. Å få AI inn i helsetjenesten, vil kanskje ikke være så rett frem som mange håper.

Til tross for dette, blir AI fortsatt fremstilt som en teknologi som kan forbedre helsetjenesten. Aung, Wong og Ting (2021) viser til at AI er foreslått som en løsning som kan være sentral for å løse problemet med underbemanning. I tillegg til dette, skriver de at AI kan både forbedre behandling av pasienter, samt redusere arbeidsmengde. De peker videre på spesielt fire fordeler AI kan bringe med seg; oppdeling av arbeidsoppgaver, styrke klinisk praksis, redusere arbeidsmengde og erstatte visse arbeidsoppgaver (Aung, Wong og Ting, 2021, s. 5, 6 og 7). Sunarti et al. (2020) trekker

frem noen andre fordeler som AI kan ha i helsesektoren. Disse er blant annet at AI kan redusere antallet henvisninger, spare tid, forbedrede både resultater og valg når det gjelder behandling av pasienter, og føre til kostnadsreduksjon (Sunarti et al., 2020, s. 68).

Optimismen rundt AI er som nevnt stor, ikke bare innen helse, men i samfunnet generelt, med en tilhørende narrativ om at AI vil fikse alle problemer og gjøre alles liv mye lettere. Optimismen og troen på AI ser vi tydelig hos digitaliseringsministeren Karianne Tung som har satt et mål om at innen 2025, så skal hele 80 % av Norges offentlige sektor bruke AI. Tung mener at for at man skal kunne både forbedre og fornye den offentlige sektoren, så vil AI være et viktig redskap (Finansavisen, 2024). At troen og forventningene er stor, kan vi også se av satsingen Regjeringen har på AI. På sine egne nettsider, skriver Regjeringen at den satser 1 milliard kroner på forskning på AI (*Regjeringen med milliardatsing på kunstig intelligens*, 2023). Det er også tydelig at man forventer at AI kommer til å være en stor del av samfunnet generelt i fremtiden, og at dette er noe man forbereder seg til. Dette kan vi blant annet se av at det har blitt utarbeidet en «*Nasjonal strategi for kunstig intelligens*» (*Nasjonal strategi for kunstig intelligens*, u.å.), som jeg nevnte tidligere.

Hva er AI?

Så langt har vi sett på hvorfor AI i helsetjenesten er blitt så relevant, men hva er egentlig denne teknologien som skal «*revolusjonere*» helsetjenesten og resten av samfunnet? AI er et begrep som de fleste har hørt om, men det finnes ikke én enkelt definisjon på hva AI er. I *Nasjonal strategi for kunstig intelligens*, står det at AI blant annet «*... utfører handlinger, fysisk eller digitalt, basert på tolkning og behandling av strukturerte eller ustrukturerte data ...*» (Kommunal- og moderniseringsdepartementet, 2020, s. 9). Når Ergen (2019) skriver om hva AI er, sier han at dette er en maskins evne til å utføre ulike kognitive funksjoner, eksempelvis som samhandling og læring (Ergen, 2019, s. 5). Panch, Mattie og Atun (2019) kommer med en annen definisjon av AI enn de to foregående. De sier at AI dreier seg om at «*algoritmer avdekker eller lærer assosiasjoner av prediktiv kraft fra data*» (Panch, Mattie og Atun, 2019, s. 1).

Hva sier det faktum at det ikke finnes én definisjon av AI oss? Hva har dette å si for AI som fenomen? Det forteller oss kanskje at AI er en veldig kompleks teknologi, men det kan også fortelle oss hvordan samfunnet tolker AI som et begrep og som teknologi. Det kan også fortelle oss at det er mange ulike teknologier som blir plassert under begrepet «AI». Kanskje selve begrepet AI eller kunstig intelligens har blitt et litt flytende begrep som fører til at mange ulike teknologier går under kategorien kunstig intelligens? Dette er også noe som Lucy Suchman (2023) skriver om i sin kommentar «*The uncontroversial 'thingness' of AI*». Der stiller hun seg spørsmålet om hva det er man egentlig mener og snakker om, når man prater om AI (Suchman, 2023, s. 1). Det finnes mange ulike typer AI-teknologier som løser mange forskjellige oppgaver, alt fra såkalte «*chat-boter*» til AI-løsninger som skal hjelpe klinikere med å analysere CT-bilder. Alle de ulike typene AI-teknologi og de ulike områdene de brukes på, kan føre til at det blir vanskelig å definere akkurat hva AI er. Som vi kan se, nevner Panch, Mattie og Atun (2019) begrepet «*algoritmer*» i sin definisjon av AI. Hva er så en algoritme? Ifølge Inga Strümke (2023) kan man tenke på algoritmer som en oppskrift. De er instruksjoner som oppnår et mål ved at de utføres i en gitt rekkefølge, og er dermed helt essensielle for at en datamaskin skal virke (Strümke, 2023, s. 18).

Maskinlæring er en tilnærming innen AI som er veldig i vinden nå for tiden (Strümke, 2023, s. 58). Hva maskinlæring går ut på, ligger litt i navnet. Det går ut på at maskiner på egenhånd lærer seg hvordan de skal både løse ulike problemer og utføre ulike oppgaver. Det handler om prøving og feiling i den forstand at maskinen forsøker å løse en oppgave og holder på helt til den lykkes. På denne måten tilegner maskinen seg kunnskap. Det er tre faktorer som må til for at dette skal være mulig. Disse er ifølge Strümke (2023) at; «maskinen må ha en oppgave å løse», «maskinen må ha data å prøve og feile på», og «maskinen må være i stand til å lære» (Strümke, 2023, s. 58). Disse maskinene trenger derfor såkalte «maskinlæringsalgoritmer». Det er disse algoritmene som gjør at maskiner i det hele tatt kan lære (Strümke, 2023, s. 58).

Som du sikkert har sett så langt, så skriver jeg «AI» og ikke «KI» når jeg skriver om kunstig intelligens. Grunnen til at jeg i denne oppgaven vil forkorte begrepet «kunstig intelligens», til den engelske forkortelsen «AI» og ikke den norske som er «KI», er fordi alle informantene mine referer til kunstig intelligens som AI i intervjuene jeg har hatt med dem. Dermed er det enklere for meg å også bruke AI slik at det ikke blir en blanding av både AI og KI i oppgaven.

Tidligere forskning: AI i helsesektoren

Mange snakker om hva AI skal bringe til helsetjenesten i fremtiden, men noen plasser har AI allerede begynt å implementeres i helsetjenester. For å finne ut av hvordan AI har blitt brukt i helsetjenester, har Yin, Ngiam og Teo (2021) gjort en litteraturstudie, og fant ut at blant de som hadde implementert AI i helsetjenesten, så var bruken av nevralt nettverk den maskinlæringsteknikken som var mest populær. Videre, så de at AI som oftest ble brukt som støtteverktøy når det kom til screening av sykdommer eller «triage», diagnose av sykdommer, risikoanalyse og behandling. Når det kommer til hvorvidt AI kan forbedre helsetjenesten, fant de at 16 av studiene viste at kapasiteten når det kommer til klinisk beslutningstaking, kan bli forbedret av AI. I tillegg til dette, viste to studier ingen tegn for at klinisk beslutningstaking ble forbedret ved bruk av AI. Når det kommer til innvirkningen AI har på klinikernes effektivitet og arbeidsflyt, fant de ut at seks av syv studier viste at arbeidsflyten ble forbedret ved bruk av AI, samt at den nødvendige tiden når det kom til kliniske oppgaver ble redusert (Yin, Ngiam og Teo, 2021, s. 1, 9 og 10). Det som er interessant å ta med seg videre fra det Yin, Ngiam og Teo (2021) har funnet, er at AI allerede er tatt i bruk i helsetjenester rundt om i verden og at den i størst grad blir brukt som et hjelpemiddel, eller støtteverktøy for klinikerne, i tillegg til at flesteparten av studiene de fant, viste at klinikernes beslutningstaking, samt arbeidsflyt og effektivitet ble forbedret ved hjelp av AI.

I likhet med Yin, Ngiam og Teo (2021), har Chomutare *et al.* (2022) også undersøkt implementering av AI i helsetjenester, men de har også sett på hvilke tilretteleggere og barrierer som eksisterer når det kommer til å implementere AI i helsetjenesten. De har gjennomgått litteratur fra ulike databaser som PubMed, ACM, IEEE, Google Scholar og the Web of Science gjennom å ha gjort et «scoping review». Et funn de hadde, var at AI som oftest ble brukt i screening. Når det kommer til tilretteleggere og barrierer, fant de ut at de aller vanligste tilretteleggerne når det kommer til AI i helsetjenesten gikk på engasjement og ledelse, den nest vanligste tilretteleggeren gikk på testing og evaluering mens den tredje vanligste tilretteleggeren gikk på teknisk design. Når det kommer til barrierer, viser studien at de vanligste barrierene for AI, er spørsmål knyttet til

interoperabilitet, datakvalitet og styring, i tillegg til åpenhet og tillit. I konklusjonen, skriver de at man ikke kan fullt forstå implementering av AI i helsetjenesten fordi man ikke har tilstrekkelig med kunnskap når det kommer til ulike faktorer som eksempelvis samarbeidsnettverk, insentiver eller retningslinjer som enten hindret eller hjalp AI i å bli implementert i helsetjenesten (Chomutare *et al.*, 2022, s. 3, 4, 7, 9 og 15).

Det å ta i bruk AI i helsetjenesten, fører også med seg ulike bekymringer. Ifølge Park *et al.* (2020) er dette bekymringer knyttet til pålitelighet og sikkerhet. De nevner blant annet flere ulike utfordringer når det kommer til å skulle bruke AI i helsetjenesten, eksempelvis problemer når det kommer til å bruke helsedata, retningslinjer og regulatoriske forhold for nye enheter, en anvendelse som er balansert med det eksisterende helsevesenet, samt ansvars- og sikkerhetsspørsmål (Park *et al.*, 2020, s. 1, 2 og 5-7). Her ser vi dermed at det er flere utfordringer når det kommer til AI i helsetjenesten. I den neste delen, skal vi derfor se nærmere på bekymringer knyttet til AI.

Bekymringer rundt AI

Det er flere bekymringer når det kommer til AI, og bias er en av dem. Under bias, finner man flere andre bekymringer. En bekymring dreier seg om representasjon, en annen dreier seg om treningsdata, en tredje bekymring dreier seg om underdiagnostiseringsbias. Man har også bekymringer knyttet til algoritmisk bias. Bias er et tema som kommer til å få en del fokus i denne masteroppgaven. Men hva er bias egentlig? Ifølge Inga Strømke (2023) brukes bias for å «beskrive systematiske avvik fra målet, som alle går i samme retning» (Strømke, 2023, s. 92). Det norske ordet for bias er «forventningsskjevhet». Dette innebærer at det er en skjevhet i hvilke svar man kan forvente at modellen vil gi. Videre legger Strømke (2023) vekt på at man må samle inn nok data og at det må samles inn på en måte som gjør at disse dataene er kapable til å estimere det som er den sanne fordelingen. Dersom dataene man har samlet inn representerer den sanne fordelingen på en dårlig måte, så vil det føre til en dårlig modell som ikke virker (Strømke, 2023, s. 85, 92 og 93).

Nå har vi sett på en generell forklaring av hva bias er, men hva vil bias i helse si? Obermeyer *et al.* (2019) gir et eksempel på hva dette kan være i en studie de har gjennomført der de har studert et stort datasett som hører til en algoritme som blir brukt i helsetjenesten i USA. Studien viste at denne algoritmen har bias når det kommer til etnisitet og at dette førte til at mange mørkhudede pasienter som trenger ekstra behandling, ikke får det. Studien viste at prosentandelen mørkhudede pasienter som mottar ekstra behandling, vil øke betraktelig dersom denne forskjellen som oppstår i denne algoritmen blir rettet opp i (Obermeyer *et al.*, 2019, s. 1). Dersom det forekommer bias i AI-løsningene som blir brukt i helsetjenesten, så vil dette gå negativt utover pasientsikkerheten. Dette kan, ifølge Celi *et al.* (2022), eksempelvis være at det vil skape helseforskjeller (Celi *et al.*, 2022, s. 12). Det er derfor viktig å ha fokus på bias når det gjelder AI i helse. Som man kan se her, kan bias når det kommer til AI i helse, føre til at man ikke får den behandlingen man trenger fordi algoritmene ikke fungerer godt nok. Derfor er bias en veldig sentral utfordring i helse.

Hvordan oppstår så bias i AI for helsevesenet? Celi *et al.* (2022) har skrevet om kilder til bias i kunstig intelligens som fører til helseforskjeller. Ved å bruke maskinlæring, har de gått gjennom kliniske artikler i PubMed som ble publisert i 2019 for å undersøke nettopp

dette. De så blant annet på forskjeller når det kom til forfatterens kjønn, nasjonalitet og ekspertise, i tillegg til klinisk spesialitet og hvilket land datasettet kom fra. De kom frem til at den kliniske spesialiteten som var mest representert var radiologi, og at USA og Kina var de landene som de fleste databasene de fant kom fra. I tillegg til dette, kom de frem til at første- og sisteforfatterne som oftest ikke var klinikere, men dataeksperter, at landene som forfatterne i størst grad kom fra var Kina eller USA og at første- og sisteforfatterne stort sett var menn. Det å mate modellene med lite mangfoldig data kan potensielt være et problem som kan føre til bias, ifølge Celi *et al.* (2022). Videre skriver de at kliniske AI-modeller kan fungere helt fint på det medisinske senteret de er blitt utviklet og bygget ved, men at de kan mislykkes når de blir tatt i bruk på andre steder. Videre understreker de viktigheten av å forstå hvem som, på grunn av underrepresentasjon, kan bli påvirket på en negativ måte av modellskjevheter (Celi *et al.*, 2022, s. 1-3 og 15).

Challen *et al.* (2019) snakker om ulike problemer med AI-algoritmer i helsevesenet som kan føre til blant annet bias. De trekker frem en utfordring med tilgjengelige treningsdata, og skriver at tilgjengelige treningsdata ikke alltid vil være representative når det gjelder normalen fordi de har blitt samlet inn på grunn av at de er interessante, ikke fordi de er representative. Dette vil da kunne føre til bias. Videre viser de til et eksempel der et maskinlæringsystem fungerte på en OCT-maskin, men ikke på en annen. Dette kan skje dersom man bruker et maskinlæringsystem i en annen kontekst på feil måte. Challen *et al.* (2019) skriver også at treningsdata ikke passer med den operasjonelle dataen på grunn av at sykdomsmønster endrer seg over tid. De nevner også at resultatene til enkelte maskinlæringsalgoritmer er «black-boxed» og gir prediksjoner hvor det er vanskelig å se hvorfor den kom frem til akkurat det resultatet. Når dette skjer, er det ifølge Challen *et al.* (2019) vanskelig å oppdage bias og feil (Challen *et al.*, 2019, s. 233).

Bender *et al.* (2021) trekker frem flere aspekter ved såkalte språkmodeller og «natural language processing» som er problematiske med tanke på bias. Dette gjelder blant annet at språkmodeller kan plukke opp språkmønstre som er fornærmende, samt det som kalles for «subtil bias» (Bender *et al.*, 2021, s. 611). Bilotta *et al.* (2019) viser til en definisjon av subtil bias, riktignok i en annen sammenheng enn det som har med AI å gjøre, og skriver at subtil bias «er en diskret fordom eller preferanse overfor en bestemt gruppe, person eller ting som kan drive ens beslutninger og handlinger» (Bilotta *et al.*, 2019, s. 228). Bender *et al.* (2021) skriver at dette kan medføre diskriminering og skaderisiko. Dette vil si at språkmodeller kan bli trent opp på språkdata som inneholder rasisme, de-humanisering og bias når det kommer til kjønn. Disse negative assosiasjonene vil bli plukket opp av språkmodeller dersom de blir trent på denne typen data. Videre sier Bender *et al.* (2021) at treningsdataene som disse språkmodellene får, kan føre til at språkmodellene koder assosiasjoner som er fornærmende og stereotypiske når det gjelder eksempelvis kjønn og rase. Det blir også trukket frem at det er en overrepresentasjon av personer som er i posisjoner som er privilegerte i treningsdata og at dette gjør at det finnes kodete bias i denne treningsdataen (Bender *et al.*, 2021, s. 611, 613 og 617). Når noen er overrepresentert, vil det føre til skjevhet i forhold til andre grupper som ikke er overrepresentert.

Det er flere som trekker frem representasjon og treningsdata når det kommer til bekymringer rundt AI, blant annet Larrazabal *et al.* (2020). De fremhever underrepresentasjon som et problem, og argumenterer for at når man skal designe og

utvikle databaser som skal brukes til å trene maskinlæringsbaserte CAD-systemer (computer-aided diagnosis), så burde mangfold prioriteres. Dette begrunner de med å vise til resultatene de fikk når de gjennomførte en studie som så på ubalanse av kjønn i medisinske bildedatasett som brukes for å trene AI-baserte CAD-systemer og hvilken påvirkning dette har. Resultatene av dette studiet, var at bruken av datasett uten balanse mellom kjønnene for å trene dyplærings-baserte CAD-systemer, kan påvirke ytelsen når det kommer til patologiklassifisering for minoritetsgrupper. Noen av deres resultater var at gjennomsnittsytelsen til en modell med 25% og 75% ubalanseforhold var «betydelig» lavere enn en modell som er trent med et perfekt balansert datasett. Videre fant de også ut at for å få den beste ytelsen for begge kjønn fra et CAD-system, så må dette CAD-systemet trenes med et balansert og mangfoldig datasett. Larrazabal *et al.* (2020) skriver også at basert på deres resultater, så har de fått en indikasjon på at mangfold fører til ytterligere informasjon eller tilleggsinformasjon, noe som øker AI-systemers generaliseringsevne (Larrazabal *et al.*, 2020, s. 12 592 og 12 593).

I avsnittet over ble det snakket om underrepresentasjon som en bekymring rundt AI. Kalantari *et al.* (2021) trekker frem det de kaller for «underdiagnostiserings-bias» av kunstig intelligens-algoritmer. For å undersøke dette, har de sett på AI-baserte forutsigelsesmodeller for røntgen for brystet og gjennomførte en systematisk studie av underdiagnostiserings-bias i disse modellene. De har sett på tre store offentlige radiologi-datasett i tillegg til at de har sett på et datasett med flere kilder som kombinerer disse tre modellene på de samme sykdommene. Kalantari *et al.* (2021) fant ut at såkalte «klassifikatorer» som blir produsert ved å bruke avanserte datasynsteknikker, underdiagnostiserte underbetjente pasient-populasjoner konsekvent og selektivt. I tillegg til dette, fant de ut at frekvensen eller hyppigheten av underdiagnostiseringen var høyere for såkalte «interseksjonelle underbetjente underpopulasjoner» (Kalantari *et al.*, 2021, s. 2176).

Mutasa, Sun og Ha (2020) trekker frem enda et stort problem for AI-teknologi; overfitting. De skriver at overfitting vil si at AI-modellen har blitt lært opp på en måte som gjør at den bare kan brukes på treningsdata, men ikke på andre typer data. Den er altså ikke generaliserbar til resten av populasjonen. Mutasa, Sun og Ha (2020) gir et eksempel med en algoritme som skal lære seg å skille mellom katter og hunder. Algoritmen har bare fått trene på bilder av én hunderase og én katterase. Da vil algoritmen gjøre det veldig bra på akkurat disse treningsdataene, men når den får se bilder av andre typer katter og hunder enn det den har blitt trent på, og skal skille mellom dem, så vil den fungere betraktelig dårligere (Mutasa, Sun og Ha, 2020, s. 96). Overfitting skjer fordi man, uten å vite det, har «trukket ut noe av gjenværende variasjon som om den variasjonen representerte underliggende struktur» (Mutasa, Sun og Ha, 2020, s. 96). Det finnes derimot måter å håndtere problemet med overfitting på. En måte å gjøre det på, er å samle inn mer treningsdata. Dersom det ikke er mer tilgjengelig treningsdata, kan man foreta «data augmentation». Data augmentation, eller augmentering, vil si at man kan øke størrelsen på et treningsdatasett på en kunstig måte ved at man eksempelvis roterer bildene, gjør bildene skjeve eller beskjærer bildene. På denne måten kan man lage flere versjoner av de eksisterende bildene man har i treningsdatasettet (Mutasa, Sun og Ha, 2020, s. 97 og 98).

Frem til nå har jeg fokusert mye på AI og bias i helsetjenesten, men dette er ikke et særegent problem når det kommer til helsetjenesten, det er et problem som gjelder AI på generell basis. Roselli, Matthews og Talagala (2019) skriver at bedrifter som ønsker å

ta i bruk AI-algoritmer vil møte på et risikomoment i form av bias i disse algoritmene. I tillegg til dette, sier de at det kan hende at det vil være en umulig oppgave å skulle klare å kvitte seg med all bias, og begrunner dette med at det blir brukt historiske data som inneholder historiske bias for å trene AI-løsninger. Videre nevner de blant annet bias i «utlånspraksis». Denne typen bias vil være uheldig fordi den eksempelvis kan gjøre at den diskriminerer mot enkelte personer, eller at noen personer blir «foretrukket» fremfor noen andre (Roselli, Matthews og Talagala, 2019, s. 539). Forskjellen på bias i helsetjenesten og bias i andre sektorer i samfunnet, er at konsekvensene av bias i helsetjenesten er mye større enn i andre deler av samfunnet, da bias i verste fall kan føre til dødsfall.

For å avslutte denne delen om bias, kan vi se på noe som kalles for «algoritmisk bias». Panch, Mattie og Atun (2019) definerer «algoritmisk bias» slik; «tilfellene når anvendelsen av en algoritme forener eksisterende ulikheter i sosioøkonomisk status, rase, etnisk bakgrunn, religion, kjønn, funksjonshemming eller seksuell legning for å forsterke dem og påvirke ulikheter i helsesystemer negativt» (Panch, Mattie og Atun, 2019, s. 1). Videre skriver de at helsesystemet vil komme til å møte på tre utfordringer når det skal takle algoritmisk bias; «utilstrekkelig kontekstuell spesifisitet», «mangel på klare definisjoner og standard for rettferdighet», samt ««black-box»-naturen til algoritmer» (Panch, Mattie og Atun, 2019, s. 2). Utfordringen når det kommer til utilstrekkelig kontekstuell spesifisitet går, ifølge Panch, Mattie og Atun (2019), ut på at det er et stort mangfold av mennesker, samt at helsesystemer varierer både når det gjelder mål og design. Her skriver de at det burde brukes en AI-modell som er «generelt anvendelig» og som blir trent på data som faktisk er representativ for mangfoldet i helsesystemer, men at dette ikke er tilfellet fordi tilgjengeligheten av data for ulike sosioøkonomiske grupper varierer slik at man ikke får samlet inn like mye data for hver gruppe. Mangel på klare definisjoner og standard for rettferdighet, handler om at det er utfordrende å på en generell måte kunne definere en «standard for rettferdighet». Panch, Mattie og Atun (2019) trekker frem at det i både helsetjenester og i samfunnet, er systematiske ulikheter, og at en «standard for rettferdighet» dermed er vanskelig å definere. Når det kommer til «black-box»-naturen til algoritmer, snakkes det her om såkalt «dyp læring». Panch, Mattie og Atun (2019) skriver at pasienter, dataforskere og klinikere har rett til, og trenger å få vite hvordan akkurat det resultatet var det som algoritmen kom frem til, men at dette er vanskelig å få til (Panch, Mattie og Atun, 2019, s. 2).

Mange av utfordringene man har innenfor AI i helse, er ikke nye, og de er heller ikke spesielt avgrenset til AI i helse. Store utfordringer som det å skaffe nok data og bias, er utfordringer som gjelder AI generelt. Inga Strümke skriver i boken «Maskiner som tenker – Algoritmenes hemmeligheter og veien til kunstig intelligens» (Strümke, 2023), at det er mange eksempler på maskinlæringsmodeller som er dårlige på grunn av dataene de er trent på. I tillegg skriver hun at dataen man bruker, kommer fra vår verden der blant annet bias og rasisme eksisterer og at dette derfor gjenspeiles i data som blir samlet inn. Hun trekker også frem Amazon sin AI-modell som skulle fungere som et rekrutteringsverktøy. Denne AI-en diskriminerte mot kvinner fordi den var trent på data hvor menn som oftest var representert. Strümke (2023) viser også til et annet eksempel med en AI-modell i den amerikanske helsetjenesten som diskriminerte afroamerikanske pasienter (Strümke, 2023, s. 88, 89, 90). Dette viser oss at problemstillingene rundt eksempelvis dataene som blir samlet inn og bias, ikke er nye.

Oppgavens struktur

Masteroppgaven er strukturert i sju kapitler, hvorav den første er innledningen som du nå har lest. Kapittel 2 tar for seg den teorien jeg bruker i oppgaven. I kapittel 3 gjør jeg rede for den metoden jeg har brukt. I kapittel 4 tar jeg for meg informantenes fremtidsvisjoner når det kommer til AI i helsetjenesten, samt hvilke fortolkninger de har av AI som teknologi. Kapittel 5 handler om den iterative arbeidsprosessen bak denne AI-løsningen, mens kapittel 6 fokuserer på forsvarlighet. Kapittel 7 er konklusjonen på det jeg har undersøkt i denne masteroppgaven.

