



Handelshøyskolen BI - campus Bergen

BTH 36201

Bacheloroppgave - Økonomi og administrasjon

Bacheloroppgave

Ansvarsplassering når en kunstig intelligens feiler

Navn: Vibeke Engeskaug, Andrea Josefine
Fiskum Graue

Utlevering: 06.01.2020 09.00

Innlevering: 03.06.2020 12.00

Ansvars plassering når en kunstig intelligens feiler



Bilde hentet fra <https://www.mime.asia/here-are-3-roles-of-artificial-intelligence-in-healthcare/>

Bacheloroppgave

Eksamenskode: BTH 3620
6 semester

Handelshøyskolen BI Bergen

Utleveringsdato: 06.01.2020
Innleveringsdato: 03.06.2020

“Denne oppgaven er gjennomført som en del av studiet ved Handelshøyskolen BI. Dette innebærer ikke at Handelshøyskolen BI går god for de metoder som er anvendt, de resultater som er fremkommet, eller de konklusjoner som er trukket”

Forord

Denne oppgaven er skrevet i forbindelse med avsluttende bachelorutdanning i studiet Økonomi og Administrasjon 2018-2020 ved Handelshøyskolen BI, campus Bergen. Oppgaven er skrevet og utført av to studenter, for vår veileder Mads Nordmo Arnestad.

Kunstig intelligens er et tema som skaper ekstreme muligheter, men også utfordringer i arbeidslivet. Hvor ansvar plasseres når kunstig intelligens svikter, er et svært aktuelt og realistisk tema. For oss vekket problemstilling stor interesse og engasjement. Det har derfor vært meget interessant og givende å få muligheten til å lære mer og fordype seg i temaet. Prosessen med å ferdigstille oppgaven har vært både tidskrevende og utfordrende, men har gitt oss et resultat og en erfaring vi ikke ville vært foruten.

Vi ønsker å takke Handelshøyskolen BI for tre innholdsrike og flotte år. Vi ønsker også takke alle deltakerne som tok del i eksperimentet.

Vi ønsker tilslutt å rette en stor takk til Mads Nordmo Arnestad for eksepsjonell veiledning og godt samarbeid. Tusen takk for muligheten og tilliten du har vist oss; vi vil uttrykke stor takknemmelighet for at vi fikk skrive et så solid og gjennomført eksperiment.

Bergen, juni 2020.

Sammendrag

I denne bacheloroppgaven skal vi undersøke temaene moralsk ansvar og kunstig intelligens. Forskningsoppgaven kaster lys over de ulike aktørene, rollene og deres ansvar. Hensikten med oppgaven er å bringe mer klarhet i problemstillingen:

Hvor plasseres ansvar når en AI gjør feil?

Vi har utført en kvantitativ undersøkelse med 158 respondenter, hvorav 121 av disse gjennomførte hele eksperimentet. Respondentene ble innledningsvis informert gjennom en randomisert vignett, der radiolog Sigurd enten hadde et valg om å utføre oppgaven selv eller ikke. Det kom tydelig frem at den kunstige roboten var 300 ganger sikrere enn Sigurd, og det riktige var derfor å la roboten utføre jobben. I undersøkelsen ble det kartlagt holdninger rundt kritikkverdighet, moralsk ansvar, straff, sinne og erstatningsansvar.

Oppgaven tar for seg eksisterende litteratur om kunstig intelligens og moralsk ansvar, i tillegg til relevant forskning. På bakgrunn av relevant teori utarbeidet vi syv hypoteser.

Vi konkluderte med at både sykehuset og produsenten er tjent med å ha på en *manual mode knapp* som kan overstyre roboten. Resultatene forteller at så lenge knappen eksisterer, endrer dette folks oppfatning av hvor ansvarlig produsenten fremstår overfor eventuelle feil. Det er viktig å poengtere at en lege ikke står fullstendig ansvarlig når vedkommende ikke er informert om ansvaret som medfølger av valget. Oppgavens konklusjon beskriver viktigheten av å informere ansatte, samt å opprettholde konkrete ruter vedrørende delansvaret ved bruk av kunstig intelligens. Samtidig forklarer resultatene hvorfor det eksisterer en manual mode knapp; produsent og sykehus er tjent med det.

Liste over tabeller og figurer

Figur 1. Eksperimentets design.....	24
Figur 2. Inntekt i løpet av et år.....	27
Figur 3. Privat eller offentlig sektor.....	27
Figur 4. Høyeste fullført utdanning.....	28
Figur 5. Er du leder på din arbeidsplass?	29
Tabell 1. Cronbachs Alpha.....	30
Tabell 2. Korrelasjonsmatrise.....	32

Innholdsfortegnelse

FORORD	2
SAMMENDRAG	3
LISTE OVER TABELLER OG FIGURER	4
INNLEDNING	7
1.0 TEORI	7
1.1 ARITIFICIAL INTELLIGENCE - KUNSTIG INTELLIGENS	7
1.2 KUNSTIG INTELLIGENS I HELSESEKTOREN.....	9
1.2.1 IBM-Watson Clinical Trial Matching.....	11
1.2.2 Norges helseplan 2020	12
1.3 ALGORITME-AVERSJON	12
1.4 MORAL OG MORALSK ANSVAR.....	14
1.4.1 Moralsk betraktning	14
1.5 ANSVARSFORDELING VED VALG	17
1.6 HYPOTESETEST.....	19
2.0 METODE	20
2.1 VALIDITET OG RELIABILITET.....	20
2.1.2 Validitet	20
2.2 FORSKINGSDESIGN	21
2.2.1 Kvalitativ vs. kvantitativ	21
2.2.2 Sekundærdata vs. Primærdata.....	22
2.3 UTVALGET.....	23
2.3.1 Invitasjon til undersøkelsen	23
2.4. EKSPERIMENTETS DESIGN	24
2.5 UTFALLSMÅL	25
3.0 RESULTATER	26
3.1 DESKRIPTIVE DATA	26
3.1.1 Alder og kjønn	26
3.1.2 Inntekt.....	26
3.1.3 Sektor.....	27
3.1.4 Høyest fullførte utdanning	28
3.1.5 Lederfordeling	28
3.1.6 Kritisk evaluering av data	29

3.2 KONSTRUKSJON AV MÅLEMODELL.....	30
3.2.1 Cronbachs alfa	30
3.2.2 Korrelasjonsanalyse	31
3.3 TEST AV MELLOMGRUPPEFORSKJELLER: HYPOTSETESTING	32
4.0 DISKUSJON	35
4.1 IMPLIKASJON	36
4.1.1 Implikasjoner for ledelse	38
4.2 BEGRENSNING OG ANBEFALINGER FOR FREMTIDIG FORSKNING	38
5.0 KONKLUSJON.....	40
6.0 KILDELISTE.....	41
7.0 VEDLEGG	45
7.1 VEDLEGG – SPØRSMÅL TIL UNDERSØKELSEN	45
7.2 VEDLEGG – HYPOTSETEST	46

Innledning

I dagens samfunn ser vi stadig flere yrker som tar i bruk kunstig intelligens gjennom digitaliserte hjelpere, for å automatisere arbeidsoppgaver. I banksektoren brukes for eksempel roboter som chatter med besøkende i nettbank, bilindustrien utvikler selvkjørt biler, og helsesektoren bruker treffsikre roboter til å utføre arbeidsoppgaver. Når en lege utfører et inngrep eller analyserer et radiologisk bilde feil, er det sykehuset som står ansvarlig. En ny problemstilling er hvem som holdes ansvarlig når en leges arbeidsoppgaver er erstattet av en robot som utfører en fatal feil. Er det legen og sykehuset som står ansvarlig, selskapet som har trent opp roboten, eller ligger ansvaret på selskapet som har utviklet roboten?

Vi skal i denne oppgaven undersøke og gjøre rede for om mennesker har en tendens til å plassere ansvar annerledes, dersom et utfall kunne vært overstyrt av et valg. Derfor er vår valgte problemstilling:

Hvor plasseres ansvaret når en AI gjør feil?

Bruken av kunstig intelligens er med på å skape etiske diskusjoner rundt moralsk ansvar. Formålet med oppgaven er nettopp å undersøke hvor ansvar plasseres når en robot gjør feil, dersom brukeren hadde et valg om å utføre arbeidsoppgaven selv. For å besvare problemstillingen skal vi bruke teori og forskning innenfor kunstig intelligens og moralsk ansvar, samt utføre en kvantitativ undersøkelse.

1.0 Teori

1.1 Artificial Intelligence - Kunstig intelligens

Kunstig intelligens, eller Artificial Intelligence (AI), er en konstellasjon av informasjonsteknologi som lar kompetente maskiner utvide menneskelige evner ved å sanse, forstå, handle og lære. Informasjonsteknologi justerer sin egen aktivitet og fremstår derfor tilsynelatende som intelligent. Dette for at mennesker skal kunne oppnå mer; både presisjon og effektivitet (Tidemann A., Kunstig Intelligens, 2020). Som fagfelt er kunstig intelligens en

sammensmelting av datateknikk, logikk, matematikk, psykologi og nevrovitenskap. Disse teknologiene inkluderer blant annet naturlig språkprosessering, maskinlæring, overvåket læring, intelligente agenter, ekspertsystemer, selvkjørende biler samt stemmegjenkjennelse (...).

Suksessformelen til kunstig intelligens er basert på to faktorer; bias og noise (Kahneman, Rosenfield, Gandhi, & Blaser, 2016). *Bias* gjør at algoritmen til roboten kun tar vurderinger med utgangspunkt i den kvantifiserte informasjonen. Det vil si at det er like uproblematisk å ekskludere en irrelevant faktor, som det er å inkludere den samme faktoren. Dette kan for eksempel være nasjonalitet eller utseende, for ikke å glemme dagsform eller humør. I et jobbintervju lar vi oss påvirke av både relevante og irrelevante inntrykk og observasjoner. Intervjueren har ofte sett for seg en karakter i forkant, og underveis i intervjuet vil felles interesser samt personlig synspunkter spille inn.

Den siste suksessfaktoren til kunstig intelligens er *noise*, også kalt støy. Denne faktoren forklarer at enkle algoritmer har mindre grad av tilfeldig variasjon, enn menneskelige beslutninger. Det vil si at en statistisk algoritme vil komme frem til samme svar hver eneste gang, gitt tilgjengelig informasjon og vurderingskriteriene (Sjåstad, 2019, s. 65). I et av studiene til Daniel Kahneman fremkommer det at erfarne radiologer som vurderer nøyaktig det samme røntgenbildet, på to forskjellige tidspunkt, gav forskjellig vurdering opptil 20% av tilfellene (Kahneman, 2011). Den radikale løsningen på støyproblemet er å erstatte menneskelige vurderinger med algoritmer som bruker datainnsamlingen til å produsere en prediksjon eller en beslutning.

Formålet med teknologien innen kunstig intelligens, er å utvikle, støtte og automatisere datasystemer som mennesker tidligere ikke har klart å løse med verken sammen hastighet eller treffsikkerhet. Dette kalles kognitive teknologier (Rowe, 2020). Kognitiv databehandling er i stand til å prosessere ustrukturert data, som for eksempel journaler, video eller bilder for så å identifisere og bryte ned dataene. Deretter defineres sammenhenger som beskriver de ustrukturerte dataene slik at dataene kan brukes videre i behandlingen (Vredenberg, 2015). De viktigste fordelene med kognitive

datasystemer er at den forstår naturlig språk, dens evne til å bygge opp sammenhenger i enorme mengde ustrukturert data og evnen til å lære av erfaring på lik linje som oss mennesker.

Computer vision og maskinlæring er to eksempler på kognitive teknologier. Computer vision handler om å lære en maskin til å tolke visuell informasjon og identifisere objekter (Dustin , 2018). ”Maskinlæring er en spesialisering hvor man bruker statistiske metoder for å la datamaskiner finne mønstre i store datamengder” (Tidemann & Elster, SNL, 07). Denne formen for kunstig intelligens har flere bruksområder, deriblant automatisk analyse av røntgenbilder i helsetjenesten.

