



Handelshøyskolen BI - campus Bergen

BTH 14111

Bacheloroppgave - Human Resource
Management

Bacheloroppgave

Bachelor 2019

Navn: Mari Mobæk Ulstad, Charlotte Larsen

Utlevering: 07.01.2019 09.00

Innlevering: 03.06.2019 12.00

Bacheloroppgave
ved Handelshøyskolen BI Bergen

Moralsk ansvar når kunstig intelligens gjør feil i kirurgi

Innleveringsdato: 03.03.2019

Eksamenskode: BTH1411



(INTUITIVE, 2019)

“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”.

Sammendrag

Hensikten med vår oppgave er å undersøke hvordan moralsk ansvar oppfattes i en situasjon der kunstig intelligens blir benyttet.

Vi vil først ta for oss relevante teorier og forskning rundt temaene kunstig intelligens og moralsk ansvar. Disse fører til vår problemstilling “hvordan oppfattes moralsk ansvar når en kunstig intelligens gjør feil i kirurgi”. Deretter presenteres hypotesene vi vil bruke for å belyse problemstillingen. Hypotesene tar for seg moralsk ansvar, kritikkverdighet, straff, sinne og erstatningsansvar.

Vi har benyttet kvantitativ metode, og utført et eksperiment for å vise forskjellen mellom våre to grupper. Alle respondentene leste om en kirurgirobot som kunne utføre operasjoner med mindre risiko enn en menneskelig kirurg. Begge grupper leser deretter at roboten utfører en fatal feil på pasienten. Respondentene ble så randomisert i to grupper og presentert med forskjellige situasjoner. Den ene gruppen fikk lese om en kirurg som kunne velge mellom å utføre inngrepet selv, men valgte roboten. Den andre gruppen fikk lese om at kirurgen ikke kunne velge, da sykehuset krevde at roboten skulle utføre det. Deretter ble det stilt en rekke spørsmål om situasjonen. Fra våre analyser kommer det frem at kirurgen får tildelt mer moralsk ansvar i gruppen der han kunne velge å utføre operasjonen selv.

Svarene fra undersøkelsen blir analysert for å gi oss større innsikt i dataene. Dette brukes videre i diskusjonen der vi går dypere inn i hvordan resultatene reelt kan påvirke samfunnet og arbeidsplasser i fremtiden.

Forord

Denne bacheloroppgaven er vår avslutning på tre fantastiske år på Handelshøyskolen BI Bergen, og er en del av fordypningen i kurset Human Resource Management.

Høsten 2018 presenterte vår foreleser, førsteamanuensis Mads Nordmo Arnestad, aktuelle eksperimenter for bacheloroppgaver. Kunstig intelligens i kirurgi fanget vår oppmerksomhet raskt, og vi begynte forskningsprosessen med Mads sitt design senere samme høst.

Prosessen med å skrive oppgaven har vært utfordrende, men svært lærerikt og spennende. Vi har arbeidet med nye temaer og med nye metoder. I løpet av de siste månedene har vi fått mye innsikt i et område vi tror vil fortsette å øke i aktualitet fremover.

Temaene og oppgaven har vært krevende, da vi måtte sette oss inn i helt nye fagfelt. Dette hadde ikke vært mulig uten vår veileder Mads, som har gitt uerstattelig rådgivning i situasjoner der vi stod fast. Engasjementet hans for temaet var smittsomt, og har gjort det givende å jobbe med oppgaven.

Vi er veldig takknemlige for alle våre respondenter som tok seg tid til å svare på undersøkelsen. Uten dere hadde ikke oppgaven vært mulig å gjennomføre. Vi håper at du som leser oppgaven vil finne den like spennende og lærerik som oss!