Kapittel 2 - Teori – STS og innovasjon, sosiotechniske forestillinger, fortolkningsmessig fleksibilitet, og tverrfaglighet

I denne masteroppgaven, skal jeg undersøke hvordan informantene forstår AI, altså hvordan tolker de AI som en teknologi, samt hvilke fremtidsvisjoner de har når det kommer til AI i helsetjenesten. I tillegg til dette, skal jeg undersøke hvordan arbeidsprosessen bak denne AI-løsningen som jeg har sett på her, fører til det resultatet som det gjør. Jeg skal også se på hvordan aktørene jobber med tanke på forsvarlighet. Disse ulike problemstillingene gjør derfor at jeg skal benytte meg av teori rundt innovasjon, sosiotechniske forestillinger, SCOT og fortolkningsmessig fleksibilitet, samt teori rundt tverrfaglighet. Siden dette er en masteroppgave innen STS, kan vi først start med å se litt på hva STS er.

STS

Science and Technology Studies (STS) studerer samfunn, teknologi og vitenskap. Innenfor teknologi og vitenskap, er STS spesielt interessert i å se på produksjon av kunnskap, hvor man fokuserer på å se på de prosessene som fører til at enkelte ting blir sett på som fakta, mens andre ting blir forkastet, og utvikling av teknologi hvor man ser på forholdet mellom teknologi, kultur, samfunn og sosiale prosesser. Ifølge Skjølsvold (2017), går STS vekk fra det som kalles «internalistiske forklaringer» når det kommer til hvordan man kan både forklare og forstå vitenskap. Istedenfor fremhever STS eksterne forklaringer når det kommer til at vitenskapelige fakta blir som de blir. Det er ikke bare vitenskapen selv som kan forklare vitenskap, den kan også forklares av eksterne forhold som samfunn og kultur (Skjølsvold, 2017, s. 9, 10, 14 og 15).

Særlig relevant for studiet av AI, er hvordan STS presenterer teknologiutvikling og hvordan teknologi påvirker samfunnet. STS forkaster den «teknologideterministiske» tankegangen. Teknologideterminisme vil si at man mener at teknologien former samfunnet, og at samfunnet ikke former teknologien. Teknologien fortsetter sin fremmarsj, uten at vi mennesker kan påvirke det. Innen STS, har man et motsatt synspunkt på teknologi. Her mener man at blant annet det historiske, sosiale og kulturelle, former teknologien, samtidig som at aktører i samfunnet er de som produserer teknologi. Videre sier Skjølsvold (2017) at man innen STS har et empirisk syn på potensielle påvirkninger som teknologien kan ha, også når det kommer til hvilken effekt teknologien kan ha for samfunnet, samt utviklingen av teknologi (Skjølsvold, 2017, s. 21).

Innovasjon

Innovasjon er et begrep som er mye brukt i mange ulike sammenhenger. På Helse Midt-Norges nettsider definerer de hva innovasjon i helse vil si på denne måten: «Utvikling av nye produkter, tjenester eller organisasjonsformer som bidrar til en styrket helsetjeneste. Det skjer i form av økt kvalitet, bedre arbeidsprosesser, økt sikkerhet for pasienter og ansatte, og slik bidra til verdiskaping» (*Innovasjon*, 2024). Denne definisjonen fokuserer kun på innovasjon innen helse, og er teknologioptimistisk. Kahn (2018) definerer innovasjon med et annet perspektiv, og skriver at innovasjon ikke bare er én ting, men heller tre ulike ting. Innovasjon er en prosess, et resultat og et tankesett

(Kahn, 2018, s. 453). Videre nevner han to ulike måter å definere innovasjon på. Den første handler om at innovasjon er at noe nytt blir introdusert, mens den andre definisjonen sier at innovasjon er «en ny idé, metode eller enhet» (Kahn, 2018, s. 454). En innovasjonsprosess kan deles inn i fem faser. Disse er ifølge Amundsen (2017); ideer, prioritering, utvikling, implementering og effekt/gevinst (Amundsen, 2017, s. 2).

Brita Solveig Pukstad, som er prodekan for innovasjon og PhD-utdanning, skriver i en fagblogg angående helse og medisin som er knyttet til NTNU om viktigheten av innovasjon i helse. Her skriver hun at ressursene i helsetjenesten må bli utnyttet på en bedre måte og at viktigheten av innovasjon i helsetjenesten derfor er stor (Pukstad, 2018). Helse Midt-Norge skriver på sine nettsider at for at helsetjenesten skal kunne forbedres og utvikles, så er innovasjon essensielt (*Innovasjon*, 2024).

Prototyper er «en tilnærming av produktet langs en eller flere dimensjoner» (Das og Das, 2021, s. 984). Kelly (2001) nevner flere fordeler med å utvikle prototyper, blant annet at det vil gjøre det lettere å ta vanskelige valg, å godta nye ideer eller å ombestemme seg eller bytte mening, i motsetning til eksempelvis en vanlig rapport, som er mye lettere å bare avvise. I tillegg til dette, skriver Kelly (2001) at det å vise folk prototyper hvor man kan se hvordan de ulike delene jobber sammen og passer sammen, formen og størrelsen, vil kunne gi folk et nytt perspektiv og få dem til å bli klar over ting de ikke hadde tenkt på tidligere. Videre mener Kelly at utvikling av prototyper er viktig for innovasjon og sier at dersom man slutter å utvikle prototyper, så vil innovasjonen stoppe. Han viser også til at prototyper gjør at man kan finne nye ting, men også at man kan finne ut hva som ikke fungerer tidligere i prosessen (Kelly, 2001, s. 39 og 40).

Das og Das (2021) viser til at det i innovasjonsfaser ofte forekommer «uønskede prosjektavslutninger» (Das og Das, 2021, s. 983), som oftest i fasene som går på utvikling og produksjon, og introduksjon til markedet, og skriver derfor at det viktig at man legger opp innovasjonsprosessen på en slik måte at man har valgt den beste ideen til når man skal inn i utviklingsfasen. Videre viser de til at forskning sier at det å, på et tidligere stadiet utvikle prototyper, kan være en god løsning på dette problemet fordi det kan være viktig faktor som kan inspirere til at man kommer med nye ideer, noe som vil fremme en innovasjonskultur (Das og Das, 2021, s. 984). I likhet med det Kelly (2001) sier om at prototyper kan føre til at man finner ut hva som ikke fungerer tidligere i prosessen (Kelly, 2001, s. 40), sier Das og Das (2021) at forskning viser at prototyper gjør at man kan feile tidlig i prosessen og at dette vil kunne sikre at man i slutten av prosessen, lykkes. I tillegg til dette, vil sannsynligheten for at man velger den beste løsningen eller ideen være større dersom man har prototyper (Das og Das, 2021, 984).

Siden utviklingen av prosjektet som jeg undersøker, er en innovativ prosess, og at prototyper spiller en viktig rolle i prosjektet, er det derfor relevant at jeg ser på teori om innovasjon og prototyper.

Sosiotekniske forestillinger

Sheila Jasanoff (2015) starter sin tekst om forestillinger om modernitet med å si at science fiction og forfatternes fantasi ligger tiår eller mer foran teknologisk innovasjon. Hun hevder videre at teknologisk innovasjon heller følger etter science fiction (Jasanoff, 2015, s. 1). Når det kommer til sosiotekniske forestillinger, sier hun at dette er «kollektivt holdt og utførte visjoner om ønskelige fremtider» (Jasanoff, 2015, s. 19) samt

at de er «formet av delte forståelser av former for sosialt liv og sosial orden som kan oppnås gjennom, og støtter fremskritt innen vitenskap og teknologi» (Jasanoff, 2015, s. 19). Sosiotekniske forestillinger kommer fra ulike grupper, eksempelvis sosiale bevegelser eller selskaper, men kan også komme fra mindre kollektiver eller enkeltpersoners visjoner. Disse sosiotekniske forestillingene kan vinne frem gjennom koalisjonsbygging eller utøvelse av makt. For at en visjon skal kunne kalles for en forestilling, må den først bli felles adoptert eller vedtatt (Jasanoff, 2015, s. 4). I tillegg til dette, sier hun at sosiotekniske forestillinger ikke bare kan omhandle hvordan man ønsker at fremtiden skal bli, men at visjonene også kan gjelde «*motstand mot det uønskede*» (Jasanoff, 2015, s. 19). Videre sier Jasanoff (2015) at det er flere faktorer som skiller sosiotekniske forestillinger fra blant annet ideer. Dette er at sosiotekniske forestillinger er slitesterke, de er kollektive og også kapable til å bli utført. I tillegg til å være alt dette, er sosiotekniske forestillinger også kulturelt bestemt og tidsmessig plassert. Sosiotekniske forestillinger er også et instrument for, og samtidig også et resultat av samproduksjonen som eksisterer mellom modernitetens samfunn, teknologi og vitenskap (Jasanoff, 2015, s. 19).

Jeg er ikke alene når det kommer til å bruke sosiotekniske forestillinger i studier av AI. Bareis og Katzenbach (2022) har blant annet brukt sosiotekniske forestillinger i sin artikkel «*Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics*» (Bareis og Katzenbach, 2022). Her fokuserer de på forestillingene og fortellingene som kommer frem «i nasjonale AI-strategidokumenter» (Bareis og Katzenbach, 2022, s. 860). Videre skriver de at forestillingene har en performativ funksjon fordi de reflekterer over sosiotekniske baner og tilbyr sosiotekniske baner. I tillegg til dette, samproduserer disse forestillingene «the installment» av disse fremtidene (Bareis og Katzenbach, 2022, s. 871). Sartori og Bocca (2022) har også brukt sosiotekniske forestillinger i sin artikkel «*Minding the gap(s): public perceptions of AI and socio-technical imaginaries*». Her skriver de blant annet om påvirkningskraften som befolkningens oppfatning av AI har på hvordan denne teknologien blir regulert, distribuert og utviklet. Videre hevder de at det er helt nødvendig å både gjøre rede for, og identifisere hvem det er som er involvert når det kommer til AI's formål, rolle og forestillinger for at man skal kunne problematisere samtalen rundt AI (Sartori og Bocca, 2022, s. 444).

Som man kan se, er det ofte knyttet visjoner til AI, noe som også er tilfellet i min studie. Siden jeg skal undersøke hvilke fremtidsvisjoner informantene har rundt AI i helsetjenesten om 10-20 år, er sosiotekniske forestillinger en relevant teori å benytte i denne masteroppgaven. Ved å bruke sosiotekniske forestillinger, kan jeg danne en bedre forståelse av hvorfor aktørene utformer teknologien som de gjør.

SCOT og fortolkningsmessig fleksibilitet

Bijker og Pinch (1987) skriver om det de kaller for Social Construction of Technology (SCOT), eller «sosialt konstruert teknologi». Ifølge Bijker og Pinch (1987), så er det ikke slik at de stadiene i en utviklingsprosess som blir vellykkede, er de eneste mulige stadiene. I SCOT-teorien, har man det Bijker og Pinch (1987) kaller for «fortolkningsmessig fleksibilitet». Med dette, mener de at ulike personer tolker og tenker på den samme teknologiske artefakten på ulike måter. Det er også fleksibilitet i hvordan en teknologisk artefakt blir utformet. Ifølge Bijker og Pinch (1987) kan en enkelt artefakt

bli tolket på helt ulike måter av ulike sosiale grupper (Bijker og Pinch, 1987, s. 21, 22, 33-36).

Siden jeg skal undersøke hvordan informantene ser på og tolker AI som en teknologisk artefakt, er det relevant å se på teori rundt SCOT og fortolkningsmessig fleksibilitet.

Tverrfaglighet

Teori om tverrfaglighet er relevant for min studie, da noe av det jeg ønsker å undersøke, er hvordan det tverrfaglige samarbeidet i dette prosjektet er organisert. I tillegg til dette kommer det også frem av intervjuene at hele utviklingsprosessen av AI-løsningen som jeg har sett på, baserer seg på et tverrfaglig samarbeid. Teorier om tverrfaglighet vil derfor hjelpe meg med å forstå måten aktørene jobber sammen på, på en bedre måte.

Knut Holtan Sørensen skriver at tverrfaglighet er et upresist begrep som har blitt gitt ulikt meningsinnhold (Sørensen, 2008, s. 10). Videre skriver han at man har tre forskjellige typer med tverrfaglighet. Disse er; a) tverrfaglighet i form av ulike fag som er sammenkoblet som fører til at forskjellene fagene seg imellom forsvinner og som gjør at måten å arbeide på blir enhetsvitenskapelig, b) samarbeid mellom ulike fag som låner både metoder, teorier og tilnærminger seg imellom og c) personer med ulike faglige bakgrunner samarbeider og bruker sin fagdisiplin, i tillegg til at sammensyningen av kunnskap må bruke perspektiver som ikke ligger innenfor fagdisiplinene til personene. Videre trekker han frem tre ting som tverrfaglighet handler om. Dette er: innstilling, som vil si hvordan man stiller seg til denne formen for arbeid – negativt eller positivt. Det andre er ferdigheter når det kommer til kommunikasjon, og det tredje er evnen man har til å kunne se helheten. Sørensen (1999) sier at begrepet tverrfaglighet oppstod fordi man var urolig for at det skulle oppstå mangel på tenking rundt helheten samt mangel når det kommer til å ha oversikt som følge av en økende grad av spesialisering. For lite innsikt i hvilke utfordringer tverrfaglig samarbeid bringer med seg, er en av grunnene til at mange har feilet når de har forsøkt å jobbe tverrfaglig. En slik utfordring, kan for eksempel være at man har forskjellig begrepsbruk (Sørensen, 1999, s. 116, 117, 121, 122).

Nå har vi sett litt på hva Sørensen (1999) skriver at tverrfaglighet er, men hva innebærer et tverrfaglig samarbeid? Ifølge Vestal og Mesmer-Magnus (2020), er tverrfaglige team, team som er satt sammen av «eksperter med unik ekspertise som er satt sammen for å generere nye forståelser av eksisterende problemer og for å utforske nye muligheter» (Vestal og Mesmer-Magnus, 2020, s. 739). I tillegg til dette, skriver Vestal og Mesmer-Magnus (2020) at tverrfaglige team er viktige når det kommer til både innovasjon og læring og snakker om det de kaller for «teamets relasjonelle ressurser». Dette dreier seg om den kunnskapen som teammedlemmene har opparbeidet seg gjennom å ha samarbeidet ved tidligere anledninger og hvor tidligere samarbeid har gjort at teammedlemmene har oversikt over hvilken ekspertise hver enkelt besitter og hvordan denne ekspertisen kan brukes. Teamets relasjonelle ressurser er også viktig når det kommer til et teams evne til å integrere medlemmenes ekspertise. Vestal og Mesmer-Magnus (2020) skriver også om det de kaller for «instrumental ekspertise», som handler om den ekspertisen som teamet har som kan brukes til innovasjon (Vestal og Mesmer-Magnus, 2020, s. 739, 743 og 744).

Kapittel 3 - Metode

AI er en teknologi som ofte kan fremstå litt mystisk og fremstilles ofte som nokså abstrakt. Det fremstår som en teknologi som det er litt vanskelig å bli klok på. Hvorfor er den som den er? Hvordan blir den til? For å kunne få en bedre forståelse av dette, intervjuet jeg seks informanter som har vært med på, og som holder på å utvikle AI-løsninger til helsetjenesten, og spurte dem om hvordan utviklingsprosessen bak en slik teknologi er. I denne delen av masteroppgaven, skal jeg forklare hvilken metode jeg brukte for å undersøke dette.

Kvalitativ metode og intervju

For å belyse de temaene som er nevnt over, altså hvordan utviklingsprosessen for denne AI-løsningen for diagnostikk i helsetjenesten var, valgte jeg å bruke kvalitativ metode hvor jeg gjennomførte dybdeintervjuer i form av semistrukturerte intervjuer med seks informanter. Jeg mener at denne metoden passer bra til mitt prosjekts formål, fordi man ifølge Tjora (2021) i dybdeintervjuer, ønsker å få informanten til å reflektere over både sine egne meninger, men også erfaringer når det kommer til det man ønsker å forske på (Tjora, 2021, s. 127). Dette var også derfor jeg bestemte meg for at intervjuene skulle være semistrukturerte. Jeg gjennomførte intervjuene på informantenes arbeidsplass når det lot seg gjøre. Når jeg skulle intervju informantene, hadde jeg fokus på å få intervjuet til å bare føles som en vanlig samtale mellom to personer slik at det ikke skulle føles så formelt. Håpet med dette, var at informantene skulle være komfortable med settingen slik at de turte å åpne seg mer i svarene de ga, og svare så ærlig som mulig. De gangene vi ikke hadde mulighet til å møtes fysisk, ble intervjuene gjennomført på Teams.

Når det kommer til intervjuene, så ble de som nevnt over, når det var mulig, gjennomført på informantenes arbeidsplass. Tre av intervjuene ble gjennomført via Teams. Jeg tok lydopptak av samtlige intervjuer som kunne hjelpe meg i transkriberingen. Når det kommer til etisk og sikker behandling av dataen som ble samlet inn, var dette i samsvar med anbefalingene og tillatelse fra SIKT (Referansenummer: 587753). I forkant av intervjuene, fikk alle informantene informasjons- og samtykkeskjema som de skrev under på. I tillegg til dette, forsikret jeg meg om at de samtykket til å være med i prosjektet før jeg startet hvert intervju. På denne måten fikk jeg sikret samtykke.

Valg av prosjekt og utvalg av informanter

For å finne et prosjekt der noen har utviklet en AI-løsning til helsetjenesten, søkte jeg på Forskningsinstitusjonen sine nettsider for å undersøke om de hadde noen AI-prosjekter som kunne være aktuelle å studere. Jeg fant to aktuelle prosjekter og sendte mail til prosjektenes kontaktpersoner og spurte om de ville være med på intervju. En av dem svarte ja til å bli med på dette. Før dette hadde jeg også sendt inn en søknad til SIKT og fått den godkjent (Referansenummer: 587753). I tillegg til dette, fylte jeg ut et informasjons- og samtykkeskjema.

Da jeg hadde funnet ut hvilket prosjekt jeg skulle studere for masteroppgaven, begynte jeg å tenke på hvem utvalget av informanter skulle bestå av. Siden jeg ønsket å studere hvordan utviklingsprosessen av AI-løsninger til helsetjenesten er, ønsket jeg å intervju

personer som hadde vært med på utviklingen av dette prosjektet, og som dermed hadde god innsikt i hvordan en slik utviklingsprosess foregår. Jeg gjorde derfor et strategisk utvalg av informanter. Dette vil, ifølge Tjora (2021), si at «(...) man velger informanter som av ulike grunner vil kunne uttale seg på en reflektert måte om det aktuelle temaet» (Tjora, 2021, s. 145). Etter at jeg hadde gjennomført det første intervjuet, spurte jeg informanten om hen visste om noen andre som jeg kunne intervju. Jeg brukte dermed også snøballmetoden. Dette gjorde jeg i de tre første intervjuene.

Etter at jeg hadde intervjuet tre informanter fra Forskningsinstitusjonen, kontaktet jeg klinikere som hadde vært med på utviklingsprosessen i dette prosjektet i tillegg, slik at jeg kunne få et utvalg som bestod av flere informanter enn kun de fra Forskningsinstitusjonen. På denne måten kunne jeg få flere ulike synspunkter og perspektiv på prosjektet jeg undersøkte. Her var en av informantene til spesiell hjelp, da hen tilbydde seg å presentere masteroppgaven min på et møte med blant annet flere klinikere. Etter hvert fikk jeg e-posten til flere klinikere som var interesserte i å la seg intervju. Fra intervjuene, fremstod det som at designere hadde spilt en viktig rolle i utviklingen av denne AI-løsningen. Derfor kontaktet jeg en av designerne som hadde vært med, og fikk intervjuet hen. Dermed hadde jeg intervjuet seks informanter, hvorav tre av dem var fra Forskningsinstitusjonen, to var klinikere og én var designer. Dermed fikk jeg flere ulike synspunkter på hvordan utviklingsprosessen var og hvordan samarbeidet dem imellom hadde fungert.

Oversikt over informantene og arbeidsplasser med kallenavn:

Informanter:	Arbeidsplass:
Forsker 1 (informant 1)	Forskningsinstitusjonen
Forsker 2 (informant 2)	Forskningsinstitusjonen
Forsker 3 (informant 3)	Forskningsinstitusjonen
Medisiner 1 (informant 4)	Sykehuset
Medisiner 2 (informant 5)	Sykehuset
Designeren (informant 6)	Designbyrået

Intervjuguide

Jeg laget tre versjoner av intervjuguiden. En til Forskerne, en til Medisinerne og en til Designeren. Det ble bare gjort noen justeringer mellom disse intervjuguidene for at de skulle være relevante for de ulike rollene som informantene hadde. Den opprinnelige intervjuguiden, som til slutt ble brukt til å intervju Forskerne, var delt inn i 7 bolker med ulike tema. Disse var følgende; «generelt og fagbakgrunn», «generelt om prosjektet», «idéprosess/arbeidsprosess», «samarbeid med andre», «beslutningsprosess», «forsvarlighetskravet», og «avslutningsspørsmål». Intervjuguiden til Medisinerne inneholdt de samme bolkene, men jeg la til en ny bolke som jeg kalte «praksis» hvor jeg spurte om hvordan behovet for denne løsningen oppstod, og hvordan de ser for seg at løsningen vil bli brukt i deres arbeid. Jeg la også til et nytt spørsmål under bolken om «samarbeid» som handlet om hvordan det er å samarbeide med andre yrkesgrupper, og hvilke styrker og utfordringer denne typen samarbeid medfører. Jeg omformulerte også spørsmålet om hvem de har samarbeidet med i dette prosjektet, og la til et spørsmål om hvem de pleier å samarbeide med i slike prosjekt. Videre la jeg til et oppfølgingsspørsmål under det første spørsmålet i bolken om «beslutningsprosess» om hva som pleier å være de viktigste eller største beslutningene som blir gjort i lignende typer prosjekter. Under bolken om «forsvarlighet», la jeg til et ekstra spørsmål om hvordan de jobber for at AI-

en eller algoritmene som blir brukt i eventuelle andre prosjekter, ikke fører til bias. Under «avslutningsspørsmål», la jeg også til et spørsmål om hva de synes er mest utfordrende med denne typen prosjekt generelt.

Intervjuguiden til Designeren bestod av 5 bolker, men jeg endret også litt på noen av spørsmålene her. Jeg tok bort spørsmålet hvor jeg spør om hvordan de ble interessert i AI og om de var interessert i AI før dette prosjektet. Jeg tok også bort bolken som handlet om prosjektet generelt og bolken om «samarbeid med andre». Under bolken «idéprosess/arbeidsprosess», la jeg til en del spørsmål. Disse handlet om hvordan møtene de deltok på sammen med klinikerne var og hva de gikk ut på, hvordan designprosessen var og hvordan de jobbet når de skulle designe løsningen, hvordan det var å jobbe med klinikerne og de fra Forskningsinstitusjonen når det gjaldt å designe løsningen, hva hen mener var hans viktigste oppgave i dette prosjektet, og hvordan hen så sin rolle mellom de ulike aktørene som var med. Under bolken om «beslutningsprosess», beholdt jeg spørsmålene, men endret litt på spørsmålet om hva som var de viktigste valgene som ble gjort underveis og knyttet dette opp mot utformingen og designet av løsningen.

Informert samtykke, konfidensialitet og konsekvenser av å delta i prosjektet

Thagaard (2018) nevner tre grunnprinsipper når det kommer til det som kalles for forskningsetiske retningslinjer. Disse er; informert samtykke, konsekvenser av å delta i forskningsprosjekter og konfidensialitet (Thagaard, 2018, s. 22). De som takket ja til å bli med, ble tilsendt et informasjons- og samtykkeskjema med mer utdypende informasjon om masterprosjektet samt informasjon rundt personvern og deres rettigheter. Når det kommer til konfidensialitet, fjernet jeg informantens navn og ga dem kallenavn, for eksempel «Forsker 1». Jeg anonymiserte også informantens kjønn og omtaler alle som «hen». I tillegg til dette, sier jeg ingenting om hvilken stilling de har og heller ikke hvor de jobber. Informantens arbeidsplass har også fått kallenavn, eksempelvis «Forskningsinstitusjonen». Det siste grunnprinsippet handler som sagt om konsekvenser av å delta i forskningsprosjekter. Her anonymiserte jeg informantene slik at ingen skulle kunne kjenne de igjen, jeg lagret lydopptakene på Nettskjema slik at de var lagret på en trygg plass, i tillegg til at jeg heller ikke sier navnet på prosjektet de jobbet med ei heller går i detalj rundt hva AI-løsningen skal gjøre.