Forskningen på dataalgoritmers evne til å utkonkurrere subjektive menneskelige vurderinger er ikke ny. Vi må helt tilbake til 1950-tallet for å finne de første studiene innenfor kunstig intelligens. To av forskerne innenfor fagfeltet var Meehl og Dawes. De studerte sammenlikningen mellom statistisk prediksjon (statistisk vektning og kombinerer av data) og klinisk prediksjon (data kombineres skjønnsmessig) (Sjåstad, 2019, s. 64). Klinisk prediksjonen var basert på å stille spørsmål om behandlingsutfall til legeeksperter, hvor de kunne bruke tilgjengelig informasjon for å besvare spørsmålene. Dette var empiriske spørsmål som for eksempel; har pasienten lungeskade? Når vil denne pasienten bli frisk? Samtidig brukte Meehl enkle regresjonsmodeller basert på samme informasjonsgrunnlag. Svaret på spørsmålene og dermed resultatet av analysen var entydig; vurderinger foretatt av maskiner med regresjonsmodeller er alltid like nøyaktige, og ofte mer nøyaktige, enn menneskelige legeeksperter (Meehl, 1954; Dawes, 1979).

1.2 Kunstig intelligens i Helsesektoren

I helsesektoren har forskning og bruk av medisinsk kunstig intelligens økt kraftig siden Meehls studie for rundt 60 år siden. Selv om mange fortsatt har tiltro til at en erfaren lege skal utføre jobben, viser det seg at intelligente maskiner og AI-algoritmer har større treffsikkerhet og gjør mindre feil (Sjåstad, 2019, s. 65). Et viktig poeng er at teknologien garanterer pasientene

lik kvalitet på behandlingen, uavhengig av kompetansen til legen eller geografisk plassering.

I løpet av de 10 siste årene er det kommet en rekke forskningsartikler som forklarer at ved hjelp av medisinsk kunstig intelligens, blir arbeidsoppgaver utført like bra, om ikke bedre enn med menneskelig arbeidskraft (Sjåstad, 2019). I artiklene fremkommer det at ved identifisering av hjertesykdom (Hutson, 2017) og øyesykdommer (Gulshan, Peng, & Coram, 2016), viser det seg at medisinsk kunstig intelligens har like stor treffsikkerhet som menneskelig ekspertvurderinger. Trolig kommer kunstig intelligens til å revolusjonere effektiviteten og treffsikkerheten i helsevesenet i fremtiden.

En utfordring i helsesektoren er at klinikkene preges av mye kortere behandlingstid per pasient, og eksperter med kompetanse ikke alltid er tilgjengelig (Johansen, 2019). Dette bærer med seg en risiko i form av at legen ikke er oppdatert på siste studie i møte med pasienten. For å skape orden i dataene blir det derfor innført teknologi som verktøy; på fagspråket kalt interoperabilitet *«evnen til å kjøre programmer, eller overføre data mellom ulike funksjonelle enheter. Dette på en slik måte at brukeren ikke trenger spesiell kunnskap om disse enhetenes karakteristikk»* (Open Geospatial Consortium). Interoperabilitet skal med dette bidra til at prosesser fungerer og rett informasjon er tilgjengelig, på tvers av virksomheter og omsorgsnivåer (Norwegian Specialist Healthcare, 2019).

Ved å bruke denne unike teknologien vil legen ha tilgang på den mest oppdaterte kliniske studie relevant for sin pasient, som gir et optimalt behandlingsgrunnlag for pasienten. Sørlandet sykehus forsker på bruk av ulike dataprogrammer som henter relevant pasientinformasjon i løpet av få sekunder (Berge, Granmo, & Tveit, 2017). Selskapet IBM jobber med å utvikle en maskin med AI-algoritmer som skal matche kreftpasienter mot riktig diagnose og behandling. I begge tilfellene brukes interoperabilitet sammen med kunstig intelligens for å skape et raskere og mer treffsikkert beslutningsverktøy.

1.2.1 IBM-Watson Clinical Trial Matching

IBM- Watson har et program som heter *Clinical Trial Matching* som går ut på å matche kreftpasienter mot kliniske studier. Dette har som formål å gi en tydelig diagnose, og ikke minst være et stødig behandlingsverktøy for legen. Dette er et typisk eksempel på en prosess som tar svært lang tid i en vanlig praksis, men som de fleste kreftpasienter etterspør da det gir pasienten en tydelig prognose.

Clinical Trial Matching vil i løpet av noen sekunder 1) ta utgangspunkt i pasienten og vise alle studier i nærområdet eller region som pasienten kan inkluderes i, eller 2) ta utgangspunkt i studie og finne pasienter som kan matche studiets spesifikke kriterier. Dette er en ekstrem forenkling av prosessene. I en rapport fra Verdens Helseorganisasjon kommer det frem at mellom 30 og 50 prosent av dødsfall forårsaket av kreft, kan unngås med forebygging, samt tidlig identifisering ved for eksempel bruk av Watson sitt program (Lein-Mathisen & Lundgreen, 2019).

University of North Carolina School of Medicine utførte en forskningsstudie ved bruk av IBM-Watson maskin på tvers av 1000 kreftdiagnoser. Resultatet fra studiet viste svært store forskjeller; Watson-maskinen avdekket behandlingsalternativer som legeekspertene gikk glipp av i 30 prosent av tilfellene (Lohr, 2016). Enkelte av behandlingene var basert på nye kliniske studier som legene ikke hadde kjennskap til. Watson-maskinen sin AI-algoritme leste alt.

En annen studie som underbygger IBM-Watson undersøkelsen er vurderingen av prioritert behandling med utgangspunkt i pasientens alvorlighetsgrad. Studiet viser at ved bruk av AI var vurderingen korrekt i 90,2 prosent av tilfellene, sammenliknet med legene som gjorde en korrekt vurdering i 77,5 prosent av tilfellene (Donnelly , 2017). Begge studiene er verdifulle for å forbedre pasientomsorgen da den høye reliabiliteten til AI-logaritmene gjør dem mer treffsikre enn fagfolk alene.

1.2.2 Norges helseplan 2020

I januar 2020 la regjeringen frem en nasjonal strategi for kunstig intelligens, hvor de fremkommer at Norge skal satse på kunstig intelligens innen blant annet helse (moderniseringsdepartementet, 2020). Regjeringens *Nasjonal helse- og sykehusplan 2020-2030* er det fremlagt at Oslo Universitetssykehus har som mål å automatisere analyser av patologibilder av kreftsvulster gjennom prosjektet *DoMore* (Helse, 2019) Vi har tidligere sett at *IBM-Watson* skal hjelpe leger med på å matche kreftpasienter med rett diagnose og behandling. Likeså er *DoMore* basert på kreftpasienter, men dette prosjektet vektlegger Computer vision; bildeanalyse og kvantifisering av pasientens DNA.

DoMore bruker informatikk og kunstig intelligens som metode for å finne kreftsvulster og studere genetiske forandringer under behandlingsperioden. Professor ved Oslo Universitetssykehus Håvard Danielsen, forteller at kreftsvulster er heterogene, som vil si at svulsten kan bestå av områder med ulike avvik og inneholde mye mer informasjon enn det leger har mulighet til å håndtere (Andresen, 2020). Professoren forteller at legenes vurdering av alvorlighetsgraden av kreftsvulster bare er riktig i om lag 60% av tilfellene (Andresen, How AI changes the prognostication of cancer, 2019).

1.3 Algoritme-aversjon

Selv om studiet til Meehl tilbake på 1950-tallet og IBM Watson sitt Clinical Trial Matching program påpeker det faktum at algoritmer er mer treffsikre enn mennesker, er det likevel flere aktuelle tilfeller hvor mennesker ikke benytter seg av algoritmen. Dette kalles algoritme-aversjon. Algoritme-aversjon beskrives som tendensen til å foretrekke en menneskelig vurdering fremfor avansert teknologi innenfor kunstig intelligens (Sjåstad, 2019, s. 63).

Særlig innenfor helsesektoren er algoritme-aversjon høy, da sårbare og omsorgskrevende pasienter står i sentrum. I etiske diskusjoner rundt kunstig intelligens blir det ofte hevdet at mennesker ser på roboter som trussel mot menneskeheten. Tematikken har lagt grunnlaget for science fiction-filmene *Terminator (1984)* og *The Matrix (1999)*. Filmene var med på å skape en offentlig bevissthet rundt en fremtid med automatiserte arbeidsoppgaver. Det

er nærliggende å tro at en rekke yrkesgrupper automatiseres bort som en følge av teknologien. Dette kan være med på å effektivisere samfunnet, men byr selvsagt på sine egne utfordringer.

Dietvorst, Simmons og Massey (2014) kastet lys over algoritme-aversjon problematikken. Deres studie gikk i korte trekk ut på å finne ut om testdeltakerne valgte å satse penger på menneskelig prediksjon eller prediksjonen til en statistisk algoritme. Det er viktig å få frem at testdeltakerne også fikk oppleve at algoritmen gjorde feil. Det mest interessante resultatet fra eksperimentet var at testdeltakerne fortsatt valgte menneskelige prediksjonen, selv om de anerkjente at algoritmen var mer treffsikker i gjennomsnitt (Dietvorst, Simmons, & Massey, 2014). Studie begrunnet det faktum at mennesker er mindre tilgivende overfor feil utført av en kunstig agent, enn tilsvarende feil hos et menneske.

Aversjon mot algoritmer er kostbar, ikke bare for testdeltakerne i studiet til Dietvorst (2014) som tapte penger på å stole på menneskelig utførelse fremfor algoritmer, men også for samfunnet for øvrig. Det er mange avgjørelser som krever en prognose, og algoritmer er i de fleste tilfeller bedre predikanter enn mennesker (Meehl, 1954; Dawes, 1979).

Uansett hvilken type algoritme som brukes er det viktig at mennesker overvåker beslutningene til algoritmen. Algoritmen må også justeres for sporadiske endringer i populasjonen av tilfellene. Samspillet mellom mennesker og algoritmen er avgjørende for suksess. Ledere må følge opp enkeltvedtak og ha myndighet til å overstyre algoritmen i tydelige beslutninger. Dette kan for eksempel være en beslutning om å godkjenne et lån, hvor det viser seg at låntakeren i etterkant har fått dramatiske betalingsanmerkninger. I et slikt tilfelle kan lånet imidlertid tilbakeføres, selv om algoritmen i forkant har godkjent lånet. I dette eksempelet er sunn skepsis og ansvarlighet viktig for at teknologien blir brukt på en etisk forsvarlig og fornuftig måte.

1.4 Moral og moralsk ansvar

Moral er internaliserte og institusjonaliserte oppfatninger av hvordan vi bør handle (Svendsen, 2018). Ansvar eller forpliktelse handler om å stå til rette for, gjøre rede for eller bære utgiftene for noe (Tranøy, 2018). Ved å være moralsk ansvarlig for en handling, sies det at man er verdig en bestemt type reaksjon for å ha utført handlingen; som ros, kritikk eller skyld (Eshleman, 2014).

Refleksjon rundt moralsk ansvar har en lang historie, og vekker generelt stor interesse. En årsak til den vedvarende interessen kan være at temaet virker tilkoblede med vår oppfatning av oss selv som person og enkeltmenneske. Hvilket vil si at våre særtrekk, men også generelle personlighetstrekk, reflekteres gjennom våre moralske, ansvarlige handlinger (Feltz & Cokely, 2009) (Mele & Cushman, 2008).

Vi skiller mellom deskriptiv og normativ betydning av moral. I den deskriptive betydningen henviser moral til personlige eller kulturelle verdier, atferdsnormer eller sedvaner (Perry & Dzhamfarov, 2012). I den normative betydningen betegner moral det som er rett eller galt (...). Følgelig angir dette hvordan vi bør handle. Etikk og moral brukes ofte om hverandre. Det er vanlig å si at etikk betegner refleksjon over moralen, eller teorier som beskriver hva moral er (Svendsen, 2018).

Det forskes mye på etikk innen kunstig intelligens og hvilke reguleringer som skal gjelde på et globalt perspektiv (Tidemann A. , 2020). *Leverhulme Centre for the Future of Intelligence* er et institutt som forsker på etiske aspekter av anvendelse av kunstig intelligens (Leverhulme Centre for the Future of Intelligence, 2020). Dette fremtrer som en svært relevant fremtidsrettet problemstilling.

1.4.1 Moralsk betraktning

Mange av valgene som kunstig intelligens gjør, kan minne om bevisste valg når man observerer oppførselen. Dette er ikke tilfelle. Som presentert i avsnitt 1.1 viser AI-algoritmer intelligent oppførsel uten at den er bevisst, blant annet gjennom bias og noise. I politiske og etiske diskusjoner rundt kunstig

intelligens blir det ofte hevdet at kunstig intelligens kan være en trussel mot menneskeheten og de økonomiske systemene vi kjenner (Tegmark, 2017).