Bergen, Våren 2019

Innholdsfortegnelse:

1.0	TEORI OG HYPOTESER.....	1
1.1	<i>Hva er kunstig intelligens?</i>	1
1.1.2	Kunstig intelligens i helsevesenet	3
1.1.2	Algoritme-aversjon.....	4
1.2	<i>Opplevelse av moralsk ansvar</i>	5
1.3	<i>Opplevd moralsk ansvar for KI som begår feil</i>	7
1.4	<i>Valg av problemstilling og hypoteser</i>	8
2.0	METODE.....	10
2.1	<i>Forskningsdesign</i>	10
2.2	<i>Seleksjon</i>	10
2.3	<i>Eksperimentets design</i>	10
2.4	<i>Utfallsmål</i>	11
2.5	<i>Statistiske analyser</i>	13
3.0	RESULTAT.....	14
3.1	<i>Deskriptive data</i>	14
3.1.1	Kjønn og antall respondenter	14
3.1.2	Alder.....	15
3.1.3	Offentlig eller privat sektor	16
3.1.4	Høyeste fullførte utdanning.....	17
3.1.5	Inntekt.....	17
3.1.6	Leder.....	18
3.2	<i>Cronbachs alfa</i>	19
3.2.1	Kritikkverdig	19
3.2.2	Moralsk ansvar	19
3.2.3	Straff.....	20
3.2.4	Sinne.....	20
3.3	<i>Korrelasjonsanalyser</i>	20
3.4	<i>Mellomgruppeskjeller</i>	21
4.0	DISKUSJON	28
4.1	<i>Hovedfunn</i>	28
4.2	<i>Teoretiske implikasjoner</i>	28
4.3	<i>Implikasjoner for arbeidsliv</i>	31
4.4	<i>Utfordringer</i>	32
4.4.1	Implikasjoner for ledere	33
4.5	<i>Begrensninger og videre forskning</i>	34
5.0	KONKLUSJON	36
6.0	BIBLIOGRAFI	37

1.0 Teori og hypoteser

1.1 Hva er kunstig intelligens?

“AI[kunstig intelligens] currently encompasses a huge variety of subfields, ranging from the general (learning and perception) to the specific, such as playing chess, proving mathematical theorems, writing poetry, driving a car on a crowded street, and diagnosing diseases. AI is relevant to any intellectual task; it is truly a universal field” (Russell & Norvig, 2016, s. 1)

Definisjonen på kunstig intelligens er kontroversiell (Hanson & Marshall, 2001), og ulike definisjoner benyttes av ulike felt (Russell & Norvig, 2016). En generell definisjon kan være at kunstig intelligens (heretter KI) kan forklares som de menneskeskaptede systemer som kan replikere menneskelige prosesser knyttet til intelligens (Shapiro, 1992).

Alan Turing utviklet sin “Turing-test” i 1950, for å tilby en operasjonell definisjon på intelligens hos maskiner (Russell & Norvig, 2016). Testen gikk ut på at en computer og et menneske skulle svare på skrevne spørsmål. I hans øyne hadde en computer kunstig intelligens dersom et menneske ikke kunne se forskjell på computerens svar, og svar skrevet av et annet menneske (Hanson & Marshall, 2001). Denne testen ble videreutviklet til den såkalte “Totale Turing-testen”, der et videosignal er inkludert for å teste subjektene perseptuelle ferdigheter.

For å bestå den Totale Turing-testen, vil en computer behøve de følgende seks egenskapene. Russell, Norvig et. al nevner i sin bok (s. 2-3, 2016) at mesteparten av KI utgjøres av seks egenskaper:

- Naturlig språkprosessering: for å kunne kommunisere suksessfullt
- Kunnskapsrepresentasjon: for å lagre den den ser/hører
- Automatisert resonnering: for å bruke den lagrede informasjonen til å respondere og trekke nye konklusjoner

- Maskinl ring: for   tilpasse seg nye omstendigheter og gjenkjenne m nstre
- Computer-sansing: for   sanse objekter
- Robotikk: for   kunne bevege seg og manipulere objekter

For   utvikle KI i en maskin, m  de f rst l res opp ved bruk av maskinl ring. Nesten alle maskinl ringsoppgaver kan forst s som at maskinen trekker slutninger fra manglende eller latent data fra det observerbare data (Ghahramani, 2015). Man kan dele opp l reprosessene i to: top-down l ringssystemer og bottom-up l ringssystemer (Hanson & Marshall, 2001). I top-down l ring mates computeren med en komplekse problemer, som brytes ned til underkomponenter. I bottom-up l ring mates computeren med store mengder data, som den trekker egne sammenhenger ut fra, s kalt “data mining” (Witten, Eibe & Hall, 2011).