Koding

Etter at jeg hadde samlet inn data fra intervjuene, begynte arbeidet med kodingen av datamaterialet. Jeg lot meg inspirere av tilnærmingen «grounded theory» som går ut på at man utvikler teori basert på empirien man har samlet inn (Thagaard, 2018, s. 18). Det første jeg gjorde, var at jeg delte inn i tre analysekapitler. Disse kapitlene ble skissert ut på bakgrunn av intervjuguiden jeg hadde laget. Kapitlene handlet om fremtidsvisjoner og fortolkninger, arbeidsprosessen og forsvarlighet. Etter dette, begynte jeg å kode datamaterialet. Jeg gikk gjennom de transkriberte intervjuene og markerte det jeg synes var mest interessant og tematikker som gikk igjen blant informantene. De temaene som gikk igjen, ble plassert inn i en tabell (en tabell for hvert kapittel), og delt inn i kategorier. Kategoriene er tett på intervjuguiden og de interessene jeg hadde da jeg startet opp dette masterprosjektet. Jeg jobbet med datamaterialet mitt rundt disse

tabellene med kategoriene. Jeg limte inn sitater fra alle informantene inn i tabellene med kategoriene, og tok med de mest interessante sitatene i analysekapitlene.

Reliabilitet, validitet og overførbarhet

Ifølge Thagaard (2018) er validitet, reliabilitet og overførbarhet sentralt når det kommer til å skulle si noe om både kvalitet og troverdighet i et forskningsprosjekt (Thagaard, 2018, s. 181). For mitt prosjekt, har jeg intervjuet seks informanter som alle har jobbet på det prosjektet jeg har undersøkt. Selv om seks informanter ikke nødvendigvis er et veldig stort utvalg, har jeg ved å ha intervjuet forskere, medisinerer og en designer dekt de ulike aktørgruppene som har vært med på å jobbe på dette prosjektet. Jeg har dermed snakket med de som har jobbet med det som jeg ønsket å undersøke for dette masterprosjektet. På den måten har jeg fått flere perspektiv på utviklingsprosessen og samarbeidet mellom de ulike aktørene, noe som øker påliteligheten til denne masteroppgaven.

Man kan si at et intervju kanskje blir mer en slags fortelling fra informantenes side, og da kan man stille seg spørsmålet om det informantene sa skjedde, faktisk skjedde, eller om det skjedde på den måten som blir fortalt av informanten. Det jeg ser, etter å ha intervjuet de ulike informantene, og fått satt de ulike fortellingene opp mot hverandre, er at informantenes fortellinger stemmer godt overens med hverandre. Dette gir en indikasjon på at det de har fortalt er sant, noe som styrker deres troverdighet, og dermed også troverdigheten til denne masteroppgaven.

Jeg mener at intervju var den beste måten for å skaffe kunnskap om hvordan det jobbes med å utvikle en AI-løsning for diagnostisering i helsetjenesten, da denne metoden har gjort det mulig for meg å få detaljerte beskrivelser av de ulike prosessene som inngår i et slikt prosjekt. Dette var nødvendig for å få svar på problemstillingene jeg har utformet.

Jeg ba om å få tilgang til prosjektsøknaden som Forskningsinstitusjonen sendte inn, samt om å få være med på et møte som aktørene pleier å ha med hverandre. Dessverre var ikke det mulig på grunn av konfidensialitet. Siden informantene er anonymiserte, og dette var noe de fikk beskjed om på forhånd, har jeg heller ingen grunn til å tro at de hadde noen agenda. Dette, i tillegg til at informantene var samstemte i det de snakket om, gjør at det de fortalte er troverdig.

Når det gjelder prosjektets overførbarhet, er det vanskelig å si om forståelsen jeg har utviklet her er overførbar til andre institusjoners arbeidsmåte når det kommer til utvikling av AI-løsninger til helsetjenesten. Jeg mener derimot at forståelsen som jeg har utviklet i dette prosjektet kan gjelde for andre AI-løsninger som blir utviklet for helsetjenesten av Forskningsinstitusjonen som jeg har fokusert på her. Jeg har studert et prosjekt som er særegent, og kan dermed ikke forvente å finne de samme funnene som jeg fant her, hos andre aktører som arbeider med å utvikle AI-løsninger til helsetjenesten. Jeg mener fortsatt at min studie kan gi gode perspektiver for å sammenligne annen type forskning på dette. Jeg studerer mer det særegne prosjektet enn AI generelt. I tillegg mener jeg at min studie kan gi innsikt i hvordan en slik utviklingsprosess kan gjøres med tanke på tverrfaglighet.

Kapittel 4 - Fremtidsvisjoner og fortolkninger av AI

Mange har tanker om AI og hva AI kan være i fremtiden. Noen ser for seg at AI nærmest skal redde verden og gjøre alle sine liv mye bedre, mens andre frykter at AI skal bety slutten for menneskeheten. Disse sprikende fremtidsvisjonene om hva AI kommer til å være i fremtiden, kan man også se i eksempelvis science fiction-filmer. Noen filmer presenterer en fremtid med storslåtte AI-løsninger som har revolusjonert verden til det bedre, mens andre filmer, som *The Matrix*, viser oss en fremtid som har blitt verre på grunn av at AI-teknologien har blitt for avansert for oss mennesker og blitt så smart at den har fått sin egen bevissthet. De ulike måtene ulike personer ser for seg at AI kommer til å være i fremtiden, tyder også på at ulike personer tolker AI på forskjellige måter.

Som jeg viste frem i introduksjonen til masteroppgaven, er AI i helsetjenesten svært aktuelt og mange ser for seg at AI kommer til å være en del av helsetjenesten i fremtiden. Derfor kan det være interessant å se på hvordan de som jobber med å utvikle AI-løsninger til helsetjenesten, samt de som jobber i norsk helsetjeneste, ser på og tolker AI og hvordan de ser for seg AI i den norske helsetjenesten i fremtiden. Disse personene kan jo være med å påvirke hvordan AI i helsetjenesten kommer til å være i fremtiden. I tillegg, er det også interessant å se på de ulike tolkningene disse aktørene har på AI.

Vi skal derfor se på hva informantene tror at AI i helse er om 10-20 år. I tillegg til dette, ser kapittelet på hvordan de ulike aktørene tolker AI som en teknologisk artefakt.

Rettferdig støtteverktøy

Støtteverktøy er et begrep som går igjen på en eller annen måte blant samtlige informanter når de snakker om hva de tror AI i helse er om 10-20 år. Medisiner 1 ser for seg at AI kommer til å fungere som et hjelpemiddel for klinikerne, og sier at «jeg tror det vil være AI-hjelpemidler til nærmest alt», og trekker spesielt frem at AI kan hjelpe klinikere når det kommer til beslutningsstøtte. Medisiner 2 trekker også frem beslutningsstøtte som et område der klinikerne kan få hjelp av AI-teknologi, men tror ikke at AI vil hjelpe i like stor grad som det Medisiner 1 ser for seg. Medisiner 2 sier at hen tror at AI kan hjelpe litt til når det kommer til «kontor- og sekretærhjelp, bestilling og skrive notater». Forsker 1 sier at hen tror at AI vil være et «veldig godt støtteverktøy, en støtteteknologi på ekstremt mange områder». Det at informantene ser for seg at AI kommer til å være et støtteverktøy, virker å ikke være så urealistisk da Yin, Ngiam og Teo (2021) i sin litteraturgjennomgang, fant ut at AI som er blitt implementert i helsetjenester, som oftest fungerer som nettopp et støtteverktøy (Yin, Ngiam og Teo, 2021, s. 9). Forsker 1 kommer med et eksempel for å forklare det hen mener med AI som støtteverktøy:

(...) Det er en AI-algoritme som skal skille mellom (...) hjerneblødning og slag, om du har propp (...) på CT. Da er hele poenget det at når det kommer inn en pasient akutt på en, la oss si en lørdag, natt til søndag, da sitter ikke den beste ekspertisen blant radiologer på jobb. Da er det vakthavende som kanskje er en person i opplæring, en person som er ny i faget og så videre. Sånn at, hvis du skal ha lik analyse av bildene uansett når en pasient kommer inn, så kan du gjerne ha en AI-algoritme som gjør den første analysen. Og det som skjedde når vi presenterte den her ideen for et par av dem på radiologisk avdeling, det var at de sa at «ja men, det her er jo en veldig lett sak for oss (...) det er ikke

noe problem, vi gjør fort det når vi ser disse bildene. Og det var da vi hadde det fine svaret til dem «ja, men det er nettopp derfor du skal slippe å gjøre det. Fordi at din ekspertise bør brukes på noe mer fornuftig enn å ta de veldig lette tingene og lette avgjørelser. Det kan en AI avhjelpe med». (...) Men en AI kan hjelpe deg til at avgjørelsen blir mer robust, mer person-uavhengig og at du får gjort den raskere.

Her virker det som at AI som støtteverktøy for Forsker 1, vil si at AI skal gjøre de lette oppgavene slik at klinikerne kan bruke tiden sin på noe mer «fornuftig». Det kan dermed virke som at hen legger vekt på at AI i helsetjenesten kan hjelpe til med å optimalisere måten klinikerne bruker tiden sin på, med at de kan gjøre andre mer krevende oppgaver, og dermed ikke trenger å bruke tiden sin på enkle oppgaver. Helt til slutt i sitatet, kan vi se at Forsker 1 sier at «AI kan hjelpe deg til at avgjørelsen blir mer robust, mer person-uavhengig og at du får gjort den raskere». Senere i intervjuet forklarer hen hva hen mener med det, og sier at det varierer hvilke kirurger som er i rommet, og at AI vil gjøre at man blir «mindre avhengig av hvilken person som er i rommet til enhver tid». Hen sier at det kan være en «veldig sterk kirurg» med «en veldig god «standing»», som kan gjøre at det blir operasjon, selv om det ikke er det beste valget. Dette er noe hen sier at de vil unngå.

Her kan man se at hen sier at AI vil kunne føre til at hvilke beslutninger som blir tatt, ikke vil variere ut ifra hvilke klinikere som «er i rommet» i hvert tilfelle. Det er AI-en som tar beslutningen, og den vil ikke påvirkes av hvilken «standing» noen av kirurgene har. Dette er et godt poeng, men hvor godt det fungerer i praksis, avhenger selvfølgelig av hvor god og pålitelig AI-en er. En AI vil ikke nødvendigvis komme til en mer riktig beslutning enn det klinikerne gjør, bare fordi den ikke vil påvirkes av hvem som er i rommet til enhver tid. AI-en vil fortsatt være påvirket av den dataen den har blitt trent på, og det er ikke sikkert at den dataen er av god kvalitet. AI er ikke en ufeilbar teknologi. Det vil være flere faktorer som spiller inn om en AI kommer frem til riktig beslutning eller ikke, for eksempel om den er trent på nok og riktige data.

Dette med objektivitet, er noe som Forsker 1 kommer inn på flere ganger når hen snakker om den AI-løsningen som jeg har studert her. Hen sier at klinikerne «mangler egentlig et objektivt verktøy». Det fremstår som at hen ser på AI som en objektiv teknologi, eller at det kan bli en objektiv teknologi i fremtiden. Dette kan tyde på at hen har veldig stor tro på denne typen teknologi. Det at hen jobber med å utvikle AI-løsninger, kan være en faktor som gjør at det fremstår som at hen har stor tro på denne teknologien. Det er jo en stor kontrovers rundt AI-algoritmer og bias, og det har vært flere tilfeller der AI-algoritmer har forskjellsbehandlet personer av ulike etnisiteter eller andre bakgrunner.

Det virker dermed som at Forsker 1 ser på AI som en teknologi som kan hjelpe til med å optimalisere arbeidsflyten til klinikerne ved at den kan gjøre de litt enklere arbeidsoppgavene. Dette er også noe som Forsker 2 trekker frem. Hen sier følgende:

AI-en kan hjelpe med letejobben og så på en måte bare lyse opp «her skal du kikke nedi, her er det noe». Så den, på en måte, vil være den der helperen som på en måte tar unna litt av, kanskje det litt kjedelige eller ensformige jobben da. Hvis du sitter der og kikker på bilder i timevis, så blir du kanskje litt sånn ... kanskje du mister fokus litt. Da er det litt lettere hvis AI-en har gjort mye av det og du bare [får det] presentert.

Her kan vi se at Forsker 2 ser for seg AI som en teknologi som kan gjøre den «kjedelige» jobben noe som dermed kan hjelpe til med å forbedre arbeidsflyten. I tillegg er hen også innom et annet tema; pasientsikkerhet. Hen sier at det er lett å miste fokus dersom man bruker flere timer av dagen til å se på bilder, men at dette er noe som AI kan gjøre i stedet. Dermed virker det som at Forsker 2 ser for seg, og tolker AI som en teknologi som kan øke pasientsikkerheten. Forsker 3 trekker også frem analysing av bilder som noe som AI kan hjelpe klinikerne med, og nevner i den anledning automatisk analyse både når det kommer til bilder, men også diagnostikk. Det virker dermed som at Forsker 1, 2 og 3 med «støtteverktøy», først og fremst tenker på analysing av bilder og AI som en teknologi som kan ta over de enkle og litt kjedelige oppgavene. AI blir dermed en teknologi som kan skape bedre flyt i arbeidshverdagen. At AI skal kunne føre til bedre arbeidsflyt i klinikerens hverdag, er ikke utenkelig. Dette kan vi se av litteraturstudien som Yin, Ngiam og Teo (2021) gjennomførte, som viste at seks av syv studier de fant, viste at AI forbedret arbeidsflyten til klinikere (Yin, Ngiam og Teo, 2021, s. 10). Designeren er også innom et lignende tema som arbeidsflyt, men snakker om AI som en teknologi som kan avlaste et helsesystem som har nådd sin grense. Hen sier at dersom helsesystemet skal være slik som det er i dag, også i fremtiden, «da trenger man AI». Her snakker hen om «forebyggende helse og AI» og mener at AI vil ha en stor rolle i å få til dette. Dermed virker det som at Designeren, på samme måte som Forsker 1, 2 og 3, ser for seg at AI vil kunne gjøre enkelte oppgaver for helsepersonell, slik at deres hverdag blir mer overkommelig. I tillegg til dette, trekker Designeren frem fastlegekrisen og at AI vil være til stor hjelp der, ikke bare for å få dagen til å gå rundt for fastlegene, men også for de som ikke har fått fastlege på grunn av en kapasitet som ikke lenger strekker til.

Forsker 1 trekker også frem «lik analyse av bildene uansett når en pasient kommer inn» og at man kan bruke AI for å få til dette. Dermed kan det også tenkes at hen med begrepene «støtteteknologi» og «støtteverktøy», også mener at AI kan hjelpe klinikerne med å gi lik behandling til pasientene uavhengig av hvem som er på jobb. Dette er noe som Forsker 3 også trekker frem:

... i forhold til opptak av data og ultralyd da, så går det nok mot at ... der snakket jeg jo om at det har mye å si på operatøren i forhold til AI-verktøy og analyse av bildene, (...) i forhold til robotikk og AI da, så går det nok mot standardiserte ultralydopptak i mange tilfeller da der det er roboter med AI som tar opp dataen, ikke hånda til legen.

Det virker dermed som at lik behandling er noe som Forsker 1 og 3 tenker på når de tenker på AI som et støtteverktøy. Kanskje de ser for seg, eller tolker AI som en slags type «rettferdighetsteknologi». Når Forsker 3 sier at det går mot at det er roboter med AI som skal ta opp data, og ikke leger, så virker det som at hen mener at dette er et område der AI kommer til å erstatte klinikerne. Det virker derfor som at Forsker 3 er ganske sikker på at AI kommer til å komme inn i helsetjenesten og ta over denne typen arbeidsoppgaver.

Forsker 1 trekker også frem at AI ikke bare vil kunne fungere som et støtteverktøy for klinikerne ved at den kan gjøre enkle oppgaver for klinikerne, men også at den kan hjelpe klinikerne med å oppnå bedre resultater i jobben sin:

... for å si det sånn, jeg tror alle klinikere som tar i bruk AI, de vil erstatte fremtidens klinikere som ikke tar i bruk AI. Det er måten å se det på. Fordi AI-løsningene kommer, og de som ikke vil ta dem i bruk,

de blir erstattet av dem som tar det i bruk – de kommer til å få mye bedre resultater, de klinikerne som tar i bruk AI for å hjelpe seg.

Av sitatet over, kan man se at Forsker 1 ser for seg at AI vil kunne være såpass mye til hjelp og kunne forbedre resultatene til klinikerne såpass mye, at klinikere som ikke bruker AI ikke vil kunne oppnå i nærheten av like gode resultater. Dermed virker det som at Forsker 1 ser på AI som en såpass solid og god teknologi, at den sammen med klinikere, vil utklasse klinikere uten AI. Det virker også som at Forsker 1, mener at det er uunngåelig at AI kommer til å komme inn i helsetjenesten og at man enten må tilpasse seg, eller finne noe annet å gjøre. Det fremstår dermed som at hen har et teknologideterministisk syn på AI og at det kommer til å bli en del av helsetjenesten og at det bare er noe man må forberede seg på. Det minner litt om det som Skjølsvold (2017) sier om teknologideterminisme; at utviklingen av teknologi vil fortsette, uten at vi kan gjøre noe med det (Skjølsvold, 2017, s. 21). Man må tilpasse seg teknologien, hvis ikke så kommer man til å miste jobben til de som bruker AI.

Det vi kan se at det som har blitt sagt så langt, er at informantenes fremtidsvisjoner om AI i helse, på et overordnet nivå, er det samme; et støtteverktøy. Forestillingen om at AI kommer til å fungere som et støtteverktøy, er dermed en forestilling som er «kollektivt holdt» og, som vi skal se i neste analysekapittel, også «kollektivt utført» som Jasanoff (2015, s. 19) sier. Siden forestillingen om at AI i helsetjenesten skal fungere som et støtteverktøy er noe som går igjen blant samtlige informanter, virker det som at dette er et syn på AI som er vedtatt eller adoptert i fellesskap blant informantene, og kan dermed kalles for en forestilling (Jasanoff, 2015, s. 4). Om dette er noe som har blitt adoptert eller vedtatt bevisst eller ikke, er vanskelig å si, men i og med at de jobber med å utvikle denne AI-løsningen sammen som skal fungere som et støtteverktøy, tyder det på at dette er noe de i fellesskap er klar over. Informantenes visjoner om at AI skal fungere som et støtteverktøy i fremtidens helsetjeneste, har også et kjennetegn som Jasanoff (2015) mener at sosiotekniske forestillinger har. Dette er blant annet at de er kollektive (Jasanoff, 2015, s. 19), noe visjonen om AI som støtteverktøy er her, da det er en felles visjon blant flere informanter fra flere ulike yrker. Det som er interessant, er at selv om de har de samme forestillingene om AI som et støtteverktøy, så har de litt ulike syn når det kommer til hvilke områder det skal være et støtteverktøy på. Noen sier at AI kommer til å være et støtteverktøy på mange ulike områder, mens andre nevner mer administrative funksjoner, som kontorarbeid og bestille ting. De er dermed samkjørte når det kommer til fremtidsvisjoner knyttet til AI på et overordnet nivå, men ikke i like stor grad når det gjelder på detaljnivå.

Når det kommer til hvordan de ulike informantene tolker AI som teknologi, virker det som de tolker AI som et verktøy – et verktøy som skal hjelpe dem. Videre virker det som at samtlige informanter, bortsett fra Medisiner 2 (som jeg kommer mer inn på senere i kapitlet), tolker AI som en veldig bra teknologi med stort potensial. Det virker også som at de tolker AI som en slags «revolusjonerende» teknologi som kommer til å endre den norske helsetjenesten til det bedre.

Erstatning

Vi har så langt sett at «støtteverktøy» er et ord som går igjen blant informantene. Et annet ord som også går igjen blant informantene når de forteller om hva de tror AI vil være i helsetjenesten om 10-20 år, er «erstatning». Selv om fem av seks informanter

sier at AI kommer til å spille en viktig rolle i helse i fremtiden, nevner de fleste av dem at de ikke tror at AI kommer til å erstatte klinikerne. Forsker 1 sier at AI ikke kommer til å ta over radiologenes jobber. Forsker 2 sier at hen tror at AI kan komme med ideer som klinikerne verifiserer, men at «det vil bestandig være klinikere der». Medisiner 1 tror heller ikke at AI kommer til å erstatte klinikerne, heller at det kommer til å hjelpe dem når det kommer til beslutningsstøtte. Hen sier at «jeg tror ikke sykehuset driftes uten personell for å si det sånn». Medisiner 2 sier at «du kan jo si at AI kan jo gjøre en doktors jobb på et vis», men at man fortsatt kommer til å ha behov for både sykepleiere og doktorer. Heller ikke Designeren ser for seg at klinikerne kommer til å bli erstattet av AI, og ser heller på AI som et verktøy.

Som man kan se her, så virker det ikke som at det er noen reell frykt blant informantene om at klinikerne kommer til å bli erstattet av AI om 10-20 år. De tolker heller på AI som et positivt hjelpemiddel, fremfor en skummel teknologi som skal ta over jobbene til klinikerne. De virker å fokusere mer på mulighetene som AI kan gi for fremtidens helsetjeneste. På en annen side, er det interessant at selv om disse informantene sier at de ikke tror at AI kommer til å erstatte klinikerne om 10-20 år, så legger de vekt på å si at AI ikke kommer til å ta over. Hva kan være grunnen til det? Kanskje det bare er naturlig å nevne i denne sammenhengen da temaet om at AI kommer til å ta over jobbene våre er ganske aktuelt og ofte blir tatt opp, eller kanskje det ligger en liten usikkerhet i underbevisstheten deres om at AI kanskje kan ta over likevel? Prøver de å overbevise seg selv? Dette blir bare spekulasjoner selvfølgelig, men det er likefullt et interessant moment å påpeke.

Nå har vi akkurat sett at informantene ikke tror at AI kommer til å erstatte klinikerne i fremtidens helsetjeneste. Det som er interessant å merke seg, er at selv om informantene sier at de ikke tror at AI kommer til å erstatte klinikerne, så nevner noen av dem fortsatt noen oppgaver som de tror AI kommer til å ta over, og dermed erstatte klinikerne i. Forsker 3 sier blant annet at det går mot «standardiserte ultralydopptak i mange tilfeller der det er roboter med AI som tar opp dataen, ikke hånda til legen». Medisiner 2 sier at AI kan hjelpe dem med «enkel kontor- og sekretærhjelp, bestille ting og skrive notater og sånt». Når Forsker 2 snakker om AI sin rolle i helsetjenesten enda lenger frem i tid enn 10-20 år, så sier hen «og sånn sett kan du jo si at, ja kanskje AI-en kjører hele prosedyren». Det virker derfor som at når informantene bruker ordet «erstatte», så tenker de på erstatning i form av at AI tar over alle arbeidsoppgaver, ikke bare noen få. Dette er interessant og sier oss noe om hvordan de tolker begrepet «erstatte» i denne sammenhengen. Kanskje de heller ser på AI som en teknologi som vil føre til delvis erstatning, men ikke som en fullkommen erstatning?

Trening og opplæring

Et annet interessant tema som ble tatt opp, er bruken av AI til trening og opplæring. Dette er noe Forsker 2 snakker om. Hen sier at man kan få mange bilder og at AI kan hjelpe der. I tillegg til dette, sier hen at AI-en kan fungere som en slags støtte under opplæring, men også når man «begynner på pasienter». Videre sier hen at dersom man har «tracking på instrumentene, så kan den hjelpe deg å finne ut «ok, hvor skal vi gå inn hen?»». Hen gir et eksempel på hvordan AI kan fungere i trening og opplæring:

(...) Så da kan AI-en hjelpe en som ikke er så erfaren eller (...) hvordan få det best mulige bildet? (...). AI-en viser deg «ok, (...) hvis du flytter ... vrir proben litte grann, flytte den litt, så får du et bilde som er

mye bedre». Og så da kan AI-en da også velge å tolke det bildet, da, for at hvis det er et bilde som er vanskelig å tolke, så kan AI-en være med å tolke det.

Her kan det se ut som at Forsker 2 ser for seg at AI kan fungere som en slags instruktør som kan veilede uerfarne klinikere gjennom en prosedyre hvor de eksempelvis skal bruke en probe, eller når det kommer til å analysere bilder. Det er interessant at hen ikke bare ser på hva AI kan gjøre for klinikere i deres arbeid, men også mulighetene AI kan ha for å trene opp klinikere. Det at hen snakker om AI som kan hjelpe til i opplæring av klinikere, gjør at det fremstår som at hen stoler på AI som teknologi, eller ser for seg at AI som teknologi, kommer til å bli såpass trygg og pålitelig at det er trygt å bruke i opplæring av klinikere, slik at klinikerne ikke får dårlig eller feil opplæring.