Kunstig intelligens har potensiale til å effektivisere samfunnet på mange vis (ScienceDirect, 2017). Dette utløser en mengde etiske spørsmål, som stiller krav til ansvarlighet dersom bruken av kunstig intelligens skader et annet menneske. *Norge er kjennetegnet av at det er høy tillit til hverandre, og til statlige og private virksomheter* (Kommunal og moderniseringsdepartementet, 2020). I regjeringens nasjonale strategirapport for kunstig intelligens (januar 2020) lyder det at kunstig intelligens som utvikles, forskes på og brukes i Norge skal bidra til ansvarlig og pålitelig kunstig intelligens, samt ivareta den enkeltes integritet og personvern. Her er det formet syv prinsipper for etisk og ansvarlig bruk av kunstig intelligens; herunder at løsninger basert på kunstig intelligens skal respektere menneskets selvbestemmelse og kontroll, samt være sikre og teknisk robuste (...).

Ett annet av prinsippene sier at systemer basert på kunstig intelligens må være gjennomsiktlige: *Beslutninger tatt av systemer basert på kunstig intelligens, skal være sporbare, forklarbare og gjennomsiktlige. Det betyr at man som enkeltperson eller juridisk person skal ha mulighet til å få innsikt i hvorfor en beslutning som gjelder dem ble som den ble. Sporbarhet muliggjør både revisjon og forklaring. Gjennomsiktighet oppnås blant annet ved å gi informasjon om behandlingen til den registrerte. Gjennomsiktighet handler også om at datasystemer ikke skal utgis for å være mennesker – mennesker skal ha rett til å få vite om de samhandler med et KI-system (...).* Med dette skal en enkeltperson, som en pasient, være klar over at det er en kunstig intelligens som handler, en skal ha mulighet til å spore beslutningstaker, samt kunne få en forklaring på hvorfor beslutningen ble som den ble.

Utvikling og bruk av kunstig intelligens er med på å skape utfordringer og reise flere vanskelige spørsmål (Kommunal og moderniseringsdepartementet, 2020). Vanligvis når noe går galt, er det en relativt grei sak å fastslå skyld. En feildiagnose vil sannsynligvis være den tilhørende leges ansvar og dermed også sykehuset sitt ansvar. I 2018 fikk for eksempel Helse Bergen en bot på 1,5

millioner etter et Djabrail Sulejmanov døde av feil medisinerings (Ruud, 2018). Ved feil i maskin eller medisinsk utstyr vil trolig produsent eller operatør bære deler av ansvaret. Spørsmålet er hva dette vil bety for en kunstig intelligens.

Det medisinske ansvar kan ved bruk av kunstig intelligens plasseres hos ulike parter; de som har utviklet teknologien, de som har solgt den, de som har trent den opp ved hjelp av data, eller de som faktisk tar den i bruk (Tidemann A. , 2020). En part av legene som bruker kunstig intelligens i dag, forventes å bruke det som et hjelpemiddel i klinisk beslutningstaking, og ikke nødvendigvis som erstatning for standard prosedyre (Yu & Kohane, 2018). Når dette er tilfelle vil legen fortsatt stå ansvarlig for feil som kan oppstå (Hart, 2017). Det er imidlertid uklart om leger faktisk vil kunne vurdere påliteligheten eller nytten av informasjon hentet ut av en kunstig intelligens, og om de kan ha en meningsfull forståelse av konsekvensene for disse handlingene (...).

Mangelen av pålitelighet oppstår som en bivirkning av hvordan maskinlæringsalgoritmer fungerer. Det er mulig å forstå hvorfor en kunstig intelligens har tatt avgjørelsen den har, basert på informasjonen den er matet med. Et problem som følger er at dataene som brukes til å trene algoritmene, er beskyttet av offentlige og utilgjengelige personvern hensyn. Personvernmuren vil sannsynligvis bli komplisert videre når leger blir nødt til å stole mer og mer på kunstig intelligens, og det blir mindre vanlig å fysisk utfordre algoritmens resultat (...).

Det er noe mer uklart hvor medisinsk ansvar ligger når legen bruker kunstig intelligens som standard prosedyre, fastslått av sykehuset. Som tidligere nevnt er det flere parter som kan pådras skylden for en beslutning gjort med kunstig intelligens. For å forstå konsekvensene ved anvendelse av kunstig intelligens i helsevesenet må vi ha en oppfatning av hvem som står ansvarlig når noe går galt. Å unnlate dette er med på å undergrave pasientens tillit, det plasserer leger i vanskelige posisjoner, kan feilaktig redusere tilliten til og anvendelsen av nyttig teknologi og derfor potensielt hindre lønnsomme og nyttige investeringer i feltet. Ufullstendige ansvarsmodeller er med på å forsinke både

utvikling og bruk av en teknologi som kan være med på å effektivisere og redde liv.

1.5 Ansvarsfordeling ved valg

Et grunnleggende moralsk ideal i vestlige samfunn er prinsippet om at mennesker skal holdes personlig ansvarlig for konsekvensene av valgene deres (Cappelen, Fest, Sørensen, & Tungodden, 2018).

I studiet gjort av Cappelen, Fest, Sørensen og Tungodden (2018), undersøkes det hvordan individer holdes personlig ansvarlig for valg. Professor i psykologi, Peter Vallentyne, har angitt to minimumskrav for at en person ikke skal kunne holdes personlig ansvarlig for resultatet av et valg. Det sier at 1) dersom personen ikke kunne ha endret sannsynligheten for utfallet ved å velge annerledes, eller 2) personen kunne bare ha unngått utfallet ved å velge et urimelig alternativ, skal vedkommende ikke kunne holdes ansvarlig (Vallentyne, 2008).

Vallentyne former også tre betingelser som må være oppfylt for at en person skal holdes ansvarlig for et utfall. Den første betingelsen handler om at personen må ta et passende autonomt valg, som vil si at valget må være rimelig responsivt og ikke manipulert av noen andre. Videre må det foreligge en årsakssammenheng, hvilket vil si at utfall må ha en passende årsakssammenheng med valget. Avsluttende foreligger det en «tro-tilstand», som vil si at personen må ha en passende eller rimelig tro på at utfallet er relatert (...). Studiet av Cappelen et. al undersøker ved hjelp av et eksperiment om deltakerne plasserer skyld på bakgrunn disse kriteriene og betingelsene, eller om det foreligger ulikheter i ansvars plassering.

Cappelen et. als eksperiment ble utført med tilfeldige roller; hvor utvalget først ble delt inn i to. Deretter ble hvert utvalg gitt en av tre tilfeldige roller. I den ene gruppen får spiller A og B et valg om de vil ha 10kr som de er garantert å få, eller om de ønsker å gamble på en 50/50 sjanse for å kunne vinne 100kr/0kr. På bakgrunn av den lave summen på 10kr velger 100% av eksperimentets deltakere å gamble. Enkelte vinner, og andre ikke. For eksempel vinner

deltaker B 100kr, og deltaker A vinner 0kr. Avslutningsvis får dommeren spørsmålet om han ønsker å gi litt av summen spiller B vant, og dele med spiller A som ikke vant. Dommeren fikk se nøyaktig hvilke instruksjoner spillerne ble gitt; og er dermed klar over at de selv valgte å satse pengene. I dette tilfellet av eksperimentet sier dommeren nei; han ønsker ikke å dele noen av pengene mellom spillerne som har tatt sine valg.

I den andre delen av utvalget blir spillerne presentert for samme problemstilling men med ulike instruksjoner; det foreligger ingen valg og de blir satt i en gamblingsituasjon hvor det er en 50/50 sjanse for å vinne 100kr/0kr. Avslutningsvis blir dommeren presentert med spørsmålet om han ønsker å gi litt av summen spiller B vant, og dele med spiller A som ikke vant. Samtlige av dommerne i dette utvalget ønsket å dele pengesummen. Dette på bakgrunn av at spillerne ikke valgte selv å gamble; men ble tvunget til det.

Dette viser at spillerne i den første gruppen blir holdt moralsk ansvarlig for sin handling da de valgte å gamble. Spillerne i den andre gruppen som ikke ble pålagt noen restriksjoner blir ikke holdt moralsk ansvarlig.

Studiet har funnet sterke bevis for at folk flest viser større aksept for flere utfall, dersom det i forkant forelå en mulighet om valg. Det vil si at det plasseres et større individuelt ansvar på personer som hadde mulighet til å påvirke situasjonen ved å foreta et valg. Dette skjer selv om valgalternativene ligger langt ifra hverandre, og det er helt klart at det ene valget er både bedre, og har høyere sannsynlighet for å lykkes, enn det andre. Det viser altså at deltakerne har lettere for å plassere skyld når det foreligger et valgalternativ.

Det å øke bevissthet rundt konsekvensene av våre valg, er med på å påvirke folks oppfatning av ansvars plassering og utfall (Greenfield, 2012). Professor i rettsvitenskap, Kent Greenfield, skriver i «The Myth of Choice» om hvordan våre samfunn er bygget på ideen om valg. Det faktum at både vårt politiske regime er basert på samtykke fra styringsmakter, og vårt rettssystem er bygget på at mennesker har frie valg, og dermed også bærer eget ansvar for dem, har en innvirkning på hvordan vi plasserer ansvar. Det viser seg at hvorvidt det

fremkommer at en person blir pålagt til å utføre en oppgave, eller gjør det av fri vilje, har en sterk påvirkning på om personen vil bli holdt personlig ansvarlig for utfallet. Greenfield viser med dette at elementet valg, er med på å spore, plassere og avgjøre personlig ansvar (...).

1.6 Hypotesetest

Vi har tidligere sett på teori som er relevant for å besvare vårt overordnede forskningsspørsmål, med hovedfokus på feltene kunstig intelligens og moralsk ansvar. På grunnlag av teori samt forskningsspørsmål har vi dermed kommet frem til syv hypoteser. Disse danner grunnlaget for å besvare hvor ansvar plasseres når en kunstig intelligens feiler.

H1: Sigurd sin fremferd blir vurdert som mer kritikkverdig når han hadde et valg.

H2: Sigurd vil holdes moralsk ansvarlig for utfallet når han hadde et valg.

H3: Det at Sigurd hadde et valg gjør at han i større grad fortjener straff enn om han ikke hadde hatt et valg..

H4: Sigurds fremferd vekket sinne.

H5: Sigurd blir sett på som mer erstatningsansvarlig hvis han hadde et valg.

H6: Sykehuset blir sett på som mer erstatningsansvarlig hvis Sigurd ikke hadde et valg.

H7: Produsenten iScan3000 blir sett på som mer erstatningsansvarlig hvis Sigurd ikke hadde et valg.

2.0 Metode

Samfunnsvitenskapelig metode dreier seg om hvordan vi går frem for å hente inn informasjon om virkeligheten, samt hvordan vi analyserer, bearbeider og tolker informasjonen. Dette er en sentral del av empirisk forskning (Johannessen, Tufte, & Christoffersen, 2016, s. 25). I denne delen av oppgaven skal vi redegjøre for fremgangsmåten for gjennomførelsen av eksperimentet, samt se kritisk på valgt forskningsdesign.

2.1 Validitet og reliabilitet

For å kunne vurdere kvalitet og troverdighet på innhentet data bruker vi begrepene reliabilitet og validitet.

2.1.1 Reliabilitet

Et grunnleggende spørsmål i all forskning er datas pålitelighet; på fagspråket kalt reliabilitet. Reliabilitet dreier seg om nøyaktigheten av undersøkelsens data, hvilke data som brukes, måten de er samlet inn på, hvordan de er bearbeidet, samt potensielle feilkilder (Johannessen, Tufte, & Christoffersen, 2016, s. 231). En mulighet for å teste datas reliabilitet på, er å utføre *test-retest-reliabilitet*. Det vil si å gjennomføre den samme undersøkelsen på samme gruppe ved to ulike tidspunkt. Oppnår man inn samme svar, er det et tegn på høy reliabilitet.

2.1.2 Validitet

Validitet dreier seg om hvor troverdig eller relevante data er. Det stilles da spørsmål om hvor godt, eller relevant data representerer fenomenet vi undersøker. Vi skiller mellom intern og ekstern validitet (Johannessen, Tufte, & Christoffersen, 2016, s. 232).