For at en maskin skal l re fra data mining, m  den trekke slutninger fra dataene. For   trekke slutninger, m  maskinen f rst fatte noen antagelser. Disse antagelser settes sammen til   bygge opp en modell. Disse modellene varierer fra   v re relativt simple, til   bli sv rt store og dype nevrale nettverk (Ghahramani, 2015). Modellene som benyttes i dag, bygges nesten alltid opp av slike nettverk.

“Et kunstig nevralt nettverk består vanligvis av sv rt mange prosesseringsenheter, satt sammen i et multi-lags m nster av parallelle koblinger” (Papadatou, 2011 [v r oversettelse])

Enhetene finnes i tre klasser: input-enheter, output-enheter og skjulte enheter. Informasjon passerer gjennom alle lagene, og hvert lags kalkulasjoner baserer seg for forrige lags (Jain, Mao & Mohiuddin, 1996). Disse nettverkens oppbygning er basert p  v r hjerne, i at hjernen v r benytter seg av en rekke parallelle systemer som jobber sammen for   ta beslutninger (Roy, 2000). Kunstige nevrale nettverk egner seg godt til generaliserte oppgaver, samt mer spesifikk gjenkjenning og bildetolkning. De kan ogs  l res opp til   bli spill-mestre, og har sl tt mennesker i sjakk, othello, dam, scrabble og backgammon (Silver, et al, 2016).

Et problem med bruken av kunstige nevralt nettverk er at det kan være vanskelig eller umulig å se hvordan nettverket kom frem til output-et. Dette skyldes de skjulte enhetene, og at man ikke kan se inter-relasjonene til nevronene i nettverket (Olden & Jackson, 2002).

1.1.2 Kunstig intelligens i helsevesenet

Innen medisin er det vanlig å dele kunstig intelligens inn i to områder, det virtuelle og det fysiske. Det virtuelle dreier seg hovedsakelig om informasjonssystemer o.l. som kan hjelpe medisinsk personell i deres behandlinger av pasienter. Det fysiske aspektet omfatter roboter, som assisterer personell i å utføre behandlingene (Tremblay, Hamet, 2017).

1.1.2.1 Virtuell gren

Maskinlæring benyttes for å gi helsesektoren større adaptiv evne, og implementeres i mange forskjellige systemer. Disse systemene dreier seg hovedsakelig om tre forskjellige funksjoner ved maskinlæring (Tremblay, Hamet, 2017):

1. Evnen til å finne mønstre i datasamlinger
2. Evnen til å klassifisere mønstre slik at prediksjoner kan utvikles
3. Læring ved forsterkning, som går ut på å finne løsninger på spesifikke problemer ved å la maskinen prøve og feile.

Disse funksjonene kan implementeres i mange forskjellige deler av helsesektoren, og dette er i utvikling hos mange helseorganisasjoner allerede. Sørlandet Sykehus har prøvd ut slik maskinlæring for å hente ut relevant sykdomsinformasjon fra pasientjournaler, og viser til at dette kan spare dyrebar tid på akuten (Christiansen, 2017).