Ingen endring

Som vi har sett av det som informantene har sagt om hva de tror AI i helsetjenesten er om 10-20 år, så kan vi se at de fleste er enige om at AI kommer til å være et viktig støtteverktøy for klinikerne. Det er derfor veldig interessant at én av informantene, Medisiner 2, ikke deler det samme synet på AI i fremtiden. Hen sier nemlig at AI vil spille en ganske liten rolle i helsetjenesten om 10-20 år og kaller seg selv for «AI-pessimist». Mens de andre informantene peker på flere områder og måter AI kan hjelpe klinikerne på, mener Medisiner 2 at AI ikke «kommer til å tilføre så veldig mye i hverdagen». Hen nevner at AI kanskje kan spille en «viss rolle i radiologi» og at man kan «få litt hjelp til sånn enkelt kontor-, sekretærhjelp, bestille ting og skrive notater og sånt». På spørsmål om hvorfor hen ikke tror at AI kommer til å spille en så stor rolle, peker Medisiner 2 på viktigheten den menneskelige faktoren har i helsetjenesten:

(...) det er jo veldig manuelt arbeid. Noen må jo ha alle opplysningene ut fra pasienten, ikke sant. Pasienten må jo fortelle hva som feiler han, og jeg er ikke sikker på om han er så interessert i å snakke i en mikrofon til en robot. Det kan tenkes, men jeg tror nok kanskje likevel at de fleste har lyst til å se en doktor. Den menneskelige faktoren er fortsatt viktig i helsevesenet. Og så selve pleien, og det å gi medisiner, og det å undersøke pasientene, kjenne på pasientene. Det kan jo gjøres med en robot, men det er jeg ikke sikker på om de er interesserte i. Og så er 10 år veldig kort tid i et sånt medisinsk-teknologisk-perspektiv, altså.

Av sitatet over, kan vi som sagt se at viktigheten av den menneskelige faktoren er en av grunnene som Medisiner 2 oppgir når hen forklarer hvorfor hen ikke tror AI kommer til å spille noen vesentlig rolle i helsetjenesten. Hen ser dermed ikke på AI som en teknologi som skal «frelse» helsetjenesten. Dette står i motsetning til det de andre informantene tenker om AI, hvor de ser for seg at det er en teknologi som virkelig vil kunne hjelpe klinikerne i fremtiden. Det virker som at hen tolker AI som en teknologi som ikke er kapabel til å erstatte klinikerne. Videre trekker hen frem en potensiell skepsis blant pasientene når det kommer til AI. Det virker som at hen mener at de fleste vil foretrekke den menneskelige faktoren fremfor teknologien. Dette er interessant, for mens de andre informantene virker å ha fokusert mer på hvor bra AI kommer til å være for klinikerne, tenker Medisiner 2 mer på hvordan pasientene vil oppleve AI. En annen faktor hen trekker frem, er tidsaspektet. Hen mener at 10 år innen medisin er «veldig kort tid». Det virker dermed som at Medisiner 2 har et annet perspektiv, eller en annen tolkning av tid innenfor helsetjenesten enn resten av informantene. Mens svarene til de andre informantene tyder på at de implisitt mener at 10 år er mer enn nok tid for AI til å utvikle seg til gode støtteverktøy, mener Medisiner 2 at dette ikke er tilstrekkelig med

tid. Dette er noe hen trekker frem flere ganger, blant annet når hen sier «Så 10 år går veldig fort innenfor medisin». Hen trekker også frem sin egen erfaring med AI i helsetjenesten og den utviklingen hen har sett de siste 10 årene. Hen sier at «AI begynte jo å piple inn som en sånn tanke hos oss for 10 år siden, og det har ikke skjedd noe». Videre sier hen at det eneste som har skjedd på den tiden, er at de har fått «litt bedre bilder». Som man kan se her, så mener Medisiner 2 at 10 år er kort tid innenfor medisin og at det derfor ikke vil være nok tid for at noen radikale endringer skal kunne finne sted. Det kan tenkes at hens egne erfaringer derfor kan ha påvirket hens tolkning av AI som teknologi, samt formet hens visjoner og litt mer pessimistiske syn på AI sin rolle i fremtidens helsetjeneste. Hen tolker eller har kanskje et annet perspektiv på tid og teknologiutvikling enn de andre informantene. Medisiner 2 snakker også om folkesykdommene og hvilken rolle AI kommer til å ha der:

... Sånn at ting som virkelig berger liv, det tror jeg ikke kommer til å bli noen radikal endring på i de nærmeste 10 årene. Det er klart, en liten operasjon kan gå bedre på akkurat den pasienten, men det er ikke folkesykdommene, ikke sant. De store tingene som malaria, der er det ikke AI som er viktig, ikke sant, annet enn hvis det kan hjelpe til å lage en vaksine litt fortere. Det er vaksinen som er viktig, ikke sant, og AI blir et lite hjelpemiddel.

Som man kan se her, virker det som at Medisiner 2 mener at AI kan hjelpe til i enkeltsituasjoner, men ikke i de mer generelle situasjonene, eksempelvis når det kommer til å få bukt med folkesykdommene. Hen tror heller at det er andre faktorer som kommer til å spille en større rolle enn AI der. Også her virker Medisiner 2 å ha et litt annet perspektiv enn de andre informantene. De andre informantene har vært mer innovert AI som støtteverktøy og hjelpemiddel for klinikere, mens Medisiner 2 tenker på større ting som folkesykdommer. Hen virker dermed ikke å være så veldig overbevist over AI-teknologi. I motsetning til de andre informantene, som virker å tolke AI som en veldig bra og lovende teknologi, virker det som at Medisiner 2 tolker AI på en helt annen måte. Det fremstår som at hen ikke tolker AI som noen spesiell eller «revolusjonerende» teknologi i så måte, det er «bare» en annen form for teknologi, eller «et lite hjelpemiddel», eller kanskje en litt kald teknologi som mangler varmen fra menneskelig interaksjon. Det virker som at hen ser på AI som en teknologi som det er knyttet en overdrevet optimisme til. Dermed får vi et eksempel på det Bijker og Pinch (1987) sier om fortolkningsmessig fleksibilitet og at ulike personer tolker og tenker på den samme teknologiske artefakten, i dette tilfellet AI, på ulike måter (Bijker og Pinch, 1987, s. 34).

Oppsummering

I dette kapitlet, så jeg på informantenes fremtidsvisjoner når det kommer til AI i helsetjenesten om 10-20 år, i tillegg til hvilke fortolkninger de har av AI som teknologi. Når det kommer til hvilke visjoner informantene har for AI i fremtidens helsetjeneste, var begrepet «støtteverktøy» et begrep som gikk igjen. Informantene så for seg at AI skulle kunne hjelpe klinikerne med ulike arbeidsoppgaver som blant annet beslutningsstøtte, kontor- og sekretærhjelp, og de litt «lette» arbeidsoppgavene. Flere av informantene virket å først og fremst tenke på analysing av bilder når de snakket om AI som en støtteteknologi. På denne måten så de for seg at AI kan hjelpe til med å optimalisere klinikernes tidsbruk ved at de skal slippe å gjøre de enkle rutineoppgavene. Noen av informantene så for seg AI som en «rettferdighetsteknologi» ved at de mente at AI kan gjøre at man kan få mer lik behandling fordi behandlingen ikke lenger vil være avhengig av hvilke personer som er i rommet på et gitt tidspunkt. Informantenes fremtidsvisjoner

rundt AI i helsetjenesten var de samme på et overordnet nivå, altså at det skal være et støtteverktøy, men de har litt ulike meninger når det kommer til akkurat hvilke områder den skal fungere som støtte på.

Videre, kunne vi se at det ikke virket som at det eksisterte noen reell frykt blant informantene om at AI skal erstatte klinikerne om 10-20 år. Informantene tolker ikke AI som en skummel eller truende teknologi, men fokuserer mer på mulighetene en slik teknologi kan gi. Likevel kan det hende at det eksisterer en liten usikkerhet blant informantene siden de legger vekt på å fortelle at AI ikke kommer til å ta over. Selv om informantene ikke så for seg at klinikerne kom til å bli erstattet av AI, trakk de frem noen oppgaver der de så for seg at AI kom til å erstatte klinikerne. Dermed virket det som at de med «erstatte» tenker på erstatning i form av at AI skal ta over alle oppgavene og ikke enkelt-oppgaver her og der.

En av informantene så for seg at AI i fremtiden kan få en rolle å spille når det kommer til trening og opplæring av klinikere, nesten i form av en slags instruktør. Det virker derfor som at hen stoler veldig på AI som teknologi, eller ser for seg at den i fremtiden kommer til å være pålitelig og trygg nok til at den kan brukes til å lære opp klinikere.

En av informantene var ikke like optimistisk som resten når det kom til hvilken rolle hen så for seg at AI kom til å ha i fremtidens helsetjeneste. Mens de andre informantene så for seg at AI kom til å være et viktig støtteverktøy i fremtiden, mente Medisiner 2 at AI ikke kommer til å spille en veldig stor rolle. Hen sa at AI kan hjelpe til med noen enkle oppgaver, men tror ikke at den vil ha så mye innvirkning i hverdagen. Hen tror ikke at AI kan erstatte viktigheten av den menneskelige faktoren i helsetjenesten, og tolker ikke AI som en «frelsende» teknologi. Videre virket Medisiner 2 å ha et annet perspektiv på tid enn de andre informantene, og mente at 10 år ikke er nok tid for at AI skal kunne utvikle seg til en teknologi som skal kunne utrette noen særlig forskjell i helsetjenesten. Her virket det som at hens egen erfaring har gitt hen et annet perspektiv på tid og AI enn resten av informantene, og som også har ført til at det virker som at hen ser på AI som en teknologi som mange per nå er overdrevent optimistisk til, og som ikke er god nok til å kunne erstatte klinikerne.

Kapittel 5 – En iterativ prosess for utvikling

Nå har vi sett på hvilke fremtidsvisjoner informantene har angående AI i helsevesenet om 10-20 år og hvordan informantene tolker AI som en teknologisk artefakt. Hovedfokuset til denne masteroppgaven, er å se på hvordan AI-løsninger til helsetjenesten blir utviklet. Jeg nevnte tidligere at det er forventet at AI kommer til å spille en større og større rolle i livene våre i årene som kommer. Derfor kan det være viktig å kunne få et lite innblikk i hvordan en slik teknologi kan bli utviklet.

Hvordan utvikler man en teknologi som skal være «intelligent»? Hvordan starter i det hele tatt en slik prosess? Starter det bare ved at noen personer tenker at det kunne vært gøy å lage en AI som kan gjøre det og det, eller søker man etter å utvikle en AI som skal løse et faktisk behov? Hvordan kommer man opp med ideer, hva er de største og viktigste valgene som blir tatt i løpet av en slik utviklingsprosess, og hvem er det man eventuelt samarbeider med? Hvis man samarbeider med noen, hvordan fungerer det samarbeidet, hvem får være med i idéprosessen og komme med forslag og ideer, hvem får være med å ta beslutninger?

I dette kapitlet skal vi se på hvordan utviklingsprosessen bak denne AI-løsningen til helsetjenesten er, som dreier seg om innovasjon i helse, da AI-løsningen skal forbedre helsetjenesten, noe Helse Midt-Norge (*Innovasjon, 2024*) trekker frem når de forklarer hva innovasjon i helse er. For å få bedre svar på spørsmålene i forrige avsnitt, skal dette kapitlet se på møtene som informantene trakk frem som viktige for deres arbeidsprosess. Videre skal kapitlet også se på det tverrfaglige samarbeidet som finner sted og hvordan miljøet for å komme med innspill er. Kapitlet skal også fokusere på det som flere av informantene nevner når de snakker om hvordan arbeidsprosessen er, nemlig at det er en «iterativ prosess», samt hvilken rolle prototyper hadde i utviklingsarbeidet.

Møtene

En viktig ting å gjøre klart før dette kapitlet begynner, er at prosjektet jeg har studert ikke er ferdig utviklet i skrivende stund. Det ble satt på vent på grunn av manglende finansiering. Jeg har derfor bare fått studert utviklingsprosessen så langt som de kom i prosjektet. Dette gjorde også at noen av informantene snakket litt mer generelt om lignende AI-prosjekt de har hatt, og det var ikke alltid like klart om de snakket om prosjektet som jeg undersøkte, eller på mer generell basis om AI prosjekt de har hatt. Det virket som at informantene hadde en generell tilnærming til denne typen prosjekt. Jeg har derfor forsøkt å velge ut sitater der de snakker om, og der det virker som at de snakker om prosjektet som jeg ønsket å undersøke. Uansett er konteksten for intervjuene AI-løsningen for diagnostikk, og det er primært det prosjektet som oppgaven fokuserer på.

Etter å ha intervjuet de ulike aktørene som har vært med på utviklingen av dette prosjektet og andre AI-prosjekter som har med diagnostikk å gjøre, er noe av det første som kommer frem møtene mellom de ulike aktørene. På spørsmål om hvordan arbeidet med dette prosjektet startet, sier Forsker 1:

Ja, det startet egentlig sånn som det veldig ofte gjør i nesten alle prosjektene vi driver med. Det er at vi har et (...) altså vi har, helt tilbake til midten av 90-tallet, så har vi et tett samarbeid med [Sykehuset] og [Forskningsinstitusjonen] og da har vi regelmessige møter i de prosjektene som går, og så har vi diskusjoner om det er andre ting vi må se på, burde søke midler til å gjøre, og så videre, finansiering for å få satt i gang. Og det er der disse ideene kommer opp, det vil si at klinikerer sier at «du, vi har en utfordring her som vi ikke har sett på enda, som vi bør se på, for det kunne ha vært nyttig». Og det er rett og slett i slike faste, jevnlig møter, dialog med klinikerne, som presenterer sine utfordringer, og så hører vi på og prøver å forstå hva de egentlig har av utfordringer (...).

Som vi kan se av sitatet over, så har de hatt slike møter i en god stund, helt siden midten av 90-tallet. Det virker derfor som at dette er en bevisst strategi de har for å starte med prosjekter som skal dekke ulike behov som klinikerne har. Det er også i slike møter at arbeidet med dette prosjektet startet. Det som kan være interessant å legge merke til her, er at Forsker 1 ikke sier at det er de på Forskningsinstitusjonen som presenterer en idé til klinikerne, det er klinikerne som forteller om ulike utfordringer de har og så forsøker Forskningsinstitusjonen å lage en løsning som skal hjelpe klinikerne med denne utfordringen. Vi kan dermed se at det her er en etterspørselsdrevet prosess hvor løsningene skal løse et faktisk behov, ikke prøve å skape et behov. Det er dermed klinikerne som har definert et problem som de ønsker en løsning på. Forsker 1 sier at det er i disse møtene at «ideene kommer opp». Disse møtene virker derfor å være veldig sentrale når det kommer til å komme med ideer til prosjektene, og det kan tenkes at idéfasen finner sted her. Dette kan man også se av det som Forsker 3 sier om møtene:

De møtene vi har, holder vi ganske uformell og der er det alltid rom for å komme opp med nye ideer. Og det er mange som er veldig kreative, spesielt noen av klinikerne – veldig kreative og kommer med stadig nye ideer lenge før vi har rukket å gjøre noe som helst med den forrige ideen de kom med. Så vi har ikke spesielle kreative møter, men vi er kreative i de møtene vi har, kan du si. Så det er litt åpne, uformelle møter der det er rom for å komme med ideer underveis, absolutt.

Av det som blir sagt i sitatet over, kan vi se at det ikke bare er klinikerne som kommer med utfordringer, og så kommer de fra Forskningsinstitusjonen med ideer til løsninger, klinikerne kommer også med ideer. Vestal og Mesmer-Magnus (2020) skriver om «instrumental ekspertise», og sier at det dreier seg om ekspertisen eller kompetansen som et team har som kan brukes til innovasjon (Vestal og Mesmer-Magnus, 2020, s. 739). Basert på sitatet over, der det blir trukket frem at det er mange som er kreative, tyder på at aktørenes instrumentelle ekspertise er høy. Idéskapingen er dermed ikke en enveisprosess, den går begge veier. Forsker 3 sier også at møtene er «ganske uformell» og sier at det «alltid er rom for å komme opp med nye ideer». Dette tyder på at møtene er lagt opp på en måte som skal gjøre at det føles trygt å komme med ideer – det er dermed et trygt miljø i disse møtene for å komme med innspill og ideer. Dette kan gjøre det lettere for de som er med i møtene å tørre å komme med ideer som de mener kan funke bra, noe som nok vil hjelpe alle aktørene med å komme opp med gode ideer til løsninger. Dette kan man også se av det Forsker 3 sier om at noen av klinikerne er «veldig kreative» og «kommer med stadig nye ideer». Det er også interessant å se at de ikke har satt av noen spesielle møter som er viet til å være kreative, men at de er «kreative i de møtene vi har». Dette kan tyde på at disse møtene fungerer såpass godt når det kommer til idéskapning, at det ikke er behov for egne idémyldringsmøter. Forsker 2 sier følgende om møtene:

(...) Så når vi møtes jevnlig, så vil vi jo på en måte si «ok, (...) da vil vi jo søke på midler». (...) noen må jo da kikke på «ok, hva er det som er mulige utlysninger?», og så snakker vi om hva som er mulige prosjekter for de utlysningene. Og da blir vi jo enige om «ok, det her er det vi vil ... fokusere på den her», og da må vi jo da skrive den søknaden. (...) men det går jo da, på en måte, de samme møtene og de samme folkene, eller ja, (...) det går jo over lang tid da, sånn at folkene vil jo da byttes ut sånn etter hvert, ja. Men det er jo mange som har vært med hele tiden.

Ut ifra det som Forsker 2 sier om disse møtene, kan det virke som at prosjektene de setter i gang, ikke alltid kommer fra at klinikere presenterer utfordringer i disse møtene, men at de også kan se på ulike utlysninger som de kan tenke seg å utvikle løsninger til. Dermed virker det ikke som at aktørene er passive og sitter og venter på at klinikerne skal presentere problemstillinger og behov som de ønsker å få dekket. De er aktivt ute og leter etter prosjekter som de kan jobbe med. Dette viser oss også at prosjektidéer ikke bare kommer fra disse møtene de har med klinikere, men at de også kan komme fra diverse utlysninger. Det fremstår også som at det ikke settes i gang prosjekter, uten at alle parter er enige om å sette i gang. Dermed virker det ikke som at én gruppe har mer makt eller innflytelse enn andre, noe som tyder på et sunt samarbeid og samarbeidsklima. En annen ting det er verdt å legge merke til, er det Forsker 2 sier om at møtene går over langt tid og at personene som er med derfor byttes ut etter hvert, samtidig som hen også sier at mange har vært med hele tiden. Det virker dermed som at det varierer om personene som er med i møtene er de samme som har vært med en stund, eller om det er nye personer som deltar i møtet. Uten at det blir sagt her, kan det tenkes at erfaring av å delta på slike møter, påvirker arbeidsprosessen i møtene. Det er mulig at de personene som har vært med på mange slike møter over en stund, er mer erfarne i denne typen arbeidsmåte og at møtene kanskje blir mer effektive og produktive, enn hvis det er noen som er med for første gang. Dersom det er noen som er med for første gang, kan det hende at noe av tiden går på å ha en introduksjonsrunde der man forteller hvem man er og hva man gjør, i tillegg til at man ikke kjenner hverandre så godt, noe som kan påvirke hvor kreative man tør å være. En annen interessant ting med det Forsker 2 sier her, er at det ikke virker som at de er låst til å ha med akkurat de og de personene i møtene, men at det varierer, eller at det er fleksibelt hvem som deltar.

Som man har sett av det som har blitt skrevet over, så er disse møtene viktige når det kommer til arbeidsprosessen med dette prosjektet, men også for andre prosjekter. Det er her behovene kommer frem og hvor ideene skapes.

Tverrfaglig samarbeid

Vi har til nå sett på møtene som Forskningsinstitusjonen og klinikerne har med hverandre for å kartlegge behov, komme med ideer og sette i stand prosjektet. Der var vi litt innom at disse møtene er et tverrfaglig samarbeid mellom teknologer og klinikerne. Dette tverrfaglige samarbeidet skal vi se på nærmere nå.

I delen over, var vi innom viktigheten av møtene mellom Forskningsinstitusjonen og klinikerne. Dette blir også tydeliggjort av det Forsker 1 sier:

Ja, det er klinikere som har problemstillinger og så har vi jevnlig møter med dem. Vi kunne sikkert ha kommet på det prosjektet her uten å gjøre det på den måten også, men spesielt dette prosjektet tror jeg ikke hadde kommet i stand uten at vi har jevnlig og tett samarbeid mellom teknologer og klinikerne.

Vi har flere eksempler på prosjekter som kommer i stand gjennom en sånn jevnlig dialog og møteplasser mellom de to.

Her fremstår det som at disse møtene var avgjørende for at dette prosjektet kom i stand. Vi kan dermed se at det er forskjell på å komme på et prosjekt og det å faktisk sette det i stand. Alene er de i kapable til å komme på ulike prosjekter, men de er avhengig av å samarbeide med klinikerne for å få satt dem i stand. Dermed får møtene også en ny dimensjon; de er ikke bare viktige for å finne behov og ideer til løsninger som skal dekke disse behovene, de er også viktige for å faktisk få utført prosjektet. Denne typen tverrfaglig samarbeid som vi ser eksempel på her, minner om en av de tre typene tverrfaglighet som Sørensen (1999) nevner, nemlig at personer som har ulike faglige bakgrunner, i dette tilfellet i form av ingeniørbakgrunn og medisinsk bakgrunn, samarbeider og bruker sine fagdisipliner (Sørensen, 1999, s. 117). Det Forsker 1 sier om det tette og jevne samarbeidet med klinikerne, viser at de ikke bare har ett møte der klinikerne presenterer utfordringer og behov, og så kommer aktørene med ideer til hvordan man skal løse dette før de går hver til sitt - de fortsetter å møtes jevnlig. Dette tyder på at det er et samarbeid gjennom hele prosjektet, ikke bare i startfasen og idéfasen. Disse møtene består av et tverrfaglig samarbeid mellom teknologer og klinikere, to grupper med ganske forskjellige faglige bakgrunner. Likevel klarer de å samarbeide med hverandre, og basert på det som har blitt sagt av informantene om disse møtene, så er begge gruppene også gjensidig avhengig av hverandre.

Det at de ulike aktørene har et tverrfaglig samarbeid, gjør at de fungerer som et tverrfaglig team. Tverrfaglige team er ifølge Vestal og Mesmer-Magnus (2020) en sammensetning av eksperter som har unik kompetanse som har som formål å utvikle nye forståelser rundt eksisterende problemer, og som skal «utforske nye muligheter» (Vestal og Mesmer Magnus, 2020, s. 739). I dette tilfellet, er det tverrfaglige teamet satt sammen av personer fra Forskningsinstitusjonen, Sykehuset, Universitetet og Designbyrået. Alle disse personene har sine unike kompetanser, enten det er medisinsk kompetanse, ingeniørkompetanse, akademisk kompetanse eller kompetanse innen design. Sammen skal de utvikle en ny forståelse rundt hvordan de kan løse et problem som klinikerne har, og som klinikerne har presentert for dem, og se etter muligheter for hvordan de kan utvikle en løsning som kan hjelpe klinikerne med dette problemet.

Fra intervjuene, kommer det altså frem at det er flere aktører som er involvert i samarbeidet om denne AI-løsningen. Disse er Forskningsinstitusjonen, Universitetet, Sykehuset og Designbyrået. Om prosjektet og det tverrfaglige samarbeidet, sier Forsker 1 at hen synes at dette prosjektet er veldig givende, men trekker også spesifikt frem «det at det er så mange disipliner og kliniske inne samtidig gjør at det blir kjempeutfordrende å lage noe som alle de får noe ut av, og har nytte av». Dermed kan vi se at det tverrfaglige samarbeidet på en side er veldig positivt, men at det også bringer med seg sine egne utfordringer i form av at samtlige aktører skal være fornøyde med det som blir utviklet.

Forsker 3 sier følgende om hvordan de ulike aktørene samarbeider (dette sitatet handler nok om denne typen prosjekt generelt, men kan tenkes å også gjelde for det prosjektet jeg studerer her):

(...) Måten vi jobber sammen på, er at vi har et møte, et ukentlig møte. Hver onsdag har vi møte med teknologer i [Forskningsinstitusjonen] og noen studenter også som skriver masteroppgave, PHD-studenter og klinikere, fem stykker, inntil fem klinikere, eller i hvert fall fem stykker som er invitert til møtet hver onsdag, ikke alle som kan delta alltid. Så det er en veldig sånn fin arena for å diskutere veien videre og de ulike delprosjekt som vi jobber på.

Av det Forsker 3 sier i sitatet over, får vi et godt bilde av hvordan de ulike aktørene jobber sammen. Vi får også vite at det ikke bare er et samarbeid mellom Forskningsinstitusjonen, klinikere og designere, men også masterstudenter og PHD-studenter. Vi får også vite at de møtes hver uke og at det blir invitert inntil fem klinikere til disse møtene. Forsker 3 sier at det ikke alltid er slik at alle kan delta i møtene. Her kan vi se noe som kan være en potensiell utfordring med denne typen samarbeid, nemlig tid. Det er viden kjent at hvis det er noe klinikere har lite av, så er det tid. Dermed kan dette være en utfordring med tanke på at det ikke er sikkert at alle klinikerne som er invitert til møtene har tid til å delta. Dersom flere av klinikerne ikke har mulighet til å delta på disse møtene, og spesielt hvis dette skjer over tid, så vil det kunne være negativt for arbeidet med prosjektet, med tanke på hvor gjensidig avhengig alle partene er av hverandre, noe vi kommer inn på litt senere. Dette kan føre til utsettelse og forsinkelser, noe som ikke er positivt.