Intern validitet, stiller spørsmål om generalisering fra utvalg til populasjon, og reflekterer om utvalget er representativt ovenfor populasjonen. Innenfor kvantitative undersøkelser stilles spørsmålet «*måler vi det vi tror vi måler*»? Altså om det er det sammenheng mellom fenomenet som undersøkes, og dataene som er samlet inn. En trussel mot validitet er bortfall i populasjonen;

det vil si at det er færre med i nettoutvalget enn bruttoutvalget. Det faktum at undersøkelser har bortfall må redegjøres for når resultatet skal presenteres. Det må også drøftes i hvilken grad bortfallet svekker undersøkelsens statistiske validitet. Statisk validitet handler om graden feilfrihet, nøyaktighet og representativitet i resultatene.

Dersom forutsetningene for statisk validitet er oppfylt, er det relevant å ta for seg ekstern validitet. Ekstern validitet er spørsmål om i hvilken grad resultater fra en undersøkelse kan overføres i rom og tid. Det vil si i hvilken grad resultater fra et forskningsprosjekt kan overføres til liknende fenomener (...) (altså generaliseres fra et utvalg til for eksempel en befolkning).

2.2 Forskingsdesign

Forskingsdesign er en grov skisse av alt som knytter seg til en undersøkelse, inkludert gjennomføringen av studien (Johannessen, Tufte, & Christoffersen, 2016, s. 69). Forskningsdesignet er derfor svært viktig for å kunne besvare den overordnede problemstillingen. Vårt eksperiment bygger på kausalt forskningsdesign hvor det er formet en digital vignettundersøkelse.

Spørreundersøkelsen er laget av Mads Nordmo Arnestad i programmet Qualtrics. Undersøkelsen måtte gjennomføres digitalt, og ble distribuert til deltakerne over privat melding og mail. Alle deltakerne ble i introduksjon til undersøkelsen informert om total anonymisering av data, samt bedt om å gi sitt samtykke til deltakelse. Undersøkelsen var aktiv og gjennomførbar i tidsperioden januar til mars 2020.

2.2.1 Kvalitativ vs. kvantitativ

Et skille som raskt dukker opp i samfunnsvitenskapelig metodelære er skillen mellom kvantitative og kvalitative metoder (Johannessen, Tufte, & Christoffersen, 2016, s. 27).

Kvalitativ metode tar utgangspunkt i detaljert informasjon, som sier noe om kvalitet eller spesielle kjennetegn eller egenskaper ved det fenomenet som skal studeres. Denne metoden er spesielt hensiktsmessig dersom man skal

undersøke et fenomen man ikke kjenner særlig godt, eller når man skal undersøke elementer man ønsker å forstå mer fylldig (...).

Kvantitativ metode befatter seg med innsamling og analyse av tall og det som er kvantitativt målbart for så å søke å trekke mer generelle konklusjoner ut i fra størrelsene og sammenhengene mellom dem. Kvantitativ metode kan også anvendes og tilpasses når det er mennesker og menneskelige fenomener som skal undersøkes. Spørreundersøkelser brukes ofte også i en kvantitativ tilnærming, hvor man oppfatt av for eksempel å kartlegge utbredelse av et fenomen (holdninger eller preferanser) blant mange personer (...). Formålet med kvantitativ metode er som oftest å teste en hypotese, for å finne ut om antakelsene om virkeligheten stemmer overens med de data vi har samlet inn (Dahlum, 2019).

I vår undersøkelse var det naturlig å ta i bruk kvantitativ metode. Vi ønsket å undersøke hva mange respondenter mente om de samme spørsmålene. Respondentene fikk oppgitt flere spørsmål, knyttet til eksperimentets tekst, med svaralternativer fra skala «helt uenig» til «helt enig».

2.2.2 Sekundærdata vs. Primærdata

Primærdata samles inn av forskeren selv, eller planlegges av forskeren for prosjektets formål (Ringdal, Enhet og mangfold, 2018, s. 124). Normalt skiller vi mellom spørreundersøkelser, dybdeintervjuer og observasjon. Våre primærdata kommer fra den kvantitative spørreundersøkelsen vi har distribuert og analysene vi har gjort. Data innhentet fra denne er med på å underbygge vårt forskningsdesign og videre overordnede problemstilling.

Sekundærdata er ikke avgrenset til forskningsdata. Dette er data som ikke er behandlet av forskeren selv, men er hentet ut fra sekundærkilder; som prosessdata, bokholderidata og forskningsdata (Ringdal, Enhet og mangfold, 2018, s. 218). Vi har benyttet sekundærdata i form av både etablert og nyere forskning om moralsk ansvar, ansvarsfordeling samt kunstig intelligens; både historisk og i helsesektoren. Videre har vi supplert med artikler og relevant pensumslitteratur.

2.3 Utvalget

I vår undersøkelse ønsket vi å finne ut hvor ansvar plasseres når en kunstig intelligens gjør feil. Vi tok utgangspunkt i et bekvemmelighetsutvalg, hvor vi rekrutterte relativt bredt blant ulike respondenter som var enkelt å få tak i. Undersøkelsen er distribuert til forskjellige aldersgrupper, med ulik yrkestittel i ulike yrkessektorer. Den ble blant annet delt med studenter, ansatte i banksektor, helsesektor, finansnæring, meklerhus, forsvaret og selvstendig næringsdrivende.

Vi henvendte oss personlig til bekjente for å rekruttere respondenter. Dette for å sikre god variasjon av målgruppen. Vi ønsket tilnærmet jevn fordeling av kvinner, menn av ulike alder samt med spredning i inntekt, yrkestittel og bransje. Vi er kjent med at enkelte av invitasjonene til undersøkelsen ble sendt videre ut på arbeidsplasser. Invitasjonene havnet dermed noe utenfor vår kontroll, noe som gjør det vanskelig å fastslå eksakt hvor mange som ble invitert til å gjennomføre undersøkelsen. Vi fikk 158 respondenter til undersøkelsen hvor 121 av disse gjennomførte hele eksperimentet. Det foreligger altså bortfall. Vi kan ikke med sikkerhet fastslå hvorfor enkelte valgte å trekke seg fra undersøkelsen underveis.

2.3.1 Invitasjon til undersøkelsen

Invitasjon til undersøkelsen ble sendt ut som privat melding og e-post.

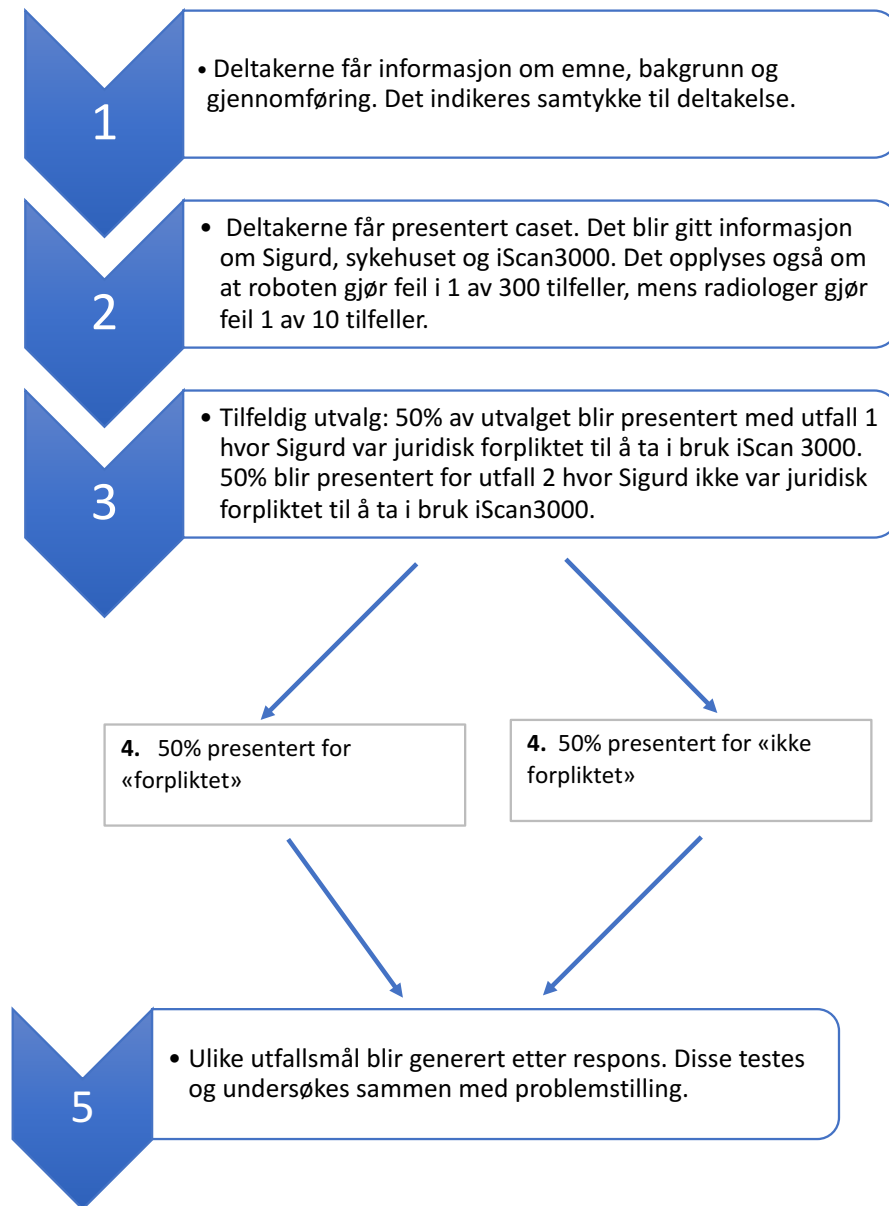
Hei! Vi er to studenter ved Handelshøyskolen BI som trenger kandidater til vårt eksperiment i forbindelse med bacheloroppgave. Studien handler om hvordan vi tenker om moralsk ansvar når kunstige agenter (artificial intelligence) gjør feil. Nyere utvikling innen kunstig intelligens og AI-algoritmer tilbyr enorme muligheter, men også betydelige utfordringer. Om bare halvparten av forhåpningene slår til, vil selvlærende algoritmer snart tilby datadrevet beslutningsstøtte på et nivå som verden aldri har sett før.

Har du mulighet til å ta deg 10 min til å svare på en undersøkelse?

Følg link: https://bino.qualtrics.com/jfe/form/SV_0USztLsm6YEytZX

Tusen takk, du er til stor hjelp.

2.4. Eksperimentets design



Figur 1: Eksperimentets design

Undersøkelsen inneholdt totalt 41 spørsmål, hvor respondentene hadde mulighet til å gå videre i undersøkelsen uten å svare på et spesifikt spørsmål. I innledningen får respondentene informasjon om at de kan avbryte undersøkelsen når som helst.

Spørreundersøkelsen i sin helhet var todelt. I den første delen ble respondentene informert om roboten iScan3000 og dens nøyaktighet

sammenliknet med menneskelig utførelse. Undersøkelser har vist at iScan3000 gjør feil i 1 av 300 tilfeller, mens menneskelig radiologer gjør feil 1 av 10 tilfeller. Respondentene blir videre introdusert for radiolog Sigurd Rosfjord, som var ansvarlig behandler for en MR-scan av lungene til en pasient. Halvparten av respondentene fikk vite at Sigurd hadde et valg om hvorvidt han ønsket å gå gjennom bildene selv, eller la iScan3000 gjøre det for han. Resterende halvpart fikk vite at Sigurd ikke hadde et valg om å bruke iScan3000. I dette senarioet hadde sykehuset stadfestet at radiologen ikke fikk bestemme selv hvorvidt han ønsket å benytte seg av roboten.

Programmet rapporterte ingen indikasjon på sykdom og pasienten ble derfor sendt hjem i begge tilfellene. Det viste seg i ettertid at pasienten hadde en voksende svulst, som iScan3000 misoppfattet, hvilket i begge tilfeller medførte at pasienten mistet livet. I den siste delen av undersøkelsen har vi kartlagt demografiske variabler hvor respondentene ble spurt om kjønn, alder, inntekt, høyeste utdanning, posisjon på arbeidsplass samt hvilken sektor respondenten arbeider i.

2.5 Utfallsmål

For å besvare vårt forskningsspørsmål har vi analysert hvordan de uavhengige variablene påvirker respondentens meninger basert på de avhengige variablene. Vi benyttet oss av følgende fem avhengige variabler; 1) kritikkverdighet, 2) ansvar, 3) straff, 4) sinne, og 5) erstatningsansvar, hvor vi formulerte tre nyanserte spørsmål tilknyttet hver variabel (se vedlegg 1 for full målemodell). Bakgrunnen for dette var primært å kvalitetssikre at respondentene ikke misforstod spørsmålene. Vi brukte en likert-skala for å måle respondentenes holdninger til de avhengige variablene. Respondentene ble bedt om å angi i hvilken utstrekning de er enig eller uenig i en serie utsagn (Gripsrud, Olsson, & Silkoset, 2016). Holdningene ble målt på en skala fra 1-7, hvor 1 representerte "helt uenig", 4 representerte "nøytral" og 7 representerte "helt enig" (Likert, 1932).