Maskinlæring er en essensiell komponent av effektive KI-systemer, og kan blant annet brukes til diagnostisk tolkning av bilder. KI har i tester vist evnen til å diagnostisere hjertefeil i EKG på samme nivå som profesjonelle kardiologer (Hannun, Rajpurkar et. al, 2019). KI har også blitt testet i gjenkjenning av lungekreft i CT-scans. Her klarer computeren også å prestere på et parallelt eller bedre nivå enn profesjonelle (Grady, 2019). Ved å bruke maskinlæring til

mammografitolkning, kan brystkreft oppdages betydelig tidligere enn uten hjelp fra KI (Yala et. al, 2019).

En potensiell begrensning ved maskinlæring, er at systemet er avhengig av tilgang til ny data for å fortsette å lære. Innen helsesektoren forventes stor fremgang i en type data mining, såkalt “Natural Language Processing”, som går ut på at maskinen tolker menneskeskapt data, som deretter implementeres i læringsprosessen (Jiang F, Jiang Y, Zhi H, et al, 2017). Pasientrapporter og kliniske notater fra leger i fortiden kan derfor supplere diagnoseprosessen i fremtiden på en svært effektiv måte.

1.1.2.2 Fysisk gren

Det globale markedet for menneskeassistert robotkirurgi ble verdsatt med 56,294.9 millioner USD i 2017, og er forventet å vokse til ca 100 millioner USD innen 2024 (Sajeev & Chandra, 2018). Disse maskinene opererer ikke selvstendig, men overfører bevegelsene fra kirurgene som styrer dem. Benyttelse av robotkirurger, kan betydelig redusere problemer knyttet til menneskelige begrensninger. Roboter i helseorganisasjoner er ikke svært utbredt ennå, men forventes å implementeres i større grad i årene fremover (Anandan, 2018).

Roboter benyttes allerede i norsk kirurgi. 3. September 2018 utførte Ahus sin første robotassisterte operasjon på en prostatakreftpasient (Ahus, 2018). Roboten, som er en DaVinci Xi, skal også brukes til å behandle urologisk kreft og nyrekreft, i tillegg til prostatakreft. Den forventes også å benyttes til andre formål i fremtiden. Det er viktig å påpeke at denne roboten kun overfører kirurgens egne bevegelser, og at det er kirurgen som tar alle medisinske avgjørelser under operasjonen.

Innen hårtransplantasjon finnes ARTAS, som kan kalkulere hvilke hårsekker som er optimale for at transplantasjonen skal være vellykket, samt transplantere disse på pasienten. Helsepersonell bestemmer hvor det nye hårfestet skal være, men roboten opererer ellers svært selvstendig (ARTAS, 2019).

1.1.2 Algoritme-aversjon

Algoritme-aversjon er en betegnelse for et menneskelig fenomen der folk velger å ikke benytte algoritmer til beslutninger dersom de oppdager at ikke er perfekte, til

tross for at algoritmene er bedre til å predikere utfall enn mennesker. Denne effekten forsterkes dersom feiltrinnene blir direkte observert av personene (Dietvorst, Simmons & Massey, 2016).

I mange tilfeller stoler mennesker heller på såkalt “ekspertise-intuisjon”, som bygger på en tro om at erfaring på et område vil gjøre personer bedre til å bedømme utfall. Denne troen på ekspertene overkjører evnen til å se utfall på et rent statistisk vis (Highhouse, 2008). Dette er ofte tilfellet i situasjoner der sterke emosjoner vekkes. Menneskene fokuserer da på uønskede utfall over sannsynligheten for dette utfallet (Sunstein, 2002). Dette kan gjøre at de velger bort statistisk ideelle løsninger fordi de ikke klarer å godta risikoen for negative utfall.

Villigheten til å benytte uperfekte algoritmer øker betydelig dersom man får muligheten til å modifisere dem, selv om modifikasjonen er svært liten. Selv ved tilfeller der modifikasjonene resulterer i en algoritme som gjør flere feil, foretrekker man denne så lenge man beholder evnen til modifikasjon (Dietvorst, Simmons, et al, 2015).