Forsker 1 gir oss et interessant innblikk i hvordan samarbeidet og rollefordelingen mellom de ulike samarbeidsaktørene er:

(...) Så vi pleier å gjøre et sånt trekantsamarbeid, hvor du har [Universitetet], [Forskningsinstitusjonen] og [Sykehuset], så er det [Sykehuset] som eier og definerer problemet og så driver [Universitetet] med basisforskning og PHD-utvikling og den biten, og så driver [Forskningsinstitusjonen] og lager selve prototypene og tester det sammen med klinikerne, og så får PHD'ene ta seg av publisering og artikler og den typen ting. Det er et veldig bra trekantsamarbeid det.

Her kan vi se at det tverrfaglige samarbeidet blir beskrevet som et «trekantsamarbeid» mellom Universitetet, Forskningsinstitusjonen og Sykehuset. Vi kan også se at de ulike aktørene har veldig klart fordelte roller og arbeidsoppgaver når det kommer til dette «trekantsamarbeidet». Forsker 2 snakker også om at de har et «veldig nært og tett samarbeid med klinikerne», og at dette ikke bare gjelder for det prosjektet vi fokuserer på her, men også i mange andre prosjekter. I tillegg til dette, sier hen at de klinikerne de samarbeider med her, også har vært med i andre prosjekter de har hatt tidligere. Dette tyder på at de har god erfaring med å samarbeide med hverandre, noe som vil kunne være fordelaktig å ta med seg videre, da partene har kunnskap om hvordan denne typen samarbeid fungerer og hvilke roller de har. Dette henger sammen med det som Vestal og Mesmer-Magnus (2020) kaller for «teamets relasjonelle ekspertise», som handler om den kunnskapen som teamdeltakerne gjennom tidligere samarbeid har opparbeidet seg og som har gjort at de har oversikt over hvilken ekspertise hver enkelt besitter og hvordan denne ekspertisen kan brukes (Vestal og Mesmer-Magnus, 2020, s. 743).

I tillegg til at Forskningsinstitusjonen samarbeider med klinikere og Universitetet, sier Forsker 1 at de samarbeider med «alle disiplinene», eksempelvis kreftleger, radiologer, patologer, lungeleger og kirurger. Dette viser oss at det ikke bare er et tverrfaglig samarbeid mellom Forskningsinstitusjonen, Sykehuset og Universitetet, men at det også er et slags tverrfaglig samarbeid mellom klinikerne som har ulike disipliner. Hen

fremhever også viktigheten av samarbeidet når hen sier at «vi var avhengige av å snakke med alle for å sikre at vi hadde all informasjonen som trengtes for å kunne lage en løsning som alle var fornøyde med». Dermed virker dette som et prosjekt som, i sin kjerne, er tverrfaglig. Forsker 2 forteller også om viktigheten av det tverrfaglige samarbeidet, og trekker blant annet frem dette samarbeidet når hen blir spurt om hva de viktigste eller største valgene som ble tatt var:

(...) men det andre er jo det der med at vi skal jobbe sammen med klinikerne. At vi skal hele tiden ha et godt samarbeid med klinikerne, at vi sammen med klinikerne prøver å finne de beste løsningene, ikke sant. Sånn at du ikke bare sitter og lager løsninger, så leter du «ok, hvordan skal vi lage noe og hva skal vi bruke det her til?». Og klinikere er jo også avhengige av oss teknologer, på en måte, på å kunne lage tekniske løsninger som de skal kunne teste ut (...). Og det er også en beslutning som er tatt langt tilbake i tid (...).

Av det som blir sagt her, kan vi se at det at Forskningsinstitusjonen skal ha et tett samarbeid med klinikerne, er en beslutning som ble tatt for en god stund siden. Det at dette er en beslutning som ble tatt for lenge siden, tyder på at dette er et klart strategisk veivalg som Forskningsinstitusjonen har tatt for hvordan de skal arbeide med prosjektene de har. Dette, kombinert med at Forsker 2 også sa at de samme klinikerne har vært med i tidligere prosjekter de har hatt, kan også tyde på at denne typen samarbeid er noe som både Forskningsinstitusjonen og klinikerne har erfaring med, noe som vil kunne være fordelaktig for begge parter. Dermed fremstår det som at aktørene er positivt innstilt til denne typen samarbeid. Innstilling, er ifølge Sørensen (1999), én av tre ting som tverrfaglighet handler om (Sørensen, 1999, s. 117). I tillegg til dette, kan vi også se at de ulike aktørene er gjensidig avhengig av hverandre. Forskningsinstitusjonen er avhengige av klinikerne for å kunne lage den best mulige løsningen, mens klinikerne er avhengige av teknologene som kan lage de løsningene som klinikerne trenger. Det at aktørene er gjensidig avhengige av hverandre, er også noe Medisiner 1 og 2 kommer inn på. Medisiner 1 sier de ikke hadde fått til det tekniske uten ingeniørene, og at ingeniørene «ikke hadde fått til det praktiske uten oss» og avslutter med å si at partene er helt avhengige av hverandre. Medisiner 2 sier:

Nei, for oss er det jo en ... altså vi hadde jo ikke fått det til. Vi driver jo med medisinsk-teknisk forskning, hvor veldig mye av det vi gjør er jo ingeniørvirksomhet. Så vi doktorer hadde ikke klart dette på egenhånd, rett og slett. Så det ... du kan si at ... ja, det er klart, vi har en litt forskjellig forståelse av muligheter, men det er en veldig grei ordning. Vi doktorer definerer kliniske behov, stort sett, og så prøver ingeniørene å tenke ut løsninger, men etter 15 år, så har vi jo ... så kjenner ... ingeniørene har blitt halvdoktorer og vi doktorer har blitt halvingeniører. Så sammen så har vi en veldig sånn felles forståelse over både ideer og begrensninger. Så det sklir helt uproblematisk.

Som vi kan se, er de enige om at de er helt avhengige av hverandre for at de skal kunne utvikle nyttige løsninger som hjelper klinikerne. Fra det Medisiner 2 sier, så kan vi også se det at begge partene, som nevnt tidligere, har fått erfaring i denne typen samarbeid, faktisk såpass med erfaring at hen kaller ingeniørene for «halvdoktorer» og doktorene for «halvingeniører». Det at hen sier dette, er veldig interessant da det peker på at det eksisterer en stor forståelse mellom ingeniørene og klinikerne. Ingeniørene vet og forstår hvordan klinikerne tenker, og klinikerne vet og forstår hvordan ingeniørene tenker. Det kan virke som at dette er et samarbeid som er veldig effektivt for aktørene, nettopp på

grunn av erfaringen de har opparbeidet seg. Alle er klar over sine roller, men klarer også å se ting fra de andre aktørenes perspektiv, noe som kan føre til større forståelse mellom aktørene og færre konflikter eller uenigheter. Også her kan vi se at dette henger sammen med det Vestal og Mesmer-Magnus (2020) sier om «teamets relasjonelle ressurser» (Vestal og Mesmer-Magnus, 2020, s. 743), som i dette tilfellet virker å være sterk på grunn av det Medisiner 2 sier om «halvdoktorer» og «halvingeniører», og at de klarer å se ting fra de andre teammedlemmenes perspektiv. Vestal og Mesmer-Magnus (2020) snakker også om viktigheten som teamets relasjonelle ressurser har for hvor godt et team klarer å integrere de ulike medlemmenes ekspertise (Vestal og Mesmer-Magnus, 2020, s. 744). Ut ifra det Medisiner 2 sier her, og det hen sier om at «det sklir helt uproblematisk», virker det som at de har klart å integrere aktørenes ekspertise på en god måte, noe som også tyder på at deres relasjonelle ressurser er gode. Dette er selvfølgelig positivt, ikke bare for stemningen mellom samarbeidspartnere, men også med tanke på at de slipper å bruke tid på konflikter og uenigheter.

Det virker som at Forskningsinstitusjonen og klinikerne synes at samarbeidet dem imellom fungerer veldig bra. Dette kan ha en sammenheng med at de har mye erfaring med å jobbe med hverandre over flere år. Men hvordan opplevde designeren som de hyret inn, det tverrfaglige samarbeidet?

Designeren syntes at det tverrfaglige samarbeidet var utfordrende på en side, og «veldig lett på en annen». Hen trekker frem at det som var utfordrende, var det faktum at de som designere ikke hadde den samme typen kompetanse på klinikerne og teknologenes fag, og at dette noen ganger førte til at det ble brukt terminologier som de ikke forstod. Dette henger sammen med det som Sørensen (1999) skriver om at forskjellig begrepsbruk kan være en utfordring i tverrfaglig samarbeid (Sørensen, 1999, s. 122). Videre sier hen at de hadde en person som pleide å være med designerne i intervjuene de hadde med klinikerne og som etter intervjuene kunne forklare designerne det som klinikerne snakket om. Hen sier at hen tror det hadde vært mer utfordrende dersom de ikke hadde hatt denne personen der og begrunner dette med at de ofte spurte klinikerne om hva ting betyr og hvorfor de sa det, men at det ikke alltid var mulig å spør om dette. Her kan man se at designerne virker å ha vært avhengige av å ha med en person som fungerte som en slags «oversetter» for dem når det kommer til det klinikerne sa i løpet av intervjuene. Det kan dermed virke som at klinikerne kanskje ikke har tenkt så mye over at designerne ikke forstår deres faglige terminologi. Hva som er grunnen til dette, er ikke godt å si. Kanskje det kan være fordi designerne var leid inn for dette prosjektet, men at de vanligvis ikke er en del av dette tverrfaglige samarbeidet, noe som kan ha gjort at klinikerne er vant til å samarbeide med personer som forstår den kliniske terminologien. Kanskje det var en utfordring å tilpasse seg til en ekstern samarbeidspartner? Det virker som at de ikke har opparbeidet seg så mye kunnskap rundt å samarbeide med Designbyrået tidligere da det fremstår som at de ikke har så mye erfaring med å samarbeide med hverandre. Alt dette tyder på at teamets relasjonelle ressurser ble svekket da det ble hentet inn en ekstern aktør i form av Designbyrået.

Videre fortalte Designeren at personen som fungerte som en oversetter for dem, alltid var med på intervjuene, og at Forskningsinstitusjonen også var med noen ganger. Noen ganger forklarte Forskningsinstitusjonen «hva som var bak teknologien». Men akkurat det med å forstå hva som var bak teknologien, trekker designeren frem som «ikke så vanskelig» fordi de bare kunne gå til teknologene for å prøve å forstå teknologien. Dette

var derimot ikke like lett med klinikerne, fordi de «ikke har tid». Dermed kan man se at det virker som at tidsaspektet, i tillegg til faglig terminologi, når det kommer til klinikerne, var noe som gjorde samarbeidet utfordrende for designerne.

Noe interessant som Designeren også trekker frem da hen snakket om hvordan hen opplevde samarbeidet, var at «vi så at teknologene ikke var helt vant til oss». Her kan det virke som at det at det har kommet inn en ny aktør i samarbeidet, i form av designerne, var noe som endret samarbeidsdynamikken til en viss grad fordi aktørene ikke var vant til å samarbeide med hverandre. Dette kan ses i lys av det som jeg snakket om tidligere, om at Forskningsinstitusjonen og klinikerne har fått god erfaring med å jobbe med hverandre over flere år, noe som gjør at samarbeidet dem imellom fungerer bra, men at når det kommer en ny aktør inn i samarbeidet, så skaper det utfordringer. Aktørene er vant til at samarbeidet fungerer på en viss måte, men når det kommer en ny aktør inn i samarbeidet, så må de samarbeide på en litt annen måte enn tidligere, noe som krever litt omstilling fra partene.

Miljø for å komme med innspill

Tidligere i kapittelet fikk vi et lite innblikk i hvordan idéene i dette prosjektet kom opp. Fra det som har blitt fortalt av informantene her, virker det som at det legges veldig opp til at det skal være lett å komme med ideer og innspill. På spørsmål om aktørene som Forskningsinstitusjonen samarbeidet med fikk komme med innspill underveis i prosessen, svarte Forsker 1 at alle sammen fikk si hva de hadde behov for i denne typen løsning. På spørsmål om innspillene ble tatt til betraktning og løsningen endret basert på disse innspillene, sa Forsker 1:

Ja, absolutt, det gjorde den, hele veien. Det var ganske viktig for prosjektet, for det var et av poengene med å ha både designere, teknologer og alle de klinikerne med – det var å få alle de innspillene som gjorde at vi klarte å lage en visualisering som alle syntes var nyttig (...).

Av det som ble sagt av Forsker 1 i sitatet over, kan man se at innspillene fra de ulike aktørene var viktig for prosjektet og at dette med å få innspill var et av «poengene» med å ha med alle de ulike aktørene. Innspillene var viktige for at de kunne klare å lage en løsning som alle parter var fornøyde med. Det at de tar innspillene til betraktning og endrer på løsningen basert på disse innspillene, tyder på at det legges opp til at man skal kunne komme med innspill og ideer. Dette kommer tydeligere frem av Forsker 3, når hen blir spurt om hvordan prosessen med å komme opp med ideer til løsninger i dette prosjektet er, og om det kommer flytende i samtaler de har, eller om de har spesifikke møter hvor fokus er idégenerering. Som vi kunne se at Forsker 3 snakket om tidligere i analysekapitlet, så sa hen at møtene de har er uformelle og at det «alltid er rom for å komme opp med nye ideer. Dette tyder på at det er lagt til rette for å skape et trygt miljø for å komme med ideer og innspill, noe som er viktig for at de menneskene som deltar skal tørre å være kreative og komme med ideer. Dette kan også tenkes å være viktig for at man skal kunne komme opp med så gode ideer som mulig, da det er mindre sannsynlighet for at noen har gode ideer, men ikke tør å si dem høyt på grunn av frykt for at ideene blir avvist.

Iterativ prosess

Frem til nå i dette kapittelet har vi sett på hvordan arbeidet med denne AI-løsningen startet og hvordan det tverrfaglige samarbeidet har vært. Denne delen av kapittelet vil fokusere mer på selve arbeidsprosessen bak dette prosjektet.

Etter å ha intervjuet de ulike informantene, er det én måte å beskrive arbeidsprosessen i dette prosjektet på som går igjen blant flere av informantene. Det er at det er en «iterativ» prosess. Kort fortalt betyr iterasjon «gjentakelse» (*Iterasjon*, 2018). Dette er noe som spesielt går igjen når informantene snakker om prosessen rundt hvordan de kom opp med ideer og laget prototyper til løsningen. Det første vi kan se på, er hvordan prosessen i idéfasen var.

Når informantene blir spurt om idéfasen, sa Forsker 1 at de hadde «en million ulike ideer» når det kommer til fremstilling og visualisering. Videre sa hen at de hadde «en svær plansje med et hundretalls av gullapper med ideer til ting som burde være med og ikke være med og måter å gjøre det på, hvem skulle styre (...)». Ut ifra dette, virker det som at det er mange kreative personer som jobbet med dette prosjektet. Basert på antallet ideer, virker det også som at deltakerne opplevde miljøet som trygt for å komme med innspill og ideer, som vi snakket om tidligere i kapittelet.

Forsker 2 sa, når hen ble spurt om hvilke ideer de hadde og om hen kunne ramse de opp, at «en god del av idéfasen er jo egentlig «ok, hva har vi laget fra før?» og bygge på det, fordi at vi, på en måte ... sånn er det jo med prosjektene, vi har jo begrenset med midler». Her kan man se at de ikke kun ser på nye ideer, men at de også ser på om de også kan bruke ting de har laget før, mest på grunn av begrensede midler. De begrensede midlene virker derfor å føre til at de må være kreative på andre måter enn kun å komme opp med nye ideer – de må se om de har laget ting før som de kan bruke i dette prosjektet og som de kan bygge videre på. Kreativitet i idéfasen, i dette tilfellet, kommer dermed i ulike former. En potensiell utfordring med å bygge på det man har fra før, er at man ikke nødvendigvis får den mest optimale løsningen, som man kanskje ville ha fått dersom man hadde hatt midler til å skape noe helt nytt.

Forsker 1 og 2 har gitt to ganske ulike svar når de blir spurt om idéfasen i prosjektet. Forsker 3 gir også et ganske interessant svar på spørsmål om hvem som var med i idéfasen. Her sier hen at «jeg vet ikke om jeg klarer å definere en idéfase» og forteller videre at de «har hele tiden en kontinuerlig idéfase av nye ideer». Det at de ikke har én fase i utviklingen av denne løsningen som er viet til idéskapning, men heller at dette er en kontinuerlig fase, er interessant. En kontinuerlig idéfase, tyder også på en iterativ arbeidsprosess, hvor man ikke gjør ferdig en og en fase som en slags sjekklister, men at man heller går litt frem og tilbake.

På spørsmål om hvordan prosessen med å komme opp med ideer til løsninger i dette prosjektet var, svarer Forsker 1:

(...) Det var en skikkelig sånn designprosess fra [Designbyrået]. Det starter med at [Designbyrået] er med på et slikt MDT-møte, multidisiplinær-teams-møte, med klinikerne, og vi sitter bare på bakerste rad og ser på og så sitter [Designbyrået] og observerer, noterer og tegner. Etterpå får de lov til å spørre og grave litt mer og forstå. Når de har gjort det, så går de og tegner og illustrerer og så tar vi et nytt møte, iterativt prøver å forstå enda mer ting som de må oppklare spørsmålene på, og så tegner man

opp et sånt flytskjema basert på alt som skjer i prosessen (...). Men det er selvfølgelig, det ligger noen rammebetingelser, det ligger noen data, og det ligger noen behov fra klinikerne, og så er det om å gjøre å iterere seg frem til en mulig løsning gjennom det. Det er en sånn (...) det heter vel en design-drevet innovasjon eller behovs-drevet innovasjon, det er litt avhengig av hvem du spør hvordan du definerer det.

Av det som blir sagt i sitatet over, får vi et godt bilde av idéprosessen i dette prosjektet. Ut ifra hvordan Forsker 1 forklarer prosessen, virker det som at dette er en prosess som er godt planlagt i forkant, hvor de har lagt opp en slags «løype» med flere runder som de skal gjennom for at idéprosessen skal være så god som mulig. Det virker som at Designbyrået har spilt en veldig viktig og sentral rolle i dette prosjektet, med tanke på hva Forsker 1 sier og det at hen kaller idéprosessen for «en skikkelig designprosess fra [Designbyrået]», samt at hen kalte det for «design-drevet-innovasjon». Medisiner 2 forteller det samme som Forsker 1 om at Designbyrået intervjuet klinikerne og laget en «visuell presentasjon» basert på behovene som kom frem og at de til slutt kom frem til en prototyp. Medisiner 2 sier at de fikk Designbyrået til å «lage en tjenestedesignprosess rundt dette».

Det kan se ut som at måten Designbyrået jobbet på, baserte seg mye på inntrykk fra møtene og samtaler med klinikerne. Som Forsker 1 sier, så «tegner og illustrerer» Designbyrået basert på hva de har observert i MDT-møtet og det klinikerne har sagt til dem når de har stilt dem noen oppklarende spørsmål. Etter dette, har aktørene et nytt møte hvor de iterativt arbeider med å oppklare eventuelle spørsmål og så tegner opp et «flytskjema basert på alt som skjer i prosessen». Her får vi et godt bilde på den iterative arbeidsprosessen i dette prosjektet hvor de går flere runder i idéprosessen for å komme frem til en god løsning som alle parter er fornøyde med. Forsker 1 sier at «det ligger noen rammebetingelser» der i form av eksempelvis data og klinikerne behov. Dette tyder på at man ikke har hatt helt frie tøyler i idéprosessen i dette prosjektet, men at man har hatt noen rammebetingelser man må forholde seg til og som er nødt til å være med i løsningen. Dette med rammebetingelser, kommer Medisiner 2 også inn på. Hen sier at de hadde en 3D-modell som de var «ganske klar på at den måtte inn» og at den nesten lå som «en av premissene for den endelige visualiseringen». I tillegg til denne 3D-modellen, nevner hen også at det var «mye annen kunnskap om pasienten som måtte inn». Dermed virker det som at de hadde noen faktorer som løsningen måtte baseres på og bygges rundt.

Prototyper

Vi har så vidt vært innom prototyper tidligere. Fra intervjuene, virker prototyper å ha vært en svært sentral del av den iterative arbeidsprosessen i dette prosjektet. Das og Das (2021) viser til at prototyper er noe som på en eller flere måter ligner på det produktet man er ute etter å skape (Das og Das, 2021, s. 984). Dette skal vi se litt nærmere på nå.

Forsker 1 sier flere ting som gir oss et godt bilde av den iterative prosessen som har vært i dette prosjektet. Blant annet nevner hen at de viste frem «en tidlig visualisering av en slags prototyp» til klinikerne og at de diskuterte med klinikerne hva de syntes om denne prototypen – hva som var bra, hva som måtte endres. Det med at de har en prototype som vises til klinikerne og som blir justert og endret på etter tilbakemeldinger fra klinikerne, nevnes også av Forsker 2 og 3. Forsker 2 sier at de hele tiden har en

iterativ prosess hvor de lager en prototype som går gjennom flere iterasjoner etter tilbakemeldinger fra klinikerne. Det at prototypen etter tilbakemeldinger fra klinikerne går gjennom flere iterasjoner, henger sammen med det Kelly (2001) sier om at man tidlig i prosessen kan finne ut hva som ikke fungerer (Kelly, 2001, s. 40). Her ser vi også et eksempel på det Kelly (2001) sier om at prototyper kan være med å gi aktørene et nytt perspektiv og bli klar over ting som de ikke hadde tenkt på tidligere (Kelly, 2001, s. 39). I tillegg til dette, forteller hen at ideene «blir justert hele tiden» og at dette skjer «fra uke til uke» og trekker frem de ukentlige møtene de har. Det virker dermed som at ideene og prototypene ikke gjennomgår store endringer, heller mer inkrementelle justeringer. Ut ifra det som har blitt sagt, virker det som at prototypene var veldig viktige for arbeidet med å komme frem til en løsning som alle parter er fornøyde med.

På spørsmål om aktørene som Forskningsinstitusjonen samarbeidet med fikk komme med innspill underveis i prosessen på akkurat hva de selv mente ville passe best til en sånn type løsning, sier Forsker 1:

Ja, de fikk alle sammen si hva de hadde behov for i en sånn løsning. (...) Og det var det de kom med underveis, at de fikk lov til å si, underveis i prosessen ... og så var det jo iterasjon, det vil si at vi viste frem en tidlig visualisering av en slags prototyp (...). Og så diskuterer man med hver enkelt av dem, hvordan ser dette ut, hva mangler her eventuelt? (...) ville det her ha brakt deg fremover, ville dette her ha forbedret møtet hvis her hadde vært løsningen? Så det var en veldig sånn åpen ... designerne var veldig flinke på å presentere på en sånn måte at de som intervjues får komme med sine innspill, behov.

Her kan vi se at Forsker 1 eksplisitt sier at det var iterasjon, men man kan også se en iterativ prosess i hvordan hen snakker om utviklingen av prototyper. Vi får dermed et eksempel på hvordan deres iterative arbeidsprosess er. I arbeidet med prototypen, vil dette si at de kommer opp med en prototype, viser den frem, får tilbakemeldinger, endrer prototypen basert på tilbakemeldingene og viser den frem igjen. Das og Das (2021) skriver at prototyper kan fremme både ideer og innovasjonskultur (Das og Das, 2021, s. 984). Det virker som at dette har vært tilfelle her også. Tilbakemeldinger virker å være en svært sentral del av den iterative prosessen. Det samme gjelder samarbeidspartnerne.

Designeren snakker også om den iterative prosessen, men med et litt annet perspektiv. På spørsmål om hvordan arbeidet til Designbyrået startet, sier hen at de ble kontaktet av Forskningsinstitusjonen og en kliniker fra Sykehuset som hadde et prosjekt der de allerede hadde en teknologi «som kunne (...) lage en 3D-modellering ut av bilder, 2D til 3D (...)». Videre forteller hen om at de bestemte seg for å observere et møte, altså et MDT-møte mellom klinikere. Der så de på hvor mange pasienter de så på i løpet av en time og hvor kort tid klinikerne hadde på å lese og se på bildene, for så å skulle fatte en beslutning. I tillegg til dette, sier Designeren at de også gjennomførte individuelle intervjuer med noen av disse spesialistene for å høre om antakelsene de hadde, stemte. Designeren sier også at de hadde med skisser som klinikerne kunne se på for å komme med tilbakemeldinger. Her trekker hen frem en viktig hendelse fra da de snakket med en radiolog. Hen sier at de tok med prototypene og skissene til denne radiologen og at dette møtet var litt «eye-opening» fordi radiologen kastet vekk mange av prototypene. Dette sier Designeren bare var fint fordi da kunne de bare gå videre. Dette sier hen også gjorde at kreativiteten deres også kunne gå videre. Det virker som at de på denne

måten, tidlig fant ut hva som ikke fungerte med prototypene, noe som Das og Das (2021, s. 984) og Kelly (2001, s. 40) trekker frem at prototyper kan gjøre.