Den uavhengige variabelen for forskningsspørsmålet er hvorvidt Sigurd hadde et valg eller ikke, til å utføre undersøkelsen selv. Vi brukte den uavhengige

variabelen til å finne ut; 1) hvorvidt Sigurd fortjente kritikk, 2) mengden ansvar, 3) hvor stor grad han fortjener straff, 4) hvorvidt det oppstod sinne hos respondenten, og 5) hvorvidt Sigurd, sykehuset eller selskapet var forpliktet til å betale erstatning til pasienten.

3.0 Resultater

I denne delen av oppgaven skal vi legge frem resultatet av eksperimentet vårt. Vi har gjennomført en spørreundersøkelse hvor samtlige av deltakerne ble informert om at roboten iScan3000 er 300 ganger sikrere enn en undersøkelse utført av radiolog Sigurd Rosfjord. Videre ble deltakerne delt inn i to grupper. Den ene halvparten fikk vite at Sigurd hadde et valg om hvorvidt han ønsket å utføre undersøkelsen selv, eller ta i bruk iScan3000. Den andre halvparten fikk vite at Sigurd ikke hadde et valg, hvorav retningslinjene til sykehuset sa at han måtte ta i bruk iScan3000 til undersøkelsen.

3.1 Deskriptive data

3.1.1 Alder og kjønn

Det var totalt 158 personer som aksepterte og valgte å delta i undersøkelsen, hvorav 121 gjennomførte hele undersøkelsen og besvarte alle spørsmål. Gjennomsnittsalderen på respondentene er 35 år, hvor den den yngste respondenten er 19 år, og den eldste er 71 år. Dette gir oss et standardavvik på 14.43 år, som forteller oss at det er god spredning i utvalgets aldersgruppe.

Når det kommer til kjønn var det totalt 123 kandidater som besvarte spørsmålet, hvorav 58 respondenter (47,15%) var kvinner og 65 respondenter (52,85%) var menn.

Videre viser vi fordeling i deskriptive data ved hjelp av diagrammer.

3.1.2 Inntekt

Alle typer inntekter er representert i undersøkelsen. Fordelingen på respondentenes lønnsnivå varierer fra intervallet 0 – 100 000, og opp til over 1 000 000 kr i året. Det er totalt 122 respondenter som har besvart spørsmålet.

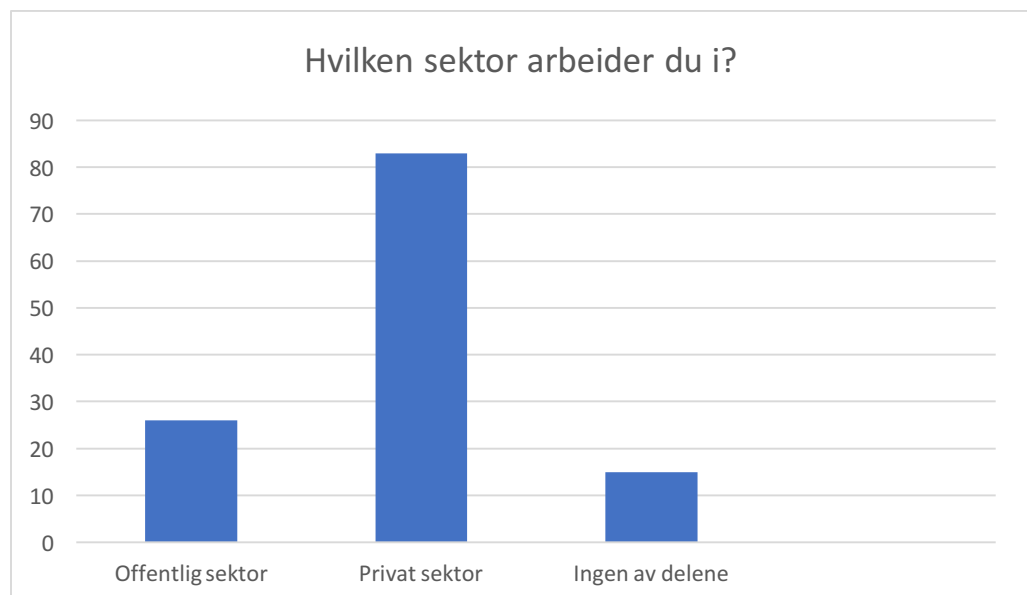
Gjennomsnittlig årslønn til respondentene i vår undersøkelse er 525 000kr, noe som reflekterer en gjennomsnittlig norsk årslønn som i 2019 var på 567 000 (SSB, 2020).



Figur 2. Inntekt

3.1.3 Sektor

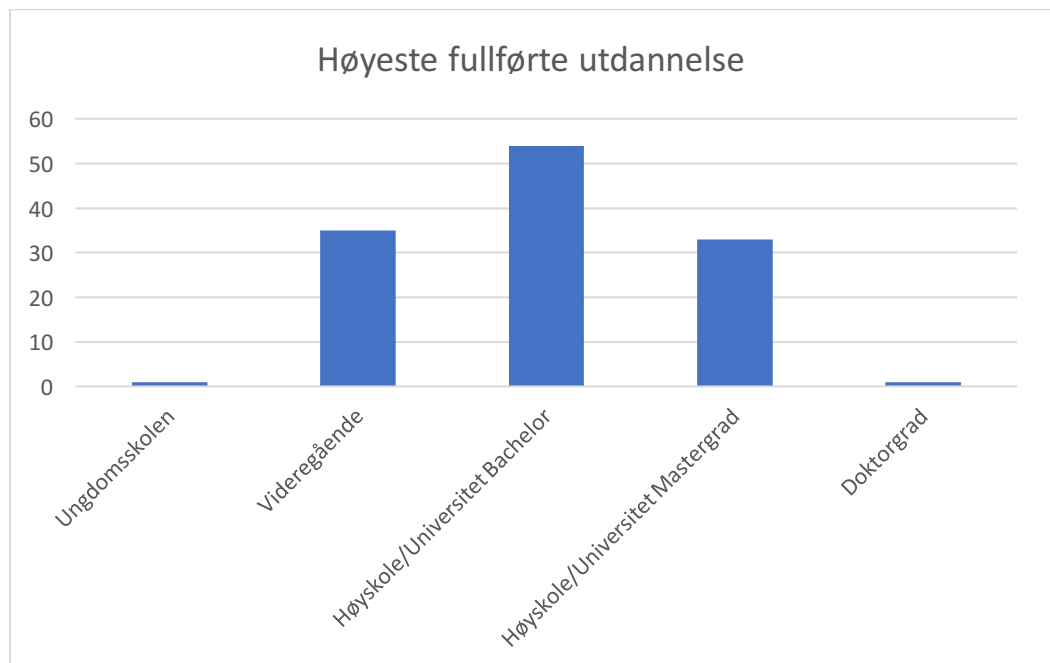
Spørreundersøkelsens respondenter arbeider både i privat og offentlig sektor. Totalt 124 stykker har besvart spørsmålet; hvor 26 stykker (20.97%) tilhører offentlig sektor, og 83 stykker (66.94%) tilhører privat sektor. 15 stykker (12.10%) av respondentene har svart at de ikke arbeider.



Figur 3. Privat eller offentlig sektor

3.1.4 Høyest fullførte utdanning

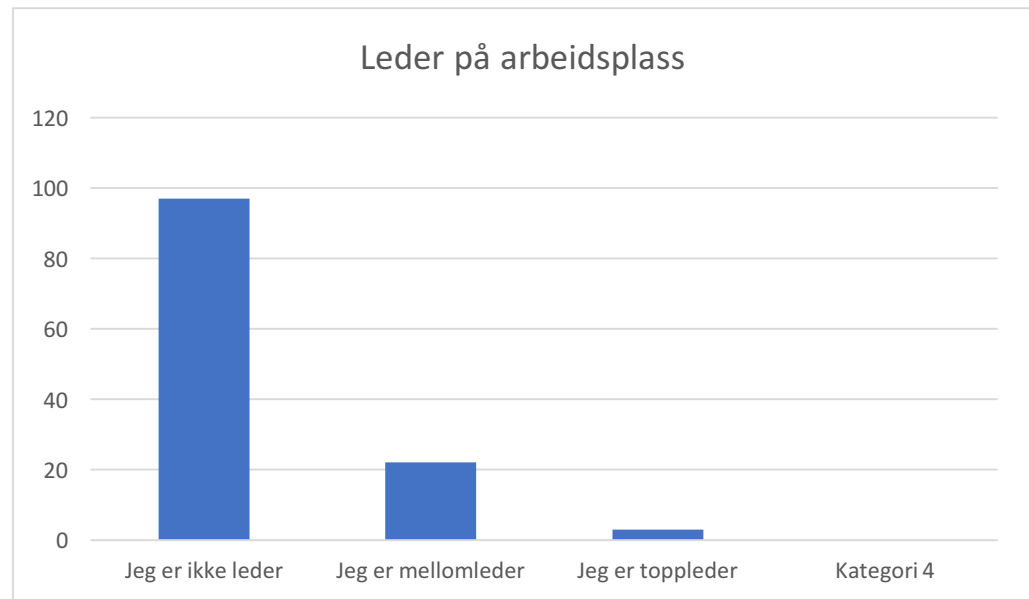
Respondentenes utdanningsnivå varierer fra fullført ungdomsskole til fullført doktorgrad. Det er totalt 124 stykker som har besvart spørsmålet. Fordelingen av høyeste fullførte grad viser at 1 person (0.83%) har fullført ungdomsskolen, 35 stykker (28.23%) fullført videregående, 54 stykker (43.55%) fullført høyskole/ universitet bachelor eller mellomfag, 33 stykker (26.61%) fullført høyskole/ universitet master/hovedfag og 1 stykk (0.81%) har fullført doktorgrad.



Figur 4. Fullført utdanning

3.1.5 Lederfordeling

I spørsmål om lederposisjon på arbeidsplassen har 122 kandidater besvart spørsmålet. Fordelingen viser at størst andel med 97 stykker (79.51%) er ikke leder, 22 stykker (18.03%) er mellomleder og 3 stykker (2.46%) er toppleder.



Figur 5. Lederrolle

3.1.6 Kritisk evaluering av data

Tidligere i oppgaven har vi lagt frem viktigheten av en studies validitet og reliabilitet; vi må derfor være kritisk til de innsamlede deskriptive data som er med å formeresultatene våre.

De innsamlede dataene tegner et bilde av et relativt representativt bekvemmelighetsutvalg. Alle demografiske karakteristika er representert, og det er derfor ingen grunn til å tro at resultatet kun gjelder en spesifikk gruppe mennesker. Representativiteten i bekvemmelighetsutvalget gjenspeiles også i data fra lederroller. Fordelingen reflekterer populasjonen; hvor flesteparten ikke er leder, enkelte er mellomleder og de færreste toppleder.

Det er viktig å poengtere at deltakerne kunne trekke seg fra studiet når de ønsket, samt la være å svare på enkelte spørsmål. Derfor varierer det totale antallet respondenter fra spørsmål til spørsmål. Vi må også være kritiske til det faktum at alle respondentene i undersøkelsen var anonyme; noe som gjør det vanskelig å fastslå om enkelte har svart uærlig, latt seg påvirke av faktorer som støy og tid, og om den enkelte respondent har gjennomført studien flere ganger. Vi må også være kritiske til at enkelte kandidater kan ha tolket spørsmål feil.

3.2 Konstruksjon av målemodell

3.2.1 Cronbachs alfa

Cronbachs alfa beregnes for å kvalitetsteste reliabiliteten til undersøkelsen vår. Den måler reliabiliteten i en summert indeks basert på et sett indikatorer, og sier dermed noe om hvor internt korrelert de ulike svarene på de ulike spørsmålene er (Ringdal, 2018, s. 367). Dersom Cronbachs Alpha er lavere enn nedre grense gitt ved 0,7, er det noe galt med målemodellen, og modellen bør dermed forkastes. Dersom verdien er høyere enn 0.7 kan målemodellen brukes. For å få riktig Cronbachs alfa måtte vi reversere 3 av 11 spørsmål, slik at verdiene ble målt samme vei.