Her får vi et bilde av den iterative prosessen som var når det kom til utarbeidingen av prototyper. Som vi kan se, ble flere av prototypene «kastet» bort av radiologen. Arbeidet med prototypene har derfor ikke bare vært rett frem, og vi kan også se, i likhet med det Forsker 1 sa, at tilbakemeldinger fra klinikerne var veldig viktige når det kommer til å utvikle en løsning. Det som er interessant med det som Designeren sier her, er at det å få så tydelige tilbakemeldinger på at prototypene ikke var bra nok, var fint, fordi da kunne både de og deres kreativitet komme seg videre. Vi kan dermed se, som Kelly (2001, s. 40) sier, at prototypene er viktige når det kommer til innovasjon, og som Das og Das (2021, s. 984) sier om at prototyper er viktig for å komme med nye ideer. Etter å ha gjort dette, forteller Designeren at de tok enda en runde med spesialistene. Her sier hen at de hadde laget nye justerte prototyper som de kunne se at spesialistene var mer fornøyde med, selv om de fortsatt fikk tilbakemeldinger om ting som ikke var bra nok.

Ut ifra dette, får vi enda et bilde på den iterative arbeidsprosessen med prototyper, samt hvordan arbeidet med prototypen utvikler seg etter hvert i den iterative arbeidsprosessen. Her kan vi se at designerne, etter å først ha fått sine opprinnelige prototyper avslått av klinikerne, kommer tilbake med prototyper som er oppdaterte etter tilbakemeldingene de hadde fått. Designeren sier at de kunne se at de nye prototypene passet bedre etter at de hadde blitt justert. Dermed får vi her et eksempel på at den iterative prosessen fungerte når det kom til å utvikle prototyper som på en bedre måte passet til klinikerne behov. Samtidig kan vi også se at to runder med iterasjon ikke var nok til at klinikerne var helt fornøyde med prototypen, det var fortsatt ting de ga tilbakemeldinger på som måtte endres.

Designeren forteller videre om at de laget en prototype «fra scratch til et nivå som man kan begynne å se på «er dette realistisk? Hva trenger vi til å lage den realistisk?»». I tillegg til dette, snakker hen om at de hadde flere runder med Forskningsinstitusjonen og teknologene der de snakket om hva som var mulig med tanke på farge og uttrykk og om dette var mulig å gjøre med denne teknologien. Fra det som blir fortalt her, virker det som at det er en kontinuerlig dialog mellom de ulike partene og at det aldri blir gjort noe uten at alle partene har forhørt seg med hverandre og er enige. Her får vi et bilde på den andre av tre faktorer som tverrfaglighet ifølge Sørensen (1999) handler om; ferdigheter når det kommer til kommunikasjon (Sørensen, 1999, s. 117). Det fremstår som at, selv om de opplevde litt utfordringer med tanke på samarbeid når Designbyrået kom inn i bildet, så har dette fortsatt fungert bra, og kommunikasjonen mellom alle aktørene virker å ha vært god. Det virker derfor som at samtlige aktører i dette samarbeidet har gode ferdigheter når det gjelder kommunikasjon.

Alt dette tyder på et sterkt samarbeid som er preget av høy grad av lojalitet og respekt. Dette gjenspeiles også fra det Designeren sier, når hen blir spurt om hva hen mener var hans viktigste oppgave i dette prosjektet. Her sier hen at det «var å sørge for at vi laget noe som var brukerorientert og at vi ikke skulle lage noe som skulle tilpasses teknologien». Det var dermed brukerne, og ikke teknologien som sto i sentrum og så må de klare å utvikle noe som passer deres behov. Dette er derfor et nytt eksempel på en etterspørselsdrevet prosess, som vi også kunne se tidligere i kapitlet. Det virker derfor som at det er konsensus mellom partene at de ikke skal utvikle noe bare for å gjøre det, men at det de utvikler faktisk skal løse et reelt problem eller behov. Behovene styrer

utformingen av teknologien, og ikke motsatt. Her virker det derfor som at deres tenkemåte, går imot en teknologideterministisk måte å tenke på.

Det man kan ta med seg fra denne delen av analysekapittelet, og som ble nevnt tidligere, er at det virker som at prototyper har vært veldig viktig for det iterative arbeidet i dette prosjektet. Prototypene virker ha gitt aktørene noe som er mer håndfast enn for eksempel bare en tegning, noe som virker å ha gjort det enklere for aktørene å se hva som fungerer, og hva som ikke fungerer. Videre virker det som at prototypene gjort det lettere for aktørene å komme med innspill og tilbakemeldinger og at det har gitt dem muligheten til å komme med mer presise tilbakemeldinger, slik at justeringene ble mer presise. Kelly (2001) snakket, som nevnt tidligere, om at prototyper er svært viktige for innovasjon (Kelly, 2001, s. 40). Det virker som at dette også har vært tilfellet med dette prosjektet, så langt som de har kommet til nå. Innovasjonen i dette prosjektet minner om det Kahn (2018) sier i sin andre definisjon av hva innovasjon er, blant annet at innovasjon er en ny enhet (Kahn, 2018, s. 454), i dette tilfellet i form av det som skal bli en ny teknologisk enhet. Det virker egentlig som at tilbakemeldingene og innspillene knyttet til prototypene, er det som har vært kjernen bak innovasjonen i dette prosjektet. Det fremstår også som at aktørene hele tiden har vært flinke til å se helheten i prosjektet, som er at de skal utvikle en AI-løsning for diagnostikk som skal fungere som et støtteverktøy for klinikerne, ikke en «fancy» teknologi som utvikles bare fordi den er kul og som prøver å tvinge frem et behov. Denne evnen man har til å kunne se helheten, er den tredje faktoren som Sørensen (1999) hevder at tverrfaglighet handler om (Sørensen, 1999, s. 117.). Dermed har vi også sett at disse tre faktorene som tverrfaglighet handler om; innstilling, ferdigheter når det kommer til kommunikasjon, og evnen man har til å kunne se helheten (Sørensen, 1999, s. 117), også går igjen i det tverrfaglige samarbeidet vi har sett på her.

Oppsummering

I dette analysekapittelet, så jeg nærmere på hvordan utviklingsprosessen av denne AI-løsningen har vært. Her fant jeg ut at møtene mellom aktørene var viktige for arbeidet med prosjektet. Disse møtene virker å være en bevisst strategi de har for å kartlegge og komme med løsninger som skal dekke klinikernes behov. Her kommer det frem at det er klinikerne som er problemeierne og som ønsker at Forskningsinstitusjonen skal utvikle en løsning som kan hjelpe dem. Prosessen med dette prosjektet er derfor etterspørselsdrevet. Det virker som at idéfasen i dette prosjektet fant sted i disse møtene. Det fremstår som at det er et trygt miljø for å komme med innspill. Videre, virker det som at både samarbeidet og samarbeidsklimaet er sunt, da det ikke fremstår som at en gruppe har mer makt eller innflytelse enn en annen. Vi fikk også vite at disse møtene har flere dimensjoner; ikke bare er de viktige for å kartlegge behov og komme med ideer til løsninger, de er også viktige for å faktisk få utført prosjektet.

Det tverrfaglige samarbeidet i utviklingen av denne AI-løsningen, virker å ha vært veldig viktig. En av informantene sa at for at de skulle klare å utvikle en løsning som alle parter ble fornøyd med, så var de «avhengige av å snakke med alle». Det virker derfor som at dette prosjektet har tverrfaglighet i sin kjerne. Det tverrfaglige samarbeidet virker å ha vært helt avgjørende i dette prosjektet. Det fremstår også som at det tverrfaglige samarbeidet er en bevisst strategi fra Forskningsinstitusjonen for hvordan de skal jobbe med prosjektene sine. Videre virker det som at det er en gjensidig avhengighet mellom de ulike aktørgruppene. Det virker også som at spesielt klinikerne og ingeniørene har en stor forståelse for hverandre. Samarbeidet mellom Forskningsinstitusjonen, Sykehuset og

Universitetet virker å ha vært veldig bra, men det ble mer utfordrende da Designbyrået kom inn i bildet. De andre aktørene virker å være godt vant til å samarbeide med hverandre, og denne erfaringen virker å gjøre at samarbeidet fungerer veldig bra. På en annen side, med en gang en ny aktør kommer inn i bildet, så fører det til at man må samarbeide på en annen måte enn før, som er utfordrende og krever omstilling.

Flere av informantene bruker ordet «iterativ» for å beskrive hvordan arbeidsprosessen i dette prosjektet har vært. En av informantene sa at de har en kontinuerlig idéfase. Dette tyder på at prosessen er iterativ og at de ikke har en «sjekklister» hvor de går gjennom en og en fase, men heller at de går frem og tilbake. Designbyrået var med i disse møtene og observerte, tok notater og tegnet basert på de inntrykkene de fikk. Dette gjelder også de inntrykkene de fikk etter å ha hatt samtaler med klinikere.

Prototyper virker å ha spilt en svært viktig rolle i utviklingsarbeidet av denne AI-løsningen. Det ble laget flere prototyper basert på tilbakemeldinger fra klinikerne, og de gikk gjennom flere iterasjoner. Ut ifra det som ble sagt, virket det som at prototypene gjennomgikk inkrementelle endringer, heller enn store endringer. Tilbakemeldingene fra klinikerne virker å ha spilt en viktig rolle i arbeidet med prototypene, da de ble mer og mer fornøyde med prototypene som Designbyrået hadde utviklet etter tilbakemeldingene. Prototypene virker å ha vært viktig for innovasjon i dette prosjektet. Designeren sa blant annet at prototypene gjorde «at vår kreativitet går videre».

Kapittel 6 – Forsvarlighet

Frem til nå, har vi sett på fremtidsvisjoner knyttet til AI i helsevesenet og hvordan utviklingen av en AI-løsning til helsetjenesten har foregått. En stor kontrovers når det kommer til AI i helsevesenet, men også AI generelt, er forsvarlighet. Ulike problematikker som eksempelvis «bias» er et problem mange frykter skal forekomme med bruken av AI. Opp gjennom årene har det vært flere tilfeller på AI-løsninger som ikke er forsvarlige. Et kjent eksempel på dette, er selskapet Amazon sitt forsøk på å utvikle en AI som skulle hjelpe dem i rekrutteringen av nye ansatte. Denne AI-modellen skulle gi dem som hadde søkt jobb hos Amazon en vurdering fra 1 til 5 stjerner. Formålet med AI-løsningen var at man skulle kunne gi den mange jobbsøknader, og så skulle AI-en vurdere seg frem til de fem beste søknadene. Det oppstod derimot et problem med denne AI-modellen. Amazon merket at AI-ens vurdering av søknadene ikke var kjønnsnøytral. Dette gjaldt ulike tekniske stillinger og programvareutvikler-stillinger. Grunnen til dette, fant de ut at var at AI-modellen hadde blitt trent på jobbsøknader som hadde blitt sendt inn over en ti-årsperiode for å se mønster i disse innsendte søknadene. Grunnen til at dette førte til at AI-modellen ikke utføre en kjønnsnøytral vurdering av jobbsøkerne, var at flesteparten av søknadene den hadde blitt trent på, kom fra menn. AI-modellen favoriserte derfor mannlige søkere og «straffet» søknader som inneholdt ord som «kvinnens». Selv om Amazon programmerte modellen slik at den skulle behandle slike ord på en nøytral måte, så betydde ikke det at AI-modellen ikke kunne finne andre måter å være diskriminerende mot slike ord og begreper. Dette prosjektet ble dermed skrotet til slutt (Dastin, 2018).

Her kan vi se et eksempel på hvordan bias kan forekomme i AI-modeller, og selv om ikke dette eksempelet har noe med AI i helsetjenesten å gjøre, viser den hvor raskt det kan gå galt når det kommer til forsvarlighet. Dette kan også skje med AI-modeller i helsetjenesten. Den største forskjellen mellom slike feil i en AI-modell som brukes i ansettelsesprosesser og en AI-modell til bruk i helsetjenesten, er at konsekvensen ved slike feil vil være mye mer alvorlig og farlig i helsetjenesten enn i en ansettelsesprosess. I helsetjenesten, vil en slik feil for eksempel kunne føre til at noen som har kreft, ikke blir diagnostisert med kreft, og dermed ikke få den behandlingen de trenger. Slike feil er dermed ikke særegne for AI i helsetjenesten, men konsekvensene når det skjer i helsetjenesten er mye større og mer alvorlig enn i de fleste andre industrier.

Det er fort gjort at eksempelvis bias forekommer i AI-modeller. Siden AI og forsvarlighet er et såpass viktig og kontroversielt tema, og forsvarlighet i helsetjenesten åpenbart er svært viktig, kunne det derfor vært interessant å se på hvordan de som utvikler AI-løsninger til bruk i helsetjenesten, jobber for at AI-løsningen skal være så forsvarlig som mulig. Hvordan påvirker forsvarlighet måten de jobber med å utvikle AI-løsninger på? Hva har det å si for deres arbeidsprosess?

I dette kapitlet skal vi derfor se på hvordan aktørene i spesielt Forskningsinstitusjonen jobber med å utvikle en AI-løsning som er trygg og forsvarlig. Kapitlet skal ta for seg hvilke konkrete grep som blir tatt i utviklingen for å unngå kjente problemer som bias, hva de gjør dersom de ikke kan samle inn tilstrekkelig mengder med treningsdata og hva som er utfordrende når det kommer til arbeidet med å gjøre AI-løsningen så forsvarlig som mulig. Det første vi skal ta for oss, er det som kalles «Forsvarlighetskravet».

Innen norsk helsetjeneste, har man noe som kalles for «forsvarlighetskravet». Dette går ut på at alle tjenester som ytes i helsetjenesten, skal være forsvarlig. Forsvarlighetskravet gjelder ikke bare for helsepersonell, men også for virksomheter. Dette gjelder blant annet at disse virksomhetene skal tilby tjenester som gjør at helsepersonellet klarer å «overholde sine lovpålagte plikter» (Helsedirektoratet, 2022, s. 18). Videre trekker denne rapporten frem det faktum at nye yrker og profesjoner i nyere tid er blitt mer involvert i helsetjenesten og dermed er med på å påvirke utviklingen av helsetjenesten. Derfor må disse yrkene og profesjonene også være klar over forsvarlighetskravet. Helsedirektoratet peker på viktigheten av at de som jobber med IKT i helsetjenesten er klar over at det de gjør vil medføre direkte konsekvenser for helsetjenesten og at det derfor er viktig at IKT-ansatte får bedre kompetanse rundt både risiko knyttet til helsehjelp og hva helsehjelp er. Helsedirektoratet ønsker at alle som er involvert skal være klar over forsvarlighetskravet. Dette gjelder også for hele utviklingsprosessen av ny teknologi som er ment for helsetjenesten (Helsedirektoratet, 2022, s. 18).

Utfordringer med forsvarlighet

Som vi har sett over, så er kontroversen rundt AI og forsvarlighet i helsetjenesten stor, og konsekvensene av en lite forsvarlig AI-modell i helsetjenesten kan bli store. Vi skal nå se nærmere på hva informantene mente var utfordrende i arbeidet med forsvarlighet. Når hen ble spurt om hva som var vanskelig når det kommer til forsvarlighet, sier Forsker 1 at «ekspert-annoteringen» er vanskelig fordi den tar tid, noe klinikerne har lite av. Hen sier også det vanskeligste er å «lage en gullstandard som vi kan måle alt opp mot». Videre sier hen noe som er interessant å legge merke til, nemlig at «det juridiske og det organisasjonsmessige med å få tak i data er selvfølgelig også en flaskehals», mens det å lage selve AI-modellen samt det å teste den, er «det letteste i hele prosessen». Det er interessant å legge merke til at det virker som at det er alt rundt selve AI-løsningen som er vanskelig, eller utfordrende, og ikke det å lage AI-løsningen. Dette er noen hen også gir et eksempel på fra et annet prosjekt:

(...) Vi har faktisk et prosjekt (...) der vi skulle ha 10 000 CT-er, og da brukte vi ett og et halvt år på det juridiske og det organisatoriske for å få dataen. Og når vi fikk dataen, så brukte AI-utvikleren to dager på å teste og validere modellen, og alt var i orden. Det sier jo litt om hva som er barrieren med innovasjon i helsesektoren.

Her får vi et bedre bilde på hva det er som tar lang tid i utviklingen av AI-modeller. Det organisatoriske og juridiske tar vesentlig lenger tid å få på plass, enn modellen i seg selv som, man kan se av sitatet over, kun tar noen dager. Til slutt i sitatet gjør Forsker 1 det tydelig hva hen mener er barrieren når det kommer til innovasjon i helsesektoren, altså det juridiske og organisatoriske.

Forsker 3 nevner flere ting når hen blir spurt om hva hen synes er mest utfordrende med å gjøre løsninger så forsvarlig som mulig. Blant annet sier hen at «når man skal trene opp AI-modeller, så blir jo modellen bedre jo mer data man har og mer differensiert dataen er da, fra ulike pasienter, ulike sykehus, kanskje tatt opp med ulikt utstyr». I tillegg til dette nevner hen flere ting, for eksempel at dersom man skal annotere, så burde dette gjøres av flere klinikere slik at man sikrer at det blir rett, og at når man skal teste modellen til slutt, så bør man teste den på data fra et annet sykehus og med en

annen lege. Her kan man se at det er flere faktorer som blir nevnt, og hvis man skal bryte det ned, så virker det som at Forsker 3 mener at det er det logistiske og alle variablene som må på plass, som er det mest utfordrende. Det blir ikke sagt eksplisitt av Forsker 3 her, men ut ifra alle faktorene hen nevner, så virker det som at dette er en prosess som kan ta litt tid.

Forsker 2 trekker frem en utfordring når det kommer til å trene opp en AI-modell på data, og sier at man kan trene modellen med mye data, men så kan det plutselig dukke opp noe som ser helt annerledes ut. Her kan man se at Forsker 2, på samme måte som Forsker 3, trekker frem utfordringen når det kommer til variasjon, og at det på grunn av denne variasjonen, kan være vanskelig å trene opp AI-modellen på realistiske data. Dette henger sammen med det Challen *et al.* (2019) sier om at operasjonelle data og treningsdata etter hvert ikke vil passe sammen på grunn av at sykdomsmønstre endrer seg over tid (Challen *et al.*, 2019, s. 233). Et annet potensielt problem hen trekker frem, med tanke på bias, er at AI-modellen kan «lure» klinikerne ved at man kan tro at AI-modellen har funnet alt den skal finne, men at det ikke er sikkert at den har det.

Medisiner 1 sier at hen ikke synes at det ikke er noe vanskelig når det kommer til punktet om forsvarlighet, og begrunner dette med at «vi har jo helt klare regler, og vi må søke og få ting godkjent før vi kan sette i gang». Her får vi et litt annet perspektiv på dette med forsvarlighet. Medisiner 1 virker å tenke mer på spesifikke regler, fremfor arbeidsprosess når det kommer til forsvarlighet.

Medisiner 2 har et litt annet perspektiv enn Forskerne på hva hen synes er utfordrende når det kommer til forsvarlighet. Hen sier følgende:

(...) det som blir vanskelig blir jo når AI's forståelse av bildene, når det gjelder kritiske ting vi er faktisk interessert i, ikke sammenfaller med vår egen. Hvis for eksempel den AI-en mener at den svulsten her er 3 cm stor, mens vi mener at den er 4 cm, for eksempel. Eller at akkurat selve svulsten er ikke så viktig, men hvis AI-en mener at den lymfeknuten der har kreft, og vi mener at den kvalifiserer ikke, så kan det medføre ekstraarbeid – at vi må gå inn og faktisk sjekke den med en vevsprøve (...). Per i dag, så finner AI i snitt litt mindre enn en røntgenlege, men den kan finne andre ting. Den kan finne ting som røntgenlegen har oversett, og så er det ting AI-en overser fordi at den er ikke programmert for å være god på akkurat det der.

Her kan vi se at Medisiner 2, mener at det at AI kan ha en annen forståelse av bildene enn det klinikerne har, kan være en utfordring fordi det vil medføre «ekstraarbeid». Dersom klinikerne er uenige med det AI-en har funnet, vil det kunne gjøre den mindre forsvarlig på en litt mer indirekte måte, ved at den medfører ekstraarbeid for klinikerne. Ekstraarbeid krever tid, noe klinikerne ikke har mye av. Dersom klinikerne må begynne å bruke mye tid på å dobbeltsjekke det AI-modellen har funnet, vil det potensielt gå utover andre pasienter. Medisiner 2 nevner både fordeler og ulemper med AI; den kan finne ting som klinikerne har oversett, men den finner fortsatt i gjennomsnitt litt mindre enn en røntgenlege samtidig som den også kan overse ting som den ikke er trent på eller «programmert for å være god på». Men det virker som at det er akkurat det med manglende samsvar mellom AI og klinikerne som er det som Medisiner 2 mener er vanskeligst.

Skafe nok data for å unngå bias

Noe som går igjen når informantene snakker om forsvarlighet og AI i helsetjenesten, er å få skaffet nok data. Når Forsker 1 blir spurt om hva hen tenker om forsvarlighet generelt med AI i helsetjenesten, sier hen følgende:

Jo, det er kjempeviktig. Når vi lager AI-modeller så handler det jo i første omgang om å skaffe nok data med tilstrekkelig variasjon for å dekke den naturlige variasjonen som finnes i den typen data som det er snakk om i hvert enkelt tilfelle. (...) Og den biten blir vi aldri ferdig med, tror jeg. Den blir vi ikke ferdig med før vi har latt modellen læres opp av alt som finnes av både CT-typer og ... ja bare produsenter av CT-maskiner, ikke sant, er jo enda en variabel osv. Men måten vi løser det på i dag, er at vi både bruker åpne datasett og søker om godkjenninger på å bruke store mengder lokale data (...).

Her kan vi se at Forsker 1 legger stor vekt på forsvarlighet når de skal utvikle AI-løsninger til helsetjenesten. Videre ser vi at det, ifølge Forsker 1, ikke bare handler om mengden data man klarer å samle inn, men også at det er nok variasjon i dataen man samler inn. Det fremstår derfor som at det er snakk om ganske mange ulike variabler som dataene de samler inn skal dekke for at algoritmene skal fungere så bra som mulig. Dette gjør at jobben med å samle inn nok data med tilstrekkelig mengde med variasjon, blir en stor jobb, som vi kan se av det Forsker 1 sier om at hen ikke tror de noen gang vil bli ferdig med den delen. Akkurat det hen sier her, er interessant da det kan tolkes som at det indirekte betyr at det er umulig å oppnå en helt komplett algoritme.

Å samle inn nok data som man kan føre algoritmene med virker å være en svært vanskelig oppgave, både fordi man må være klar over mange ulike variabler som kan spille en rolle med tanke på forsvarlighet og fordi at modellen burde bli lært opp av alt som finnes av variabler, noe som ikke er realistisk. I tillegg til dette, forteller hen om hvordan de løser dette i dag, nemlig med å bruke data fra «åpne datasett» samt at de søker om å få godkjenning på å kunne bruke lokale data. Her får vi et bilde på at det ikke bare er å gå ut å samle inn data man trenger til AI-løsninger, man må først søke om godkjenning. Dermed blir også tidsaspektet en faktor som kan påvirke utviklingene av denne typen løsninger, da man må vente på å få godkjenning for å samle inn data. Det at Forsker 1 nevner at hvem det er som produserer CT-maskiner er en variabel man også må tenke over, minner litt om det eksempelet som Challen *et al.* (2019) nevner i sin artikkel, hvor et maskinlæringsystem fungerte på en OCT-maskin, men ikke en annen (Challen *et al.*, 2019, s. 233). Dette tyder på at produsenter av eksempelvis CT-maskiner er noe man må ta høyde for på generell basis når det kommer til bias.

Dette med å samle inn nok data, går spesielt igjen når de blir spurt om hvordan de jobber for at AI-en som de bruker i denne løsningen ikke fører til «bias», altså skjevhet i data eller i resultat. Ifølge Barua (2023) kan bias forekomme i flere ulike varianter og at det som er hovedgrunnen til at bias oppstår i maskinlæringsmodeller, er skjevheter i datasettet (Barua, 2023, s. 231). Forsker 1, 2 og 3, samt Medisiner 1 og 2 snakker om det å samle inn nok data når de jobber for at AI-en eller algoritmene ikke fører til bias. Forsker 2 sier at «det går jo egentlig på å prøve å føre den med mest mulig realistiske data». Om hvordan de jobber for å unngå bias, forteller Forsker 1 at det også her handler om å samle inn nok data med nok variasjon, i tillegg til å også «få tak i data fra ulike leverandører av CT-maskiner, og på tvers av befolkningsgrupper». Hen sier også at «bias er et veldig sentralt element vi har i utviklingen hos oss, så vi er klar over den ja».