Kritikkverdighet, moralsk ansvar og straff ble målt med 3 spørsmål, og sinne med 2 spørsmål. I tabell 1 ser vi at kritikkverdighet har en Cronbachs Alpha på .890, moralsk ansvar .931, vurdering av rettmessighet for straff .874, og opplevd sinne .920. Alle våre spørsmål ga høye utslag på skalaen, som forteller oss at målemodellen har en høy reliabilitet.

<i>Variabel</i>	<i>Cronbach's Alpha</i>
1. Kritikk mot radiolog Sigurd	.890
<i>Jeg syns Sigurd fortjener kritikk for den inntrufne hendelsen.</i>	
<i>Sigurds fremferd kan og bør kritiseres.</i>	
<i>Det er ikke riktig å kritisere Sigurd i dette tilfellet.</i>	
2. Moralsk Ansvar	.931
<i>Sigurd er moralsk ansvarlig for dette.</i>	
<i>Det moralske ansvaret for hendelsen må radiolog Sigurd Rosfjord ta.</i>	
<i>Det blir galt å holde Sigurd moralsk ansvarlig for det som skjedde.</i>	
3. Ønske om å straffe Sigurd	.874
<i>Sigurd Rosfjord fortjener straff for dette</i>	
<i>Det er riktig at Sigurd utsettes for en disiplinær reaksjon etter dette</i>	
<i>Det blir feil å straffe Sigurd for det som skjedde</i>	
4. Sinne mot Sigurd	.920
<i>Det Sigurd gjorde vekket sinne i meg</i>	
<i>Jeg ble sint av å lese om denne radiologen</i>	

Tabell 1. Cronbachs alfa

3.2.2 Korrelasjonsanalyse

Korrelasjon er den statistiske sammenhengen mellom to variabler. En robust sammenheng mellom årsak og virkning betyr at den empiriske sammenhengen (korrelasjonen) mellom X og Y ikke kan skyldes andre forhold (Z), som er årsak til begge to (Ringdal, 2018, s. 50). Dette er med på å forklare at sammenhengen mellom X og Y ikke er tilfeldig, eller spuriøst, men derimot et statistisk mål på samvariasjon. Korrelasjon uttrykkes tallmessig gjennom Pearson correlation (r).

En sterk korrelasjon vil være en korrelasjon ($r=.6$) eller høyere. En moderat korrelasjon er mellom ($r=.3$) og ($r=.6$). Hvis tallet derimot er lavere enn ($r=.3$) så eksisterer det ikke noen sammenheng.

Figur 7 viser tydelig at kritikkverdighet, moralsk ansvarlig, straff og sinne er alle moderat til sterkt korrelert med hverandre ($r > .5$). Korrelasjonsmatrisen forklarer at jo høyere verdi deltakerne har gitt på en av de fire verdiene, desto høyere er tendensen til å score høyt på de tre andre verdiene. Det vil si at de som mente at Sigurd opptrådte kritikkverdig, også mente han var moralsk ansvarlig og fortjente straff.

Spørsmålet om erstatning har en mildere korrelasjon ($r < .46$) med de nevnte faktorene. Det ser ut til at vurderingen av erstatningsansvar er mer uavhengig av hvorvidt Sigurd er moralsk ansvarlig, fortjener straff og opptrer kritikkverdig. En mulig forklaring på dette er at det ikke er kultur å kreve erstatning fra enkeltstående medarbeidere i norsk arbeidsliv. I de fleste tilfeller er det arbeidsgiver som står ansvarlig for feil og avvik; ikke arbeidstaker selv. Vi ser en mild korrelasjon for denne variabelen.

I korrelasjonsmatrisen ser vi at jo høyere inntekt deltakerne har, i desto mindre grad mener de at selskapet skal stå erstatningsansvarlig. Vi ser at samme fenomen gjelder for de eldre i utvalget.

Måling	1	2	3	4	5	6	7	8	9
1. Kritikkverdig									
2. Moralsk ansvarlig	.721								
3. Straff	.707	.553							
4. Sinne	.641	.514	.606						
5. Pårørte burde få erstatning fra Sigurd	.347	.254	.367	.464					
6. Pårørte burde få erstatning fra sykehuset	.076	.110	.120	.192	.195				
7. Pårørte burde få erstatning fra produsent	.031	.030	.160	.205	.290	.371			
8. Hvor gammel er du?	-.022	-.123	-.101	.078	.048	-.101	-.330		
9. Hvor mye tjener du i året?	-.108	-.138	-.210	.014	.015	-.110	-.335	.789	

Tabell 2: Korrelasjonsmatrise

3.3 Test av mellomgruppeskjeller: Hypotesetesting

I denne delen av oppgaven skal vi gjennomføre en independent-samples t-test for alle 7 hypotesene. Her skal vi teste om gruppen som leste at Sigurd hadde et valg plasserer ansvar annerledes, enn gruppen som leste at Sigurd ikke hadde et valg. Vi har brukt kontrollspørsmål 1 som uavhengig variabel. Se vedlegg 2 for fullstendig hypotesetest.

H1: Sigurd sin fremferd blir vurdert som mer kritikkverdig når han hadde et valg.

Hypotese 1 vektlegger kritikkverdighet når Sigurd *hadde et valg*. Gruppen som leste at Sigurd *ikke hadde et valg* mener at han i snitt hadde en kritikkverdighet på $M=2,791$. Dette er på en skala fra 1-7, hvor 1 er overhode ikke kritikkverdig og 7 er svært kritikkverdig. Standardavviket er 1,565.

Gruppen som leste at Sigurd *hadde et valg*, mener han har opptrådt på et nivå på $M=3,95$ som da er veldig nært 4 som er nøytral. Nøytral er hverken kritikkverdig eller ikke kritikkverdig. Standardavviket er på 1.58 som er nesten det samme som forrige gruppe.

Dersom $p < .05$ er forskjellen mellom gruppene stor nok til at den er statistisk signifikant, og vi finner støtte for hypotesen. Hypotesen 1 har en verdi på $p < .001$ og resultatene gir dermed støtte for den første hypotesen.

H2: Sigurd vil holdes moralsk ansvarlig for utfallet når han hadde et valg.

Gruppen som leste at Sigurd *ikke hadde et valg* mener at han har et moralsk ansvar på $[M=3,04, SD=1.695]$. Gruppen som leste at Sigurd *hadde et valg*, gir han nesten et helt tall høyere $[M=3,9, SD=1.501]$. Denne forskjellen mellom gruppene er signifikant fordi $p=.004$, ergo finner vi støtte for hypotesen. Den absolutte scoren for begge svarene er under nøytralt nivå ($M < 4$).

H3: Det at Sigurd hadde et valg gjør at han i større grad fortjener straff enn om han ikke hadde hatt et valg.

Gruppen som leste at Sigurd *ikke hadde et valg*, oppnådde et nivå på $[M=1.933, SD=1.051]$. De mener det ikke er riktig å straffe Sigurd for handlingen, da han har fulgt sykehusets rutiner. Gruppen som leste at Sigurd *hadde et valg* de gir han nesten et helt tall høyere $[M=2.88, SD=1.442]$. Denne forskjellen er signifikant fordi $p < .001$, ergo finner vi støtte for hypotesen. Den absolutte scoren for begge svarene er godt under nøytralt nivå ($M=4$) som betyr at ingen av gruppene mener at Sigurd fortjener straff.

H4: Sigurds fremferd vekket sinne.

Gruppen som leste at Sigurd *ikke hadde et valg*, oppnådde et nivå på $[M=2.589, SD=1.519]$. Gruppen som leste at Sigurd *hadde et valg* oppnådde nesten samme nivå: $[M=2.686, SD=1.665]$. Hypotesen er ikke signifikant fordi

$p=.745$, som vil si at vi ikke finner støtte i hypotese 4. Det er ingen forskjell i hvor sint deltakerne i de ulike gruppene blir.

H5: Sigurd blir sett på som mer erstatningsansvarlig hvis han hadde et valg.

Gruppen som leste at Sigurd *ikke hadde et valg*, oppnådde et nivå på $[M=2.13, SD=1.779]$. Gruppen som leste at Sigurd *hadde et valg* oppnådde nesten samme nivå: $[M=2.07, SD=1.363]$. Forskjellen er ikke signifikant fordi $p=.847$, som vil si at vi ikke finner støtte for hypotese 5. Funnene forklarer at uansett om Sigurd hadde et valg eller ikke, syntes ikke respondentene at radiologen alene skal stå erstatningsansvarlig for hendelsen.

H6: Sykehuset blir sett på som mer erstatningsansvarlig hvis Sigurd ikke hadde et valg.

Gruppen som leste at Sigurd *ikke hadde et valg* oppnådde et nivå på $[M=5.36, SD=1.873]$. Gruppen som leste at Sigurd *hadde et valg* oppnådde et nivå på $[M=4.50, SD=1.857]$. Forskjellen mellom gruppene er signifikant fordi $p=.016$. Resultatene gir dermed støtte i vår hypotese om at sykehuset blir sett på som mer erstatningsansvarlig hvis Sigurd ikke hadde et valg.

H7: Produsenten av iScan3000 blir sett på som mer erstatningsansvarlig hvis Sigurd ikke hadde et valg.

Gruppen som leste at Sigurd *ikke hadde et valg* sier at *produsenten av iScan3000* er mer erstatningsansvarlig enn om Sigurd hadde fått et valg om å bruke maskinen $[M=4.48, SD=2.089]$. Gruppen som leste at Sigurd *hadde et valg* mener at produsenten er mindre erstatningsansvarlig enn om Sigurd ikke hadde fått et valg $[M=3.53, SD=1.933]$. Denne forskjellen er signifikant fordi $p=.012$. Resultatene gir dermed støtte for hypotesen om at produsenten av iScan3000 blir sett på som mer erstatningsansvarlig dersom Sigurd han ikke hadde et valg.

4.0 Diskusjon

Målet med studiet har vært å teste hvor ansvaret plasseres når en kunstig intelligens utfører en fatal feil. Forsking og studier på tvers av fagfelt har vist at mennesker har en tendens til å tenke at hvis du har et valg, så har du også ansvar for utfallet (Cappelen, 2018). Vi har testet seks ulike hypoteser basert på de ulike utfallsmålene i spørreundersøkelse.

Hypotese én forklarer i hvilken grad deltakerne av undersøkelsen mener Sigurd fortjener kritikk. Hypotese to tar for seg i hvilken grad Sigurd står moralsk ansvarlig for hendelsen. Respondentene var informert i innledningen til spørreundersøkelsen om at roboten var betydelig mer treffsikker enn menneskelig utførelse. Til tross for denne informasjonen, viser resultatet at Sigurd stiller *mer* moralsk ansvarlig for feilen, hvis han hadde et valg om å utføre jobben selv. Resultatet viser også at han *i større grad* fortjener kritikk dersom han hadde valget om å utføre jobben selv. Dog tilsier svarene at Sigurd verken fortjener kritikk eller bør holdes moralsk ansvarlig, absolutt sett (M er under 4 for alle alternativ).

I hypotese tre undersøker vi i hvilken grad Sigurd fortjener straff. Vi finner også støtte for denne hypotesen. Respondentene synes han fortjener straff i større grad, dersom han hadde et valg om å utføre jobben selv. Absolutt sett er det interessant at respondentene mener at han ikke fortjener straff uansett om han har et valg eller ikke (M er godt under 4 for begge alternativ).

Hypotese fire tar for seg i hvilken grad hendelsen vekket sinne hos respondentene. Vi finner ikke støtte for hypotesen, da det ikke er signifikant forskjell i hvor sint respondentene i de ulike gruppene blir. Grunnen til dette er kanskje at eksempelet virker noe hypotetisk og urealistisk.

Hypotese fem undersøker i hvilken grad Sigurd blir sett på som mer erstatningsansvarlig hvis han hadde et valg. På lik linje som i hypotese 4, finner vi heller ikke støtte for dette. Funnene forklarer at uansett om Sigurd hadde et valg eller ikke, synes ikke deltakerne at radiologen personlig skal være erstatningsansvarlig. Funnet harmonerer også med norsk arbeidskultur

hvor en ikke holder enkeltansatte ansvarlige for feil de gjør. Det er oftest arbeidsgiver som blir erstatningspliktig og må ta ansvar for arbeidstaker sine feil.