Her kan vi se at det igjen handler om at man klarer å samle inn nok data med tilstrekkelig mengde variasjon. Det er interessant å se at dette ikke er nok, men at de også må samle inn data fra forskjellige befolkningsgrupper og leverandører av CT-maskiner. Dermed kan vi se at det er mange variabler som spiller inn i arbeidet med å forhindre at bias skal oppstå i AI-modeller. Det at Forsker 1 trekker frem viktigheten av å samle inn data fra flere befolkningsgrupper kan tenkes å ha med at man skal unngå såkalt «underrepresentasjon» som Larrazabal *et al.* (2020, s. 12 592) skriver om. For at aktørene skal kunne samle inn data fra ulike befolkningsgrupper, er det viktig at det, som Larrazabal *et al.* (2020) trekker frem, eksisterer datasett som har data som både er balansert og mangfoldig. De skriver også at, ut ifra studien de har gjennomført, så kan de se at mangfold i datasettet fører til at generaliseringsevnen til AI-systemer øker (Larrazabal *et al.*, 2020, s. 12 593). Det kan tenkes at AI-modellen som aktørene utvikler her, også vil kunne få en bedre generaliseringsevne hvis den har data fra ulike befolkningsgrupper, noe som vil kunne forhindre bias. Celi *et al.* (2022) trekker, på samme måte som Larrazabal *et al.* (2020), frem viktigheten av mangfold i dataene som man trener AI-modellene med, og sier at for lite mangfold i disse dataene kan føre til bias (Celi *et al.*, 2022, s. 3). Det at aktørene legger vekt på å samle inn data med mangfold, er dermed noe som også trekkes frem som viktig i litteraturen. Måten hen snakker om at bias er sentralt i deres utvikling, tyder på at dette er noe som er med å prege utviklingen av AI-modeller og at de legger ned mye innsats for å unngå bias.

Forsker 3 kommer med et eksempel på hvordan de jobbet i et annet prosjekt med AI for å unngå bias. Her forteller hen at de har tatt opp data ved et sykehus fra rundt 40 pasienter. I tillegg til dette, sier hen at de også har tatt opp data ved et annet sykehus «for å kunne sammenligne og se». Videre sier hen at modellen først kun blir trent på data som kommer fra det første sykehuset, «og så blir det holdt igjen litt data fra [det første sykehuset] som den blir testet på». I tillegg til dette, sier hen at modellen også blir testet på det andre sykehuset slik at de kan sammenligne og se om modellen gjør det bedre på data fra det første sykehuset «fordi at den er tilpasset de som har tatt opp der, eller kanskje er det forskjell på utstyret de har der, i forhold til data fra et annet sykehus». Dette sier Forsker 3 kan være en måte man kan benytte seg av for å «se at modellen er generell» og hvis det viser seg at den ikke er generell «så må man kanskje ta opp data fra flere sykehus og flere operatører for å få en mer generell modell».

Her får vi et godt bilde på hvordan de kan jobbe for å unngå bias i AI-modellene de utvikler. Vi kan blant annet se at modellen testes på data fra flere ulike sykehus. Dette henger muligens sammen med det som blant annet Forsker 1 har sagt om variasjon i dataen de samler inn. Det virker dermed igjen som at mengde data og variasjon i dataen er det sentrale i arbeidet med å gjøre AI-modellen sikker. Celi *et al.* (2022) skriver at AI-modeller til bruk i klinisk sammenheng, kan fungere godt der det er blitt utviklet og bygget, men at det kan mislykkes når det brukes på andre steder (Celi *et al.*, 2022, s. 15). Det kan virke som at aktørene, ved å teste modellen på data fra flere sykehus, forsøker å unngå det som Celi *et al.* (2022) skriver her.

Forsker 1 snakker også om det hen kaller for «augmentering», og sier «vi utvider med syntetiske data når dataen vi har mangler en del variasjon som vi typisk mistenker at finnes». Augmentering innebærer, ifølge Mutasa, Sun og Ha (2020), at man, ved å for eksempel rotere bildene, gjør bildene skjeve eller beskærer bildene, kan øke størrelsen på treningsdatasettet man har. På denne måten kan man lage flere versjoner av de eksisterende bildene man har i treningsdatasettet (Mutasa, Sun og Ha, 2020, s. 98).

Forsker 1 sier videre at de blant annet kan simulere dårlig bildekvalitet og augmentere dette, og «plassering av svulster kan du også augmentere med å lage syntetiske plasseringer rundt omkring i lungene og så videre». Hen sier at man kan «gjøre mye på augmentering» og at dette kan føre til at man får en «mer robust modell». Dermed kan man se at de også har alternativer dersom de ikke skulle klare å få tak i nok data som dekker den nødvendige variasjonen. Det at de også har alternativer til dette, viser viktigheten av arbeidet for å unngå bias. Dette med augmentering er heller ikke noe spesielt vanskelig, ifølge Forsker 1.

Overfitting som risiko

Når Forsker 3 blir spurt om hvordan de jobber for å gjøre løsningen så forsvarlig, sikker og trygg som mulig, for å forhindre eksempelvis bias, nevner hen «overfitting» og at «de ser at de ikke «overfitter» modellen til akkurat de dataene de har». Dette er noe de, ifølge Forsker 3, gjør «i analysen underveis». Overfitting vil si at AI-modellen har blitt lært opp på en måte som gjør at den bare kan brukes på treningsdata, men ikke på andre typer data. Den er altså ikke generaliserbar til resten av populasjonen (Mutasa, Sun og Ha, 2020, s. 96). Det finnes derimot måter å håndtere problemet med overfitting på. En måte å gjøre det på, er å samle inn mer treningsdata. Dersom det ikke er mer tilgjengelig treningsdata, kan man foreta «data augmentation» (Mutasa, Sun og Ha, 2020, s. 97 og 98), som jeg var inne på litt tidligere. Her virker det som at dette er et risikomoment som aktørene er klar over, og at de forsøker å jobbe proaktivt for å forhindre det.

Akkurat dette med augmentering, er noe Forsker 3 snakker om når hen trekker frem et eksempel på hvordan de jobber for å forsøke å unngå bias i AI-løsninger. Eksempelen handler om luftveissegmentering som de tidligere har jobbet med, som de gjorde på et offentlig tilgjengelig datasett med rundt 500 CT-er. «Vi laget en modell først og så testet vi den på lokale data som er mer klinisk relevant for praksis her». Forsker 3 sa at modellen ikke fungerte så bra og at for at den skulle både fungere og passe bedre «på de lokale dataene vi har», så gjorde de en augmentering. Her får vi et slags bilde på hvordan de tester om en løsning fungerer slik den burde, og hva de gjør dersom de finner ut at den ikke fungerer bra nok. Her kan vi se at de bruker augmentering for å gjøre AI-modellen bedre, noe som passer sammen med det Mutasa, Sun og Ha (2020, s. 98) sier om hvordan man kan håndtere et problem som overfitting. Dette var et annet prosjekt enn det som denne masteroppgaven hovedsakelig fokuserer på, men uansett får vi et bilde på hvordan de gjør dette og at det de gjør samsvarer med det litteraturen sier at man kan gjøre dersom man opplever overfitting. I tillegg til Forsker 3, nevnte Forsker 1 også augmentering, som nevnt tidligere i kapittelet.

Medisiner 1, trekker også frem innhenting av data som viktig for å forhindre bias og sier at de henter inn data «etter spesifikke kriterier vi har satt på forhånd for at de dataene skal være så gode som mulig». Videre sier hen at de også vurderer bildene etter «definerte kriterier sånn at det da blir så likt og riktig som mulig da», og at de også markerer opptakene slik at «det skal være på riktig plass, så det ikke blir bias der». I tillegg til dette, snakker hen om viktigheten av hvordan dataen blir håndtert. Her nevner hen spesifikt at «det blir tatt opp likt og håndtert likt og at det da er etter litteraturen i dag». Her kan vi se at hen trekker frem litt andre ting enn det de fra Forskningsinstitusjonen gjør i tillegg til å hente inn data. Vi kan se at de har satt kriterier på forhånd for hvilke data de skal hente inn og for hvordan de skal vurdere bildene.

Dermed kan vi se at de ikke bare ukritisk samler inn mye data bare for å ha så mye data som mulig, men at de har det klart for seg hvilke data de er ute etter.

Medisiner 2 sier som Forskerne og Medisiner 1, at det er «om å gjøre å ha størst mulig datasett». I tillegg til dette sier hen at bias «i praksis for oss, er det jo om den finner for lite eller for mye» og at «biasen for oss går på at algoritmene blir for dårlige». Dette mener hen som oftest skjer på grunn av «for lite bakgrunnsmateriale» og at AI-en ikke har sett nok pasienter med tilstrekkelig med variasjon. Igjen, kan vi se at det å samle inn nok data med tilstrekkelig med variasjon er noe som informantene legger stor vekt på for å forhindre bias. Videre sier Medisiner 2 at de samler inn data selv, men at det også er store databaser fra andre steder i verden som de kan bruke. Her er det derimot ikke bare rett frem, ifølge hen. Dette på grunn av «at pasientgrunnlaget i de databasene, er jo ikke nødvendigvis overlappende med vårt pasientgrunnlag (...) det er riktige pasienter, men forskjellige problemstillinger». Det de jobber med nå, ifølge Medisiner 2, er «å skaffe oss selv et svært bildemateriale». Her kan vi se at de aktivt er ute og leter etter data som dekker nok variasjon slik at dataen er så bra som mulig. Dette henger sammen med det Forsker 1 har sagt om at de må ha data «med tilstrekkelig variasjon for å dekke den naturlige variasjonen» og det Forsker 2 sa om at de prøver å gi AI-modellene «mest mulig realistiske data». Fra det Medisiner 2 sier, kan vi se at det å samle inn data er utfordrende. Selv om dataen de samler inn er fra «riktige pasienter», så betyr ikke det at de dataene vil fungere optimalt, da de som de henter disse dataene fra, kan ha hatt andre problemstillinger enn det man selv har. Nettopp dette med hvilke problemstillinger dataen er samlet inn til, er også noe som Challen *et al.* (2019) skriver om. De skriver at data som er tilgjengelig, ikke nødvendigvis har blitt samlet inn fordi de er representative når det gjelder normalen, men fordi de er interessante. Dermed kan det oppstå bias (Challen *et al.*, 2019, s. 233). Det at det er såpass mange variabler og faktorer man må passe på når man samler inn data, tyder på at arbeidet med å forhindre at det skal forekomme bias i AI-løsninger, er krevende.

Så langt, virker som at informantene mener at hovedløsningen for å gjøre AI-modellen forsvarlig, bare er å ha nok data, og dersom de ikke har nok data, kan de bruke syntetiske data for å gjøre løsningene mer forsvarlig. Men er dette nok? Bruken av syntetiske data vil jo kunne bidra til at en AI-modell vil kunne bli mer forsvarlig, men spørsmålet er i hvilken grad? Hadde syntetiske data løst alle problemer når det kommer til forsvarlighet, hadde man jo ikke brukt tid på å samle inn data. Man hadde bare brukt syntetiske data, men dette er ikke tilfellet. Man samler fortsatt inn data, noe som tyder på at det å bruke syntetiske data ikke er helt optimalt.

Som vi har sett i dette kapittelet, har både Forskerne og Medisinerne hatt fokus på forsvarlighet i arbeidet med utviklingen av denne løsningen. Hva med Designerne? Designeren ble spurt om de også hadde et fokus på forsvarlighet når de holdt på å designe løsningen. Hen sier at «vi snakket ikke så mye om akkurat hvordan løsningen skulle være sikker og datasikkerhet og sånt» fordi dette var en mer «kreativ fase». Selv om de ikke hadde et så stort fokus på forsvarlighet, sier Designeren at forsvarlighet mest sannsynlig «skulle være neste steg om vi skulle gå videre [til] implementeringsprosess». Her nevner hen blant annet en faktor som «risiko for datalekkasje». Som vi da kan se av det Designeren sa, så har ikke de fokusert noe særlig på forsvarlighet så langt som de kom i dette prosjektet, men at de nok skulle jobbe med det i neste fase av prosjektet, som hen kaller for «implementeringsprosess». Dermed kan vi se at samtlige av aktørene som har vært med på utviklingen av denne løsningen, skulle ha fokus på forsvarlighet,

men at de rett og slett ikke kom langt nok i prosjektet til at designerne skulle fokusere på det.

Bare en forskningsplattform

En annen faktor som man kan si går på det med forsvarlighet, som gjelder for flere av prosjektene aktørene har jobbet med, men også dette prosjektet, er at det er en forskningsplattform. Dette er et poeng som, ifølge Forsker 2 er noe de er veldig opptatt av å gjøre klart for klinikerne, blant annet med å fortelle klinikerne at de ikke kan ta noen beslutninger basert på denne forskningsplattformen, men at de må bruke det godkjente utstyret som allerede er inne på operasjonsstuen. I tillegg til dette, forteller Forsker 2 at denne forskningsplattformen kun brukes i forskningsprosjekt. Et annet element som kan ha noe å si for klinikerne og pasientene, er det Forsker 2 sier om at de ikke «vil sløse med tiden til noen av dem som er inne på operasjonsstuen». Dette er, ifølge hen, en av grunnene til at de har brukt mye tid på å utvikle en løsning som er brukervennlig. Det at Forsker 2 sier at de er opptatt av å være tydelige overfor klinikerne om at denne løsningen bare er en forskningsplattform, kan også sies å være et grep de gjør med tanke på forsvarlighet.

Det at denne løsningen, og andre løsninger aktørene har jobbet med, er forskningsprosjekt og ikke medisinsk godkjent utstyr, er også noe Forsker 3 trekker frem:

(...) vi må være veldig tydelige overfor klinikerne at det ikke nødvendigvis er riktig alt en AI-modell gir ut og vi kan ikke forklare helt alt hvorfor den finner det den finner på den måten den gjør det. Så de kan bruke det som et hjelpeverktøy, men at de også må forsikre seg om at det er riktig med de standardbildene og metodene som de har, sånn i første omgang i hvert fall. For vi jobber jo på et sånn forskningsutviklingsnivå der vi ikke har ... altså vi lager ikke medisinsk godkjent utstyr, det er en litt senere del av prosessen (...).

Her kan vi se at Forskerne er opptatte av at klinikerne skal være klar over at AI-en ikke er ufeilbar, noe som gjør at de ikke skal basere sine beslutninger på den. Dermed kan vi se at det ikke bare er hvordan AI-løsningen er utviklet og hvor bra algoritmene er, som har noe å si for hvor forsvarlig den er. Challen *et al.* (2019) skriver om at resultatene som noen maskinlæringsalgoritmer kommer frem til, er «black-boxed», altså at det er vanskelig å si hvorfor algoritmen kom frem til akkurat det resultatet (Challen *et al.*, 2019, s.233). Dette er også noe Forsker 3 trekker frem når hen sier at de ikke nødvendigvis kan forklare hvorfor og hvordan AI-modellen finner det den finner. Her får vi et bilde av «black-boxen» når det kommer til AI, nemlig at vi ikke vet akkurat hvorfor den kommer frem til de resultatene som den gjør. Dette er noe Inga Strümke (2023) også nevner. Hun sier at data ikke nødvendigvis kan fortelle hvorfor akkurat det og det skjer (Strümke, 2023, s. 91). Vi kan også se at Forsker 3 sier at det med å lage medisinsk godkjent utstyr er «en litt senere del av prosessen». Dette tyder på at de har planer eller visjoner om å etter hvert få til å lage godkjent utstyr som klinikerne kan bruke og ta beslutninger på.

Oppsummering

I dette analysekapittelet, har jeg undersøkt hvordan aktørene jobber når det kommer til forsvarlighet. Her undersøkte jeg først utfordringer knyttet til arbeidet med forsvarlighet. Her virket det som at en av de største utfordringene, ikke hadde med AI-løsningen i seg

selv å gjøre, men heller alt annet rundt som det organisasjonsmessige og det juridiske. Videre kunne vi se at tidsaspektet, var en annen utfordring når det kommer til forsvarlighet. Dette gjelder, ifølge Forsker 1 «ekspert-annoteringen», som hen forteller er vanskelig fordi det er noe som tar tid, og klinikerne har ikke mye tid til rådighet. Hen trekker også frem det å skulle «lage en gullstandard som vi kan måle alt opp mot». Manglende samsvar mellom det AI finner og det klinikerne finner, er en annen utfordring og kan føre til ekstraarbeid. Det kan derfor tenkes at dette er noe som kan gå utover pasientsikkerheten dersom klinikerne må bruke mye tid på ekstraarbeid.

Når det kommer til hvordan aktørene jobber for å gjøre løsningen så forsvarlig som mulig, er det å skaffe nok data noe som går igjen blant informantene. Her fant jeg ut at det er mange variabler som skal dekkes av dataene som blir samlet inn. Det handler ikke bare å mengden data som blir skaffet, men også at dataen inneholder nok variasjon. Det å samle inn data, er derfor en stor jobb. Forsker 1 sier at «den biten blir vi aldri ferdig med, tror jeg». Det kan dermed virke som at det å skulle klare å lage en algoritme som er helt komplett, egentlig er umulig. Vi fikk også sett hva aktørene gjør dersom dataene de har samlet inn, ikke har tilstrekkelig med variasjon. Da gjør de en augmentering hvor de «utvider med syntetiske data». Dermed kan vi se at aktørene har en «plan-B» dersom de ikke har nok data, eller mangler variasjon i dataen de har. Det at de har en «plan-B» viser oss også at de tar dette med forsvarlighet seriøst og at dette er noe de er opptatte av.

Videre i analysen, kom vi inn på noe som heter «overfitting». Dette er noe som Forsker 3 forteller at de forsøker å unngå. En måte å håndtere dette problemet på, er med augmentering, noe aktørene gjør. Dermed kan vi se at de er klar over ulike risikomomenter som kan oppstå, og at de har metoder for hvordan de skal håndtere dette. Vi kunne også se at aktørene henter inn data etter forhåndssette kriterier. I tillegg kunne vi også se at de har «definerte kriterier» for å vurdere bildene slik at «det da blir så likt og riktig som mulig». Vi kan dermed se at data ikke bare blir ukritisk samlet inn for at man bare skal ha «nok» data, de har kontroll på hvilke data de er ute etter. Vi får også vite at de ikke kun samler inn data selv, men at de også benytter seg av databaser rundt om i verden. Ifølge Medisiner 2, er det ikke bare rett frem når de benytter seg av databaser for å skaffe data. Hen sier at dette er på grunn av «at pasientgrunnlaget i de databasene, er jo ikke nødvendigvis overlappende med vårt pasientgrunnlag (...) det er riktige pasienter, men forskjellige problemstillinger». Dette gjør at vi får enda et bilde av at det er utfordrende å samle inn data som er bra nok. Problemstillingene som dataene de henter inn er samlet inn etter, kan være forskjellig fra de problemstillingene som de selv har, noe som gjør at selv om dataene kommer fra «riktige pasienter», så vil ikke det automatisk si at disse dataene vil fungere optimalt.

Helt til slutt i kapittelet, kunne vi se at jobben aktørene gjorde med tanke på forsvarlighet, ikke bare gikk ut på hvordan de utviklet AI-løsningen rent teknisk, men også i form av at de var svært opptatt av å gjøre det klart for klinikerne at løsningen bare er et forskningsprosjekt, og ikke noe de skal basere sine beslutninger på. Teknologene la også vekt på at klinikerne skulle forstå at AI-en ikke er ufeilbar.

Kapittel 7 - Konklusjon

I denne masteroppgaven har jeg sett på utviklingsprosessen bak AI-løsninger til helsetjenesten i Norge. I den anledning, utviklet jeg følgende problemstilling: «*Hvordan utvikles AI for diagnostikk i et norsk tverrfaglig forskningsprosjekt?*». I tillegg, utarbeidet jeg tre underproblemstillinger, én til hvert analysekapittel:

«Hvilke fortolkninger og fremtidsvisjoner har prosjektdeltakerne rundt AI?»

«Hvordan er prosjektet organisert?»

«Hvordan forholder prosjektet seg til forsvarlighetskrav til helseteknologi?»

For å få svar på problemstillingen og underproblemstillingene, valgte jeg å fokusere på idéprosess, innovasjonsprosesser, arbeidsprosess, beslutningsprosesser og hvordan de jobber med tanke på forsvarlighet, som er et meget sentralt tema når det kommer til AI, og spesielt når det gjelder AI til bruk i helsetjenesten. I tillegg så jeg på informantenes fremtidsvisjoner og fortolkninger av AI. Målet med dette masterprosjektet er dermed å skape en større forståelse for hvordan prosessen bak utviklingen av AI-løsninger til den norske helsetjenesten er. Dette innebærer en større forståelse for hvem som får delta i idéprosessene, hvordan de gjør det i idéfasen med tanke på blant annet idéskaping, hvem som får delta i beslutningsprosessene, og hvordan de jobber for å gjøre AI-løsningen så trygg og forsvarlig som mulig for klinikere og pasienter, spesielt med tanke på bias. Under, følger det sammendrag fra analysekapitlene hvor jeg trekker frem hovedfunnene som svar på problemstillingen og underproblemstillingene. Sammendragene til analysekapitlene under, vil dermed fungere som svar på hovedproblemstillingen i sin helhet, mens hvert enkelt sammendrag også vil fungere som svar på hver sine underproblemstillinger.

I kapittel 4, undersøkte jeg hvilke fremtidsvisjoner informantene hadde om AI i helsetjenesten, samt hvordan de tolket AI som en teknologisk artefakt. Her viste analysen at alle informantene hadde de samme visjonene om hva AI i fremtidens helsetjeneste kommer til å være; et støtteverktøy for klinikerne. Analysen viste også at de fleste av informantene tolket AI som en nærmest «revolusjonerende» teknologi med et stort potensiale. En av informantene virket å ha en annen tolkning av AI og virket ikke å tolke det som en like «spesiell» eller «revolusjonerende» teknologi, heller bare en annen form for teknologi.

I kapittel 5, undersøkte jeg hvordan arbeids- og utviklingsprosessen i dette prosjektet var. Her viste analysen at strategien som Forskningsinstitusjonen har om å ha faste møter med klinikerne, er viktig når det kommer til å skulle starte med prosjekter som skal dekke klinikernes behov. Analysen viste også at prosessen er etterspørselsdrevet og at løsningene som utvikles, utvikles for å dekke et faktisk behov, og ikke for å skape et behov. Møtene er lagt opp på en måte som skal legge opp til et miljø hvor det er trygt å komme med både ideer og innspill. Videre viste analysen at disse møtene har en annen dimensjon, som går på at de er viktige for at dette prosjektet faktisk kunne blir utført. Det tverrfaglige samarbeidet mellom Forskningsinstitusjonen, Sykehuset, Universitetet og Designbyrået, har vært svært viktig for utviklingsprosessen av dette prosjektet. Viktigheten av å få tilbakemeldinger fra alle aktørene var også stor for å kunne utvikle en

så god løsning som mulig. Analysen viste også at det eksisterer en gjensidig avhengighet mellom aktørene. En av informantene trekker frem at Forskningsinstitusjonen og klinikerne har hatt et samarbeid i over 15 år og sier at ingeniørene har blitt «halvdoktorer» og doktorene har blitt «halvingeniører». Dette viser at aktørene har fått veldig god erfaring i denne typen samarbeid og at de har tilegnet seg stor forståelse for hverandres roller og kompetanse. Videre kom det frem at utviklingsprosessen av denne AI-løsningen, er iterativ. Det virket heller ikke som at de hadde én idéfase i dette prosjektet, men heller at den er kontinuerlig. Den iterative prosessen kom spesielt godt frem i arbeidet med prototypene. Analysen viste at prototypene var viktig i utviklingsprosessen. De utviklet flere prototyper som ble endret etter tilbakemeldinger fra klinikerne. På den måten, var tilbakemeldinger, og dermed også de ulike samarbeidsaktørene, viktig for den iterative prosessen. Designere fremstod også som å ha vært viktig i utviklingen, spesielt når det kom til prototypene. Nettopp disse prototypene virker å ha gjort det lettere for de ulike aktørene å se hva som fungerer og hva som ikke fungerer ved at prototypene ga dem noe som var mer håndfast enn en tegning. Kjernen bak innovasjonen i dette prosjektet, ser ut til å ha vært innspillene og tilbakemeldingene rundt prototypene. Alt i alt, virket det som at samarbeidet mellom de ulike aktørene er sterkt, og også preget av en stor grad av respekt og lojalitet. I tillegg viste analysen at det er brukerne som er i sentrum, og ikke teknologien, samt at det er behovene som former teknologien, og ikke motsatt.