Hypotese seks tar for seg om sykehuset blir sett på som mer erstatningsansvarlig hvis Sigurd ikke hadde et valg. Vi finner støtte for dette. Hvis sykehuset sine retningslinjer tilsier at Sigurd ikke har et valg, synes utvalget at sykehuset fortjener å være erstatningspliktig i større grad enn hvis Sigurd hadde et valg. De absolutte verdiene tilsier at sykehuset bør være erstatningspliktig uansett om Sigurd har et valg eller ikke (M er godt over 4 i begge alternativene).

Med andre ord er sykehuset tjent med å gi Sigurd et valg. For det første vet sykehuset at Sigurd trolig ikke kommer til å gjøre jobben selv; fordi Sigurd vet at roboten utfører oppgaven mer treffsikkert. For det andre vet sykehuset at dersom Sigurd har et valg; velger å bruke roboten, og denne så gjør feil, holdes sykehuset tross alt noe mindre erstatningsansvarlig enn om Sigurd ikke gis et valg. Av alle svar er det sykehuset som oppnår høyest score med hensyn til å holdes ansvarlig; høyere enn Sigurd og høyere enn produsenten av iScan3000.

På lik linje med hypotese seks, finner vi også støtte for hypotese syv. Respondentene mener at produsenten av iScan3000 er mer erstatningsansvarlig hvis Sigurd ikke hadde et valg. Dette kan ses i lys av studiet til Cappelen (2018). På lik linje som med sykehuset i hypotese 6, ser vi at produsenten er tjent med å ha på en *manual mode knapp*. Det viser seg at så lenge knappen eksisterer, endrer dette folks oppfatning av hvor ansvarlig produsenten er for at roboten gjør feil. Dersom knappen er tilstede holdes verken sykehuset eller produsenten like mye ansvarlig, da legen gis et valg.

4.1 Implikasjon

Resultatene fra eksperimentet tyder på at Sigurd blir holdt ansvarlig for utfallet dersom han gis et valg, sammenliknet med tilfellet der han ikke gis et valg. Våre resultater er i samsvar med Cappelen et. Al (2018) sin studie om hvordan individer holdes personlig ansvarlig for sine valg. Studien viser til Professor i

psykologi, Peter Vallentyne, som har angitt to minimumskrav for at en person ikke kan holdes ansvarlig. Det første kriteriet er at dersom personen ikke kunne ha endret sannsynligheten for utfallet, skal vedkommende ikke kunne holdes ansvarlig. Det andre kriteriet sier at dersom personen bare kunne ha unngått utfallet ved å velge et urimelig alternativ, skal vedkommende heller ikke holdes ansvarlig (Vallentyne, 2008). Utfallet av vår studie viser at begge Vallentynes kriterier frikjenner Sigurd for å holdes ansvarlig; når han ikke gis mulighet til å påvirke utfallet ved å gis et valg, eller når han måtte valgt et klart dårligere alternativ for å unngå utfallet, så holder respondentene ham ikke ansvarlig eller i mindre grad ansvarlig.

Sigurd er utdannet radiolog, jobber i helsesektor og vet han opererer med menneskers liv. Det er derfor ingen grunn til at han ikke skal velge roboten som er det mest gunstige og tryggeste for pasienten. Utvalget fikk opplysningen om at det er betydelig forskjell på robotens og Sigurds treffsikkerhet; det er hele 300 ganger mer sikkert å ta i bruk roboten, enn å gjennomføre undersøkelsen selv. En vurdering av dette gjør det klart at det er umoralsk og urimelig av Sigurd å gjennomføre undersøkelsen selv. Dersom det ikke hadde vært så stor forskjell i sannsynligheten for sikkerhet mellom å bruke roboten og gjennomføre eksperimentet selv; hadde valget om å gjennomføre selv ikke vært urimelig.

Eksperimentet vi har gjennomført reflekterer også disse funnene. I resultatet vårt foreligger det positiv korrelasjon mellom ansvars plassering, straff og kritikk. Resultatet gir oss dermed svar på at deltakerne som tillegger Sigurd ansvar, også mener han fortjener mer straff og mer kritikk når det foreligger et valg. En forklaring på dette kan være fenomenet algoritmeaversjon. Algoritmeaversjon er en av årsakene til at mennesker velger å ikke benytte seg av kunstig intelligens (Dietvorst, 2014). Dette sammenfaller med forskning fra Dietvorst et al. (2014), som viser at mennesker anser datamaskiner og roboter som trusler og dermed ikke tildeler de tillitt. Forskningen reflekterer at pasienter finner det fremmedgjørende og upersonlig, samt skremmende å la en robot gjennomføre en klinisk undersøkelse. Dette kan også være med på å underbygge hvorfor enkelte respondenter mener det er kritikkverdigg av Sigurd

å benytte roboten, når han har et valg om å gjennomføre selv; til tross for at respondentene vet roboten er 300 ganger mer treffsikker.

4.1.1 Implikasjoner for ledelse

Det er viktig for en sykehusledelse å være klar over at de indirekte gir en ansatt økt personlig ansvar når det foreligger et valg om å overstyre eller erstatte en kunstig intelligens. Dersom den ansatte selv har valget om å benytte seg av roboten eller gjennomføre undersøkelsen manuelt, er det nødvendig at den ansatte kjenner til risiko og konsekvenser ved sin posisjon. Det viser seg at dersom det foreligger en manual mode knapp har leder plasseres det økt ansvar til den ansatte, simpelthen fordi vedkommende gis et valg. En leder bør være oppmerksom på dette og informere den ansatte om risiko, feilmargin og implisitt ansvar ved bruk av roboten.

Dersom den ansatte er juridisk forpliktet av sykehuset til å ta i bruk roboten, vil den ansatte altså reelt sett tillegges et mindre ansvar for utfallet. På denne måten fraskriver ledelsen den ansatte en del av ansvaret rundt bruken av roboten.

4.2 Begrensning og anbefalinger for fremtidig forskning

Trusler og begrensninger er med på å svekke generaliserbarheten til forskningseksperimentet vårt. Videre ser vi på svakheter ved oppgaven, samt anbefalinger for fremtidig forskning.

Den første begrensningen i oppgaven er at alle i utvalget er norske. Vi kjenner til at andre land og kulturer tenker annerledes om valgfrihet og ansvar, sammenliknet med Norge. Vi kan derfor ikke generalisere resultatene på tvers av kulturer. Anbefalingene for fremtidig forskning er å utforske liknende problemstillinger basert på et utvalg fra andre land og kulturer.

Den andre begrensningen er at scenarioet trolig virker urealistisk for respondentene. Dette forklarer trolig hvorfor hypotesetest fire ikke ble signifikant; respondentene opplevde ikke sinne. Utvalget opplever tanken om en omfattende kunstig intelligens litt fjern, og dermed fremstår

problemstillingen fiktiv. En slik innvending kan være en trussel mot validiteten til resultatene. En anbefaling for fremtidig forskning blir derfor å gjenta disse studiene, i takt med at kunstig intelligens i større grad implementeres i samfunnet. Undersøkelsens validitet styrkes når kunstig intelligens blir normalisert; scenarioet vil da trolig oppleves mer realistisk.

En tredje begrensning for oppgaven er at utvalget består av 158 deltakere, hvorav 121 besvarte samtlige spørsmål. Et mer omfattende utvalg ville styrket variabelens reliabilitet, og gitt oss mer data til å sammenlikne og trekke konklusjoner. Flere nyanser i utvalgets synspunkter ville da blitt undersøkt nærmere. Anbefalinger for fremtidig forskning er å fortsette forskning på samme problemstilling med et større utvalg respondenter.

Avslutningsvis ønsker vi å trekke frem viktigheten rundt utvikling av lovverk og regler omkring ansvarsfordeling ved bruk av kunstig intelligens. Som kjent viser resultatene fra oppgaven at Sigurd tildeles mer skyld og straff dersom det eksisterer en manual mode knapp. Kunstig intelligens tilbyr et verdifullt tilskudd til mange felter av samfunnsutviklingen; både i diagnostisering og behandling. Samtidig har vi de siste årene sett stadig økt tendens til rettsliggjøring i samfunnet, også i helse og omsorgstjenester. Hvilket betyr at interesser underlegges rettslige lover, og dermed lovfestes.

Samme tendens kan tenkes å bli gjort gjeldende innen feltet vi omhandler i vår oppgave. Som oppgaven viser er det mange hensyn å ta for å kunne sørge for at interessene til både pasient, behandler (lege) og institusjon (sykehus) sågar industrien (selskapet) ikke gir seg ubalanserte eller uønskede utslag. Ikke alle interessekonflikter og maktforhold er åpenbare i et hurtig utviklende felt som dette. Vi tenker at en oppgave som vår er med på å belyse betydningen av å utrede slike problemstillinger grundig før konklusjoner trekkes om en god praksis i fremtiden.

5.0 Konklusjon

I denne studien har vi prøvd å konkretisere problemstillingen ”*Hvor plasseres ansvaret når en AI gjør feil?*”. Som kjent oppstår det flere utfordringer knyttet til bruk av kunstig intelligens, som for eksempel algoritme aversjon. Algoritme aversjon bærer preg av mangelfull kunnskap og tillit hvilket gjør at folk flest er å foretrekke en menneskelig vurdering fremfor bruk av kunstig intelligens. Det foreligger også mangelfullhet i prosedyrer rundt plassering av ansvar når en kunstig intelligens feiler. Som en følge av dette bør det diskuteres hvordan ansatte skal holdes ansvarlig, samt tydeliggjøre risikoen for ansatte ved bruk av kunstig intelligens med mulighet for å overstyre.

Ut fra resultatene fra forskningseksperimentet og hypotesetesting kan vi konkludere med at både sykehus og produsent er tjent med en *manual mode knapp* som kan overstyre roboten. Resultatene forteller at så lenge knappen eksisterer, endrer dette folks oppfatning av hvor ansvarlig sykehus og produsent fremstår overfor eventuelle feil. Når feilen inntreffer holdes verken sykehus eller produsent i like stor grad ansvarlig dersom knappen er til stede, og legen er gitt et valg om å overstyre roboten. Dette fremmer aktualiteten av å implementere ansvar og risikoplassering overfor de ansatte ved bruk av kunstig intelligens.

6.0 Kildeliste

- Andresen, K. (2019, 06 07). *How AI changes the prognostication of cancer*.
Hentet fra Medium.com: <https://medium.com/@kariandresen/how-ai-changes-the-prognostication-of-cancer-e9f5d945b2ba>
- Andresen, K. (2020, 03 18). *Internasjonalt samarbeid på tarmkreftforskning*.
Hentet fra Radiumhospitalets Legater:
<https://radiumlegat.no/Prosjekter/Danielsen>
- Arnold, S. (2020, 04 01). *How Netflix changed what we watch on our screens*.
Hentet fra Rte: <https://www.rte.ie/brainstorm/2019/03/13/1036204-how-netflix-changed-what-we-watch-on-our-screens/>
- Berge, G. T., Granmo, O.-C., & Tveit, T. (2017, 04 04). Tester ut kunstig intelligens på pasientjournaler. *Aftenposten*.
- Cappelen, A., Fest, S., Sørensen, E., & Tungodden, B. (2018, september 12). Choice and personal responsibility; What is a morally relevant choice .
- Dahlum, S. (2019, 09 24). *kvantitativ metode*. Hentet fra www.snl.no:
https://snl.no/kvantitativ_metode
- Dawes, R. M. (1979). *The robust beauty of improper linear models in decision making*. *American Psychologist*, 34(7), 571–582.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2014). Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. *Academy of Management Proceedings*, 1, 12227.
- Donnelly , L. (2017, 03 07). *Forget your GP, robots will 'soon be able to diagnose more accurately than almost any doctor'*. Hentet fra Telegraph.com:
<https://www.telegraph.co.uk/technology/2017/03/07/robots-will-soon-able-diagnose-accurately-almost-doctor/>
- Dustin* . (2018, 08 21). Hentet fra Computer vision – en hjørnestein i kunstig intelligens:
<https://www.dustin.no/tjenester/kunnskapsbanken/archive/computer-vision-en-hjoernestein-i-kunstig-intelligens/>
- Eshleman, A. (2014, 07). *Moral Responsibility*. Hentet fra Stanford Encyclopedia of philosophy archive:
https://plato.stanford.edu/cgi/viewcontent.cgi?article=1001&context=phl_facpubs

- Feltz, A., & Cokely, E. (2009). Personality differences in intuitions about compatibilism and incompatibilism. 10.
- Greenfield, K. (2012). *The Myth Of Choice, personal responsibility in a world of limits*. Yale University Press.
- Gripsrud, G., Olsson, U. H., & Silkoset, R. (2016). Metode og dataanalyse : Beslutningsstøtte for bedrifter ved bruk av JMP, Excel og SPSS. I G. O. Gripsrud (Red.). Cappelen Damm.
- Gulshan, V., Peng, L., & Coram , M. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *Journal of American Medical Association*, s. 2402.
- Hart, R. D. (2017, 05 23). *When artificial intelligence botches your medical diagnosis, who's to blame?* Hentet fra www.quartz.com:
<https://qz.com/989137/when-a-robot-ai-doctor-misdiagnoses-you-whos-to-blame/>
- Helse, O. (2019, 11 22). *Nasjonal helse- og sykehusplan 2020-2023*. Hentet fra Regjeringen: <https://www.regjeringen.no/no/dokumenter/nasjonal-helse--og-sykehusplan-2020-2023/id2679013/?ch=1>
- Hutson, M. (2017, 04 14). *Self-taught artificial intelligence beats doctors at predicting heart attacks*. Hentet fra Sciencemag:
<https://www.sciencemag.org/news/2017/04/self-taught-artificial-intelligence-beats-doctors-predicting-heart-attacks>
- Johannessen, A., Tufte, P., & Christoffersen, L. (2016). *Introdiksjon til samfunnsvitenskapelig metode*. Abstrakt forlag.
- Johansen, R. (2019, 01 03). #01 Kreftlege Roar Johansen i IBM Watson. *Heath Talk* . (H. Anderssen, Intervjuer)
- Kahneman , D. (2011). *THINKING, FAST AND SLOW*. Farrar, Straus and Giroux.
- Kahneman, D., Rosenfield, A. M., Gandhi, L., & Blaser, T. (2016). Noise: How to Overcome the High, Hidden Cost of Inconsistent Decision Making. *Harvard Business Review*.
- Kommunal og moderniseringsdepartementet. (2020). *Nasjonal strategi for kunstig intelligens*. regjeringen.no: Kommunal- og moderniseringsdepartementet.

- Lein-Mathisen, K., & Lundgreen, L. (2019, 04 05). *Pionerarbeid med kunstig intelligens i helsesektoren*. Hentet fra Dagens Medisin:
<https://www.dagensmedisin.no/artikler/2019/04/05/pionerarbeid-med-kunstig-intelligens-i-helsesektoren/>
- Leverhumle Centre for the Future of Intelligence. (2020). *This project focuses on the collaborations between technologists, policymakers, and other stakeholders needed for the responsible development of AI*. Hentet fra www.lcfi.ac.uk: <http://lcfi.ac.uk/projects/ai-futures-and-responsibility/policy-and-responsible-innovation/>
- Likert, R. (1932). *A Technique for Measurement of Attitudes*. Hentet fra Archives of Psychology, 140(6), 5-23.:
<http://dx.doi.org.ezproxy.library.bi.no/10.1037/0021-9010.78.4.662>
- Lohr, S. (2016, 10 17). *IBM Is Counting on Its Bet on Watson, and Paying Big Money for It*. Hentet fra The Ny York Times:
<https://www.nytimes.com/2016/10/17/technology/ibm-is-counting-on-its-bet-on-watson-and-paying-big-money-for-it.html>
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a re- view of the evidence*. Minniapolis : University of Minnesota .
- Mele, & Cushman. (2008). Intentional action: Two-and-a-half folk concepts? 17.
- moderniseringsdepartementet, K. o. (2020, 01). *Nasjonal strategi for kunstig intelligens*. Hentet fra [Regjeringen.no](http://www.regjeringen.no) :
<https://www.regjeringen.no/contentassets/1febbb2c4fd4b7d92c67ddd353b6ae8/no/pdfs/ki-strategi.pdf>
- Norwegian Specialist Healthcare. (2019, 06 24). *Interoperabilitet (evne til samhandling)*. Hentet fra www.kilden.sykehusene.no:
<https://kilden.sykehusene.no/display/AR/4.+Interoperabilitet>
- Open Geospatial Consortium. (u.d.). GeoPortal - Rammeverk introduksjon. Kartverket.no, Norge.
- Perry, & Dzhamfarov. (2012). *Descriptive and Normative Approaches to Human Behavior*. singapore: World Scientific.
- Persvold, A. Z. (2020, 02 10). *Idiosynkratisk*. Hentet fra SNL:
<https://snl.no/idiosynkratisk>

- Ringdal, K. (2018). *Enhet og mangfold, samfunnsvitenskapelig forskning og kvantitativ metode*. Fagbokforlaget.
- Rowe, J. (2020). *Deloitte*. Hentet fra Tre ting du må vite om kunstig intelligens (AI):<https://www2.deloitte.com/no/no/pages/technology/articles/tre-ting-vite-kunstig-intelligens-ai.html>
- Ruud, T. T.-E. (2018, 10 05). *Sykehus får bot på 1,5 millioner etter at Djabrail (6) døde av feilbehandling*. Hentet fra VG:
<https://www.vg.no/nyheter/innenriks/i/5Vg1wm/sykehus-faar-bot-paa-15-millioner-etter-at-djabrail-6-doede-av-feilbehandling>
- ScienceDirect. (2017, juni). *The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms*. Hentet fra www.sciencedirect.com:
https://www.sciencedirect.com/science/article/pii/S0016328717300046?casa_token=mO9_BpJTOSQAAAAA:PdI-HCPGRxzyQch0QGbMLo3hZ6T0O7fQVrBh3rwo-DLgPueoenzYUdl_bx2jBUHTm7gEDpc2-w
- Sjåstad, H. (2019, 08). Algoritme-aversjon. *Magma*, 64.
- SSB. (2020, mars 3). *Lønn*. Hentet fra www.ssb.no: <https://www.ssb.no/arbeid-og-lonn/statistikker/lonnansatt>
- Svendsen, L. (2018, 11 14). *Hva er moral*. Hentet fra www.civita.no:
<https://www.civita.no/politisk-ordbok/hva-er-moral>
- Svendsen, L. (2019, 11 15). *Hva er etikk*. Hentet fra www.civita.no:
<https://www.civita.no/politisk-ordbok/hva-er-etikk>
- Tegmark, M. (2017). *Life 3.0: Being human in the age of Artificial Intelligence*. United States: Knopf.
- Tidemann, A. (2020, 01 08). *Kunstig Intelligens*. Hentet fra SNL:
https://snl.no/kunstig_intelligens
- Tidemann, A., & Elster, A. C. (07, 06 2019). *SNL*. Hentet fra Maskinlæring:
<https://snl.no/maskinlæring>
- Tranøy, K. (2018, 11 08). *Ansvar*. Hentet fra www.snl.no: <https://snl.no/ansvar>
- Vallentyne, P. (2008). Brute luck and responsibility. *Politics, philosophy & economics* .
- Vredenberg, L. (2015, 01 27). *Kognitiv databehandling: Hva er det og hvordan gjør det verden smartere?* . Hentet fra IBM:

<https://www.ibm.com/blogs/think/no-no/2015/01/27/kognitiv-databehandling-hva-er-det-og-hvordan-gjor-det-verden-smartere/>
Yu, K.-H., & Kohane, I. (2018, October 05). *www.qualitysafety.bmj.com*.
Hentet fra Framing the challenges of artificial intelligence in medicine :
<https://qualitysafety.bmj.com/content/qhc/28/3/238.full.pdf>

7.0 Vedlegg

7.1 Vedlegg – Spørsmål til undersøkelsen

Kritikk spørsmål:

- Q1: **Jeg syns Sigurd fortjener kritikk for den inntrufne hendelsen**
- Q2: **Sigurds fremferd kan og bør kritiseres**
- Q3: **Det er ikke riktig å kritisere Sigurd i dette tilfellet**

Moralsk ansvar spørsmål:

- Q4: **Sigurd er moralsk ansvarlig for dette**
- Q5: **Det moralske ansvaret for hendelsen må radiolog Sigurd Rosfjord ta**
- Q6: **Det blir galt å holde Sigurd moralsk ansvarlig for det som skjedde**

Straff spørsmål:

- Q7: **Sigurd Rosfjord fortjener straff for dette**
- Q8: **Det er riktig at Sigurd utsettes for en disiplinær reaksjon etter dette**
- Q9: **Det blir feil å straffe Sigurd for det som skjedde**

Sinne spørsmål:

- Q10: **Jeg ble sint av å lese om denne radiologen**
- Q11: **Det Sigurd gjorde vekket sinne i meg**

Erstatningskrav spm:

- Q12: **De pårørende burde få erstatning fra Sigurd**
- Q13: **De pårørende burde få erstatning fra sykehuset**
- Q14: **De pårørende burde få erstatning fra selskapet som lager iScan3000**

Kontrollspørsmål:

Q15: Kontrollspørsmål: I tilfellet du leste om så tilsa sykehusets retningslinjer at radiologer ikke kunne utføre analysen selv, men måtte benytte seg av dataprogrammet.

Q16: Kontrollspørsmål: I beskrivelsen stod det at menneskelige radiologer gjør flere feil enn iScan3000.

Demografi spørsmål:

Q17: Hvor gammel er du

Q18: Hvilket kjønn har du?

Q19: Hvor mye tjener du i året (i NOK)?

Q20: I hvilken sektor arbeider du?

Q21: Er du leder på din arbeidsplass?

Q22: Hva er din høyest fullførte utdanning?

7.2 Vedlegg – Hypotesetest

Hypotesetest 1

T-Test

Group Statistics					
Kontrollspørsmål: I tilfellet du leste om så tilsa sykehusets retningslinjer at radiologer ikke k...					
		N	Mean	Std. Deviation	Std. Error Mean
Kritikkverdig	Korrekt	56	2.7917	1.56549	.20920
	Ikke korrekt	59	3.9548	1.58472	.20631

Independent Samples Test										
		Levene's Test for Equality of Variances			t-test for Equality of Means				95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Kritikkverdig	Equal variances assumed	.148	.701	-3.957	113	.000	-1.16314	.29391	-1.74543	-.58084
	Equal variances not assumed			-3.959	112.815	.000	-1.16314	.29382	-1.74525	-.58102

Hypotesetest 2

T-Test

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
Moralskansvarlig	Korrekt	56	3.0417	1.69499	.22650
	Ikke korrekt	59	3.9153	1.50044	.19534

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Moralskansvarlig	Equal variances assumed	1.032	.312	-2.930	113	.004	-.87359	.29815	-1.46427	-.28290
	Equal variances not assumed			-2.921	109.696	.004	-.87359	.29910	-1.46635	-.28082

Hypotesetest 3

T-Test

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
Straff	Korrekt	55	1.9333	1.05096	.14171
	Ikke korrekt	59	2.8814	1.44206	.18774

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Straff	Equal variances assumed	10.674	.001	-3.987	112	.000	-.94802	.23778	-1.41916	-.47688
	Equal variances not assumed			-4.030	105.971	.000	-.94802	.23522	-1.41437	-.48167

Hypotesetest 4

T-Test

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
Sint	Korrekt	56	2.5893	1.51989	.20310
	Ikke korrekt	59	2.6864	1.66583	.21687

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Sint	Equal variances assumed	.229	.633	-.326	113	.745	-.09715	.29784	-.68724	.49293
	Equal variances not assumed			-.327	112.829	.744	-.09715	.29713	-.68583	.49152

Hypotesetest 5-7

T-Test

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
De pårørende burde få erstatning fra Sigurd	Korrekt	56	2.13	1.779	.238
	Ikke korrekt	59	2.07	1.363	.177
De pårørende burde få erstatning fra sykehuset	Korrekt	56	5.36	1.873	.250
	Ikke korrekt	58	4.50	1.857	.244
De pårørende burde få erstatning fra selskapet som lager iScan3000	Korrekt	56	4.48	2.089	.279
	Ikke korrekt	59	3.53	1.933	.252

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
De pårørende burde få erstatning fra Sigurd	Equal variances assumed	3.250	.074	.194	113	.846	.057	.295	-.527	.641
	Equal variances not assumed			.193	103.017	.847	.057	.297	-.531	.646
De pårørende burde få erstatning fra sykehuset	Equal variances assumed	.095	.759	2.454	112	.016	.857	.349	.165	1.549
	Equal variances not assumed			2.453	111.785	.016	.857	.349	.165	1.549
De pårørende burde få erstatning fra selskapet som lager iScan3000	Equal variances assumed	.521	.472	2.551	113	.012	.957	.375	.214	1.700
	Equal variances not assumed			2.546	111.126	.012	.957	.376	.212	1.701