I det siste analysekapittelet, undersøkte jeg hvordan aktørene jobbet med tanke på forsvarlighet. Analysen viste her at viktigheten av å skaffe nok data med nok variasjon, var svært stor i dette arbeidet. Videre fant analysen ut at det er flere utfordringer når det kommer til å jobbe med forsvarlighet. Det er ikke nødvendigvis det å lage AI-løsningen som er vanskelig, men heller alt annet rundt, som eksempelvis det juridiske og organisasjonsmessige. I tillegg er det er veldig mange variabler som skal på plass og som man må være klar over når det kommer til å samle inn data, i tillegg til mye logistikk. Ut ifra det noen av informantene sa, virker det som at det er umulig å skulle lage en komplett algoritme, samt det å gi mest mulig realistiske data til en algoritme. Dette på grunn av man kan trene modellen på mye data, men så kan det komme opp noe som er helt nytt. Dette viser at variasjon utgjør en utfordring. Videre fant analysen at aktørene har alternative løsninger, i form av syntetiske data og augmentering, dersom de ikke klarer å få tak i nok data som inneholder nok variasjon. Aktørene er også klar over ulike typer risiko, eksempelvis i form av «overfitting», og har bevisste tiltak de gjør underveis i utviklingen for å sikre at de ikke «overfitter» AI-modellen. Her virker det som at de er proaktive for å hindre at modellen skal bli overfittet. Analysen viste også at aktørene har andre måter å jobbe med forsvarlighet på, enn kun det som har med utvikling av AI-modeller, algoritmer og innsamling av data og gjøre, og det er å gjøre det tydelig for klinikerne at AI-løsningen bare er et forskningsprosjekt og at de ikke skal basere beslutningene sine på denne løsningen.

Alt henger sammen

Selv om oppgavens analysekapitler står hver for seg og tar for seg ulike tema, kan jeg se at det er en sammenheng mellom dem. Det ser ut som at det er en sammenheng mellom informantenes fremtidsvisjoner, den AI-løsningen de har jobbet med å utvikle, og hvordan de har organisert seg i utviklingsprosessen. Det virker også å være en

sammenheng mellom forsvarlighet og det tverrfaglige samarbeidet. Dette skal vi se litt nærmere på nå.

De fleste informantene så for seg at AI i fremtidens helsetjeneste, kom til å fungere som et støtteverktøy for klinikerne. Og det er nettopp en slik løsning de har jobbet med å utvikle; et støtteverktøy. Fremtidsvisjonene gjenspeiles også i det samarbeidet som aktørene har. Det er et samarbeid mellom de teknologene som kan utvikle en slik AI-løsning, og klinikerne som AI-løsningen skal fungere som støtte til. Det er ikke godt å si om det er fremtidsvisjonene som gjør at AI-løsningen de utvikler også er et støtteverktøy og at de samarbeider med klinikere fordi løsningen skal hjelpe dem, eller om det er prosjektet de har jobbet med som har formet deres fremtidsvisjoner om AI. Det kan hende at grunnen til at teknologene samarbeider med klinikerne, er fordi deres fremtidsvisjoner om at AI skal fungere som en støtte for klinikerne, gjør at de tenker at dette best kan oppnås ved å samarbeide med klinikere. Samtidig kan det også hende at det faktum at de samarbeider med klinikere, gjør at det blir vanskelig å se for seg at man skal utvikle en AI-løsning som skal erstatte dem. Det kan jo også tenkes at hvem det er som er med på å utvikle denne typen AI-teknologi, påvirker hvilket formål denne teknologien skal ha. De som eventuelt hadde blitt erstattet av AI, altså klinikerne, er også de som kan sies å sitte i førersetet i dette prosjektet, og det kan derfor tenkes at denne AI-løsningens formål, på grunn av dette, er å fungere som et støtteverktøy for klinikerne. Men igjen, det er ikke godt å si om informantenes fremtidsvisjoner er som de er på grunn av prosjektet de har jobbet med, eller om prosjektet de har jobbet med har det formålet det har på grunn av informantenes fremtidsvisjoner. Det kan jo tenkes at de henger sammen.

En annen sammenheng jeg kan se mellom de ulike analysekapitlene, er når det kommer til forsvarlighet og måten aktørene har organisert utviklingsprosessen på i dette prosjektet. Som vi kunne se i kapittel 6, som handlet om forsvarlighet, så forteller aktørene at forsvarlighet er noe som de er opptatt av. Det tverrfaglige samarbeidet kan derfor være et resultat av dette. Måten aktørene har lagt opp utviklingsprosessen på, som er at det er en iterativ prosess som baserer seg på hyppige møter, stor grad av deltakelse blant alle aktørene, samt vekten de legger på tilbakemeldinger fra hverandre, kan også henge sammen med forsvarlighet. Ved å legge stor vekt på å gi tilbakemeldinger og endre på løsningen basert på disse tilbakemeldingene, er dette en måte å utvikle en AI-løsning på som fungerer best mulig ut ifra det aktørene har kapasitet til, noe som igjen vil gjøre løsningen mer trygg og forsvarlig. Ved å samarbeide med klinikere, vil de kunne få gode og spesifikke tilbakemeldinger på akkurat hva det er klinikerne har behov for og hva de synes er bra og mindre bra i løsningen som blir utviklet. Det å hyre inn Designbyrået som er gode på digitale løsninger, er også en måte å kunne utvikle en så god løsning på som mulig. Dette vil føre til at Forskningsinstitusjonen på en bedre måte kan utvikle en AI-løsning som kan dekke klinikernes behov, og på den måten være mer forsvarlig. Det at det ikke virker som at det blir tatt noen beslutninger før alle aktørene er enige, kan også føre til en løsning som er mer forsvarlig, da man vil kunne unngå å gå glipp av viktige innspill. Denne måten å organisere utviklingsprosessen på, kan derfor ses på som en måte å organisere seg på for å kunne utvikle en så forsvarlig løsning som mulig. Dette samarbeidet er et eksempel på en løsning som er «nedenfra og opp». Dette er en måte å både jobbe og samarbeide på som jeg mener er veldig fordelaktig, både når det kommer til å dekke behov, men også for å lage en forsvarlig AI-modell.

Det at denne arbeidsmåten er fordelaktig, forsterkes av et relativt ferskt og dagsaktuelt eksempel hvor det virker som man har gjort det motsatte av det som er tilfellet her, nemlig Helseplattformen. Helseplattformen er en annen type løsning enn en AI-modell som skal brukes som et støtteverktøy, men er like fullt en teknologisk løsning, akkurat som en AI-modell, til bruk i helsetjenesten. Helseplattformen er mer en løsning som kommer «ovenfra og ned» i motsetning til prosjektet som jeg har sett på i denne masteroppgaven. I Helseplattformens tilfelle, er løsningen kjøpt utenfra fra en ekstern aktør. Dette har skapt problemer for brukerne, altså ansatte i helsetjenesten. En «superbruker» av Helseplattformen, skriver blant annet på Dagens Medisin at Helseplattformen er lite brukervennlig, og at Helseplattformens arbeidsflate er kompleks. I tillegg skriver hun at måten Helseplattformen er bygget på, gjør at det er vanskelig å finne den informasjonen man er ute etter. Videre sier superbrukeren at informasjon har forsvunnet når hun har brukt filter for å redusere mengden informasjon, som hun kaller «journalstøy», og at dette potensielt er en fare for pasienter (Grandaunet, 2023).

Hvorfor et forskningsprosjekt?

En ting som går igjen blant noen av informantene, er at denne AI-løsningen er et forskningsprosjekt eller forskningsplattform, og ikke medisinsk godkjent utstyr. Hva har dette å si for måten de jobber på, og hva de ønsker å oppnå med dette prosjektet? Det at dette er et forskningsprosjekt, gjør at de holder på med noe annet enn direkte produktutvikling. Målet med dette forskningsprosjektet, er ikke nødvendigvis i første omgang at de skal utvikle et teknologisk produkt som kan selges til sykehus rundt om i Norge eller i andre land, men heller at de skal utvikle kunnskap og ekspertise rundt denne typen AI-løsninger til helsetjenesten. Hva fungerer, og hva fungerer ikke? Dermed vil ikke prosjektet nødvendigvis være mislykket dersom AI-løsningen viser seg å ikke virke slik som de håper, fordi de fortsatt har fått opparbeidet seg kompetanse rundt bruk av AI i helsevesenet. Det kan tenkes at siden dette er et forskningsprosjekt, så er ikke målet bare å utvikle en AI-løsning som kan fungere som et støtteverktøy for klinikerne. Et annet viktig bidrag en slik løsning kan gi, er at det kan gi viktig kunnskap eller kompetanse om AI-løsninger i helsetjenesten som kan hjelpe aktørene på veien videre.

Veien videre - AI som svaret på alle utfordringer?

I innledningen til denne masteroppgaven, skrev jeg at det er mange som ser på AI som en teknologi som skal gjøre alt bedre og lettere for oss mennesker. Her stilte jeg også spørsmålet om det virkelig er slik? Er det så lett? Av det som har kommet frem fra intervjuene jeg har gjort, kan jeg se, spesielt når det kommer til forsvarlighet og AI, at det er veldig mange ulike faktorer som skal på plass for at en AI-modell skal bli forsvarlig og at det er en del utfordringer når det kommer til forsvarlighet. For det første, er det vanskelig å få tak i nok data til å trene AI-modeller. Noen av informantene fortalte at de har løsninger for hva de skal gjøre dersom de ikke får samlet inn nok data, og det er at de gjør en augmentering og benytter seg av syntetiske data. Selv om dette er en løsning som fungerer, så kan man tenke seg at bruken av syntetisk data ikke vil være like bra som «vanlig» data. Hvis det hadde vært like bra med syntetiske data, så hadde jo ingen tatt seg bryet med å samle inn virkelige data. Det å bruke syntetiske data er derfor ikke ideelt og man kan derfor ikke slå seg til ro med at det å bruke syntetiske data, dersom man ikke har klart å få tak i nok data, vil løse alle problemer.

Og dersom man får tak i nok data, så betyr heller ikke det at AI-modellen blir forsvarlig. Som noen av informantene var inne på, så kan de dataene man samler inn, opprinnelig

være laget for en annen problemstilling enn det man selv skal bruke dem til. Dermed er det ikke sikkert at dataene passer så bra til andre formål. Og selv om man har fått samlet inn mye god data som man kan trene AI-modellen på, så kan det oppstå situasjoner hvor man finner noe nytt som man ikke har data på og som man kanskje ikke finner data på noen annen plass heller. Da vil man være på bar bakke igjen. AI-modeller bygger på data som er samlet inn tidligere fra tidligere caser, noe som er et grunnleggende problem med AI. Dersom man skulle finne nye medisinske tilfeller som man ikke har data på, så må man klare å integrere denne nye dataen i AI-modellen. Da er man avhengige av at AI'en kontinuerlig blir oppdatert med nye data, og da må AI-modellen også være kompatibel til å ta imot og samle inn nye data. Det er ikke utenkelig at dette vil bli et problem som berører alt som har med AI i helsetjenesten å gjøre, og feil her kan i ytterste konsekvens føre til at liv går tapt.

Tidligere i oppgaven, snakket jeg om alle problemene som innføringen av Helseplattformen har brakt med seg, og at dette er et system som ikke er like komplekst som det som AI-systemer i helsetjenesten kan bli. Derfor kan dette tyde på at det å skulle implementere AI i helsetjenesten ikke nødvendigvis kommer til å gå helt problemfritt. Alt som er nevnt i avsnittet over, vil også føre til at AI i helsetjenesten blir et veldig stort og komplekst system. Her dukker flere utfordringer opp; hvem er det som skal jobbe med dette? Hvem skal ha ansvaret for systemet? Hvor i verden skal dette gjøres – lokalt eller globalt? Alle disse faktorene gjør at man kan stille spørsmålet om AI i helsevesenet vil kunne bringe kunnskapen og helsetjenesten fremover, eller om den heller vil fungere som en slags bremse? Kan AI i helsetjenesten bli like mye et problem som en løsning?

Denne masteroppgaven er en studie av hvordan en AI-løsning til helsetjenesten har blitt utviklet. I tiden fremover, er det ikke utenkelig at AI i helsetjenesten kommer til å bli mer og mer aktuelt, og at flere og flere AI-løsninger til helsetjenesten kommer til å bli både utviklet og implementert. Som forslag til videre forskning, kan det derfor være interessant å gjøre flere studier av utviklingsprosessen av AI-løsninger til bruk i helsetjenesten. På den måten kan man få et enda bedre bilde, og skape en enda større forståelse for hvordan denne typen teknologiske løsninger blir utviklet. Kanskje ser man likheter i hvordan andre AI-løsninger blir utviklet med det som jeg har funnet her, eller kanskje er utviklingsprosessen helt annerledes? Hvordan er det denne litt mystiske teknologien blir «intelligent»?

Referanseliste

- Amundsen, O. (2017). Hva er innovasjon? Tilgjengelig fra: <https://o365addins.it.ntnu.no/RadUtvalg/Widget/DownloadWidgetDataFile?url=/sites/meeting-instituttledermotet-su-fakultetet/Delte%20dokumenter/13/Hva%20er%20innovasjon%20-%20modell-definisjon-begrep%20-inst-lederm%C3%B8te%20051217.pdf&filename=Hva%20er%20innovasjon%20-%20modell-definisjon-begrep%20-inst-lederm%C3%B8te%20051217.pdf> (Hentet: 6. juni 2024).
- Aung, Y., Y.M., Wong, D., CS., og Ting, D., SW. (2021) The promise of artificial intelligence: A review of the opportunities and challenges of artificial intelligence in healthcare, *British Medical Bulletin*, volum 139 (1). Doi: <https://doi.org/10.1093/bmb/ldab016>
- Bareis, J., og Katzenbach, C. (2022) Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics, *Sage Journals*, vol. 47 (5), s. 855-881. Doi: <https://doi.org/10.1177/01622439211030007>
- Barua, I. (2023) *Kunstig intelligens redder liv – AI er legenes nye superkrefter*. 1. utgave. Oslo: Cappelen Damm AS.
- Bender, E., M., McMillan-Major, A., Gebru, T., og Schmitz, S. (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Conference on Fairness, Accountability, and Transparency (FAccT'21)*, 3.-10. Mars, 2021, Virtual Event, Canada. ACM, New York, USA. Side 610-623. Doi: <https://doi.org/10.1145/3442188.3445922>
- Bijker, W., og Pinch, T. (1987) The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. I W. Bijker, T. Hughes, og T. Pinch (red.), *The social construction of technological systems*, (s. 17-50). The MIT Press. Tilgjengelig fra: https://bibsys-almaprimo.hosted.exlibrisgroup.com/permalink/f/11qff65/BIBSYS_ILS715240392800022_01 (Hentet: 12. juni 2024).
- Bilotta, I., Corrington, A., Mandoza, S., A., Watson, I., og King, E. (2019) How Subtle Bias Infects the Law, *Annual Review of Law and Social Science*, vol. 15, s. 227-245. Doi: <https://doi.org/10.1146/annurev-lawsocsci-101518-042602>
- Celi, L., A., Cellini, J., Charpignon, M-L., Dee, EC., Derroncourt, F., Eber, R., et al. (2022) Sources of Bias in Artificial Intelligence that Perpetuate Healthcare Disparities – A Global Review, *PLOS Digital Health*, vol. 1 (3), s. 1-19. Doi: <https://doi.org/10.1371/journal.pdig.0000022>
- Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., og Atanasova, K., T. (2019) Artificial intelligence, bias and clinical safety, *BMJ Quality & Safety*, vol. 28 (3), s. 231-237. Doi: <https://doi.org/10.1136/bmjqs-2018-008370>
- Chomutare, T., Tejedor, M., Svenning, T., O., Ruiz., L., M., Tayefi, M., Lind, K., Godtliebsen, F., Moen, A., Ismail, L., Makhlysheva, A., og Ngo., P., D. (2022) Artificial

Intelligence Implementation in Healthcare: A Theory-Based Scoping Review of Barriers and Facilitators, *International Journal of Environmental Research and Public Health*, vol. 19 (23), s. 1-18. Doi: <https://doi.org/10.3390/ijerph192316359>

Das, S., og Das, A., K. (2021) Prototype-Driven Innovation: Propositions Based on Challenges and Opportunities. In: Chakrabarti, A., Poovaiah, R., Bokil, P., Kant, V. (eds) *Design for Tomorrow – Volume 2. Smart Innovation, Systems and Technologies*, vol. 222. Springer, Singapore. S. 981-992. Doi: https://doi.org/10.1007/978-981-16-0119-4_79

Dastin, J. (2018) Insight – Amazon scraps secret AI recruiting tool that showed bias against women, *Reuters*. Tilgjengelig fra: <https://www.reuters.com/article/idUSKCN1MK0AG/> (Hentet: 5. Mai 2024).

Ergen, M. (2019) What is Artificial Intelligence? Technical Considerations And Future Perception, *Anatol J. Cardiol*, s. 5-7. Doi: DOI:10.14744/AnatolJCardiol.2019.79091

Finansavisen.no (2024) Regjeringen vil ha 80 prosent AI innen 2025, *Finansavisen*. ... Tilgjengelig fra: https://www.finansavisen.no/nyheter/2024/04/16/8119577/regjeringen-vil-ha-80-prosent-ai-innen-2025?zephyr_sso_ott=505FUP (Hentet: 22. Mai 2024)

Grandaunet, B. (2023) Helseplattformen: Erfaringer fra en superbruker, *Dagens Medisin*. Tilgjengelig fra: <https://www.dagensmedisin.no/helseplattformen-erfaringer-fra-en-superbruker/597372> (Hentet: 9. mai 2024).

Helsedirektoratet.no (2022) *Kunstig intelligens i helsetjenesten – Status og veien videre for det nasjonale koordineringsarbeidet*. Tilgjengelig fra: https://www.helsedirektoratet.no/tema/kunstig-intelligens/Kunstig%20intelligens%20i%20helsetjenesten%20-%202022.pdf/_attachment/inline/7290856c-450a-47d6-aaa7-d660c205d4d4:f6c3e75605639f3d2c4750f2a6d8196a6e1a40d1/Kunstig%20intelligens%20i%20helsetjenesten%20-%202022.pdf (Hentet: 25. oktober 2023).

Innovasjon (2024) Tilgjengelig fra: <https://www.helse-midt.no/innovasjon#prioriterte-omr%C3%A5der> (Hentet: 6. juni 2024).

Iterasjon (2018) Tilgjengelig fra: <https://snl.no/iterasjon> (Hentet: 11. april 2024).

Jasanoff, S (2015). Future Imperfect: Science, Technology, and the Imaginations of Modernity. I S. Jasanoff & S.-H. Kim (red.), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power* (s. 1-33). University of Chicago Press. Tilgjengelig fra: https://bibsys-almaprimo.hosted.exlibrisgroup.com/permalink/f/11qff65/BIBSYS_ILS715369091700022_01 (Hentet: 12. mai 2024).

Kahn, K., B. (2018) Understanding innovation, *ScienceDirect*, vol. 61 (3), s. 453-460. Doi: <https://doi.org/10.1016/j.bushor.2018.01.011>

Kalantari, L., S., Zhang, H., McDermott, M., B., A., Chen, I., Y., og Ghassemi, M. (2021) Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations, *Nature Medicine*, vol. 27, s. 2176 – 2182. Doi: <https://doi.org/10.1038/s41591-021-01595-0>

Kelly, T. (2001) Prototyping is the Shorthand of Design, Prototyping is the Shorthand of Innovation, *Design Management Journal*, vol. 12(3), side 34-42. Doi: <https://doi.org/10.1111/j.1948-7169.2001.tb00551.x>

Kommunal- og moderniseringsdepartementet (2020) *Nasjonal Strategi for Kunstig Intelligens*. Tilgjengelig fra: <https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/no/pdfs/ki-strategi.pdf> (Hentet: 10. mai 2024)

Larrazabal, A., J., Nieto, N., Peterson, V., Milone, D., H., Ferrante, E. (2020) Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis, *PNAS*, Vol. 117 (23), s. 12 592 – 12 594. Doi: <https://doi.org/10.1073/pnas.1919012117>

Linstad, L., Silsand, L., Kristiansen, T., T., Næss, S., N., Ekeland, A., G. (2021) Kunstig intelligens i helsetjenesten – en stille revolusjon? *Dagens Medisin*. Tilgjengelig fra: <https://www.dagensmedisin.no/debatt-og-kronikk/kunstig-intelligens-i-helsetjenesten-en-stille-revolusjon/371793> (Hentet: 18. januar 2024).

Mutasa, S., Sun, S., og Ha, R. (2020) Understanding artificial intelligence based on radiology studies: What is overfitting? *Clinical Imaging*, vol 65, s. 96-99. Doi: <https://doi.org/10.1016/j.clinimag.2020.04.025>

Nasjonal strategi for kunstig intelligens (u.å.) Tilgjengelig fra: <https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/?ch=1> (Hentet: 10. mai 2024).

Nordanger, K, og Ludvigsen, E. (2023) Kunstig intelligens: Kan ha svaret på vår tids største helsetrusler, *TV2.no*. Tilgjengelig fra: <https://www.tv2.no/nyheter/innenriks/kan-ha-svaret-pa-var-tids-storste-helsetrusler/16223310/> (Hentet: 18. januar 2024).

NOU 2023: 4 (2023) *Tid for handling – Personellet i en bærekraftig helse- og omsorgstjeneste*. Oslo: Departementenes sikkerhets- og serviceorganisasjon Teknisk redaksjon. Tilgjengelig fra: <https://www.regjeringen.no/contentassets/337fef958f2148bebd326f0749a1213d/no/pdfs/nou202320230004000dddpdfs.pdf> (Hentet: 11. mai 2024).

Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S. (2019) Dissecting racial bias in an algorithm used to manage the health of populations, *Science*, vol. 366 (6464), s. 447-453. Doi: [DOI: 10.1126/science.aax2342](https://doi.org/10.1126/science.aax2342)

Panch, T., Mattie, H., og Atun, R. (2019) Artificial intelligence and algorithmic bias: Implications for health systems, *National Library of Medicine*, vol. 9 (2), s. 1-5. Doi: [10.7189/jogh.09.020318](https://doi.org/10.7189/jogh.09.020318)

Park, C., W., Seo, S., W., Kang, N., et al. (2020) Artificial Intelligence in Health Care: Current Applications and Issues, *Journal of Korean Medical Science*, vol. 35 (42), s. 1-11. Doi: <https://doi.org/10.3346/jkms.2020.35.e379>

Pukstad, B., S. (2018) Innovasjon – et viktig samfunnsoppdrag for alle som arbeider med helse, *NTNUhelse*, 23. november 2018. Tilgjengelig fra: <https://blog.medisin.ntnu.no/innovasjon-et-viktig-samfunnsoppdrag-for-alle-som-arbeider-med-helse/> (Hentet: 6. juni 2024).

Regjeringen med milliardatsing på kunstig intelligens (2023) Tilgjengelig fra: <https://www.regjeringen.no/no/aktuelt/regjeringen-med-milliardsatsing-pa-kunstig-intelligens/id2993214/> (Hentet: 22. mai 2024).

Roselli, D., Matthews, J., og Talagala, N. (2019) Managing Bias in AI, *ACM Digital Library*, s. 539-544. Doi: <https://doi.org/10.1145/3308560.3317590>

Sartori, L., og Bocca, G. (2022) Minding the gap(s): Public perceptions of AI and socio-technical imaginaries, *Springer Link*, vol. 38, s. 443-458. Doi: <https://doi.org/10.1007/s00146-022-01422-1>

Skjølsvold, T., M. (2017) *Vitenskap, teknologi og samfunn – En introduksjon til STS*. 1. utgave. Cappelen Damm AS.

Strümke, I. (2023) *Maskiner som tenker – Algoritmenes hemmeligheter og veien til kunstig intelligens*. 4. opplag. Oslo: Kagge Forlag AS.

Suchman, L. (2023) The uncontroversial 'thingness' of AI, *Sage Journals*, s. 1-5. Doi: <https://doi.org/10.1177/20539517231206794>

Sunarti, S., Rahman, F., F., Naufal, M., Risky, M., Febriyanto. K., Masnina, R. (2020) Artificial Intelligence in Healthcare: Opportunities and Risk for Future, *Gaceta Sanitaria*, vol. 35 (1), s. 67-70. Doi: <https://doi.org/10.1016/j.gaceta.2020.12.019>

Sørensen, K., H. (1999) Disiplinering og tverrfaglighet, i Knutsen, T., L. (red.) *Virkelighet og vitenskap – Perspektiver på kultur, samfunn, natur og teknologi*. 2. opplag. Ad Notam Gyldendal AS, s. 116-129.

Sørensen, K., H. (2008). Dialog og tverrfaglighet som kunnskapspolitikk. I Sørensen, Gansmo, Lagesen og Amdahl (red.) *Vitenskap som dialog – kunnskap i bevegelse*. Tapir akademiske forlag. S. 9-25

Thagaard, T. (2018). *Systematikk og innlevelse – en innføring i kvalitative metoder*. 5. utgave. Vigmostad og Bjørke AS.

Tjora, A. (2021). *Kvalitative forskningsmetoder i praksis*. 4. utgave. Oslo: Gyldendal Norsk Forlag AS.

Vestal, A., Mesmer-Magnus, J. (2020) Interdisciplinarity and Team Innovation: The Role of Team Experiential and Relational Resources, *Sage Journals*, vol. 51 (6), s. 738-775. Doi: <https://doi.org/10.1177/1046496420928405>

Yin, J., Ngiam, K., Y., og Teo, H., H. (2021) Role of Artificial Intelligence Applications in Real-Life Clinical Practice: Systematic Review, *Journal of Medical Internet Research*, vol. 23 (4), s. 1-17. Doi: <https://doi.org/10.2196/25759>