1.2. Opplevelse av moralsk ansvar

“The principles that people should be held personally responsible for the consequences of their choices is a fundamental moral ideal in the Western societies, but the interpretation and application of this principle have been a heated political issue for centuries” (Greenfield, 2011).

Det finnes ingen fasit for hva følelsen av moralsk ansvar kommer av. Av denne grunn presenteres noen ulike teorier rundt dette.

Ifølge Galen Strawson(1994) er det et grunnleggende argument som viser seg å bevise at ingen kan helt bli holdt moralsk ansvarlig for deres handlinger. Dette går ut på at; (1) Ingenting kan være *casa sui*, altså at ingenting kan være årsaken til seg selv. (2) For å være virkelig moralsk ansvarlig for ens handlinger, må en være *casa sui* ihvertfall i viktige avgjørende mentale hensyn. (3) Derfor kan ingenting være virkelig moralsk ansvarlig.

I praksis er ikke denne tankegangen vanlig å forholde seg til, da vi lever i et samfunn der moralsk ansvar spiller en stor rolle i hverdagen. Andre teorier baserer seg på at moralsk ansvar finnes, og at det er en integrert del av menneskers tankegang. Teoriene som følger, er av denne tankegangen.

Aristoteles er en av de første som kom ut med en teori om moralsk ansvar, han mente at det noen ganger kunne være riktig å tildele ros eller ris på grunnlag av ens handlinger. Han nevner også at moralske personer er underlagt ansvarsfraskrivelse fordi de har kapasitet til å kunne bestemme hvilket valg de tar. Han mente at en kunne bare få ros eller ris dersom handlingen hadde vært frivillig og ikke påtvunget.

Aristoteles mente at en handling har to karakteristiske trekk. Det første er en kontroll tilstand; handlingen må ha sin opprinnelse i personen, altså at det er opp til personen om å utføre denne handlingen eller ikke, den kan ikke bli tvunget på fra noen andre. Det andre er en epistemisk tilstand; altså at personen må være klar over hva det er han gjør eller får med seg av informasjon.

Forskere har kommet frem til to konkurrerende syn på begrepet moralsk ansvar, hvor den første er et meritbasert syn; altså hvilken ros eller skyld ville vært en passende reaksjon mot personen hvis, og da bare hvis han fortjener det. Det andre synet er det konsekvensetiske synet; altså hvilken ros eller skyld ville vært hensiktsmessig hvis og da bare hvis en slik reaksjon ville ført til en ønsket forandring eller endring av oppførsel (Eshleman, 2014). Forskjellen ligger altså i at man tar personens indre motivasjon i betraktning, eller konsekvensene av personens handling.

Hva mennesker mener er et moralsk relevant valg kan komme an på om det er et valgfritt eller tvunget valg. Det er velkjent at mennesker i forskjellige situasjoner holder individer personlig ansvarlig for valgene deres (Cappelen, Fest et al, 2016), men det er ikke alltid at nominelle og tvungete valg møter minstekravet for vilkårene for når vi burde holde noen personlig ansvarlige. Dette gjelder da spesielt hvis en person ikke kunne ha endret utfallet av valget hvis et annet valg ble tatt. Om mennesker oppfatter et valg som fritt eller tvunget, blir påvirket av graden emosjoner som vekkes, altså oppfattet alvorlighetsgrad av situasjonen (Nichols & Knobe, 2007).

Hvordan mennesker ser på en person som fri eller tvunget til å ta en farlig jobb er med på å bestemme om de holder ham personlig ansvarlig for konsekvensene av å ta den jobben. Det påvirker også om de mener at han fortjener hjelp hvis han havner i en dårlig situasjon (Greenfield, 2011).

Ifølge Savani & Rattan (2012) er det slik at ved å påpeke konseptet av valgene som er mulig, gjør at mennesker blir mindre støttende til sosialstønad på grunn av at livssituasjonen din skyldes dine tidligere valg.

1.3. Opplevd moralsk ansvar for KI som begår feil

Evnen til å velge er underliggende for alle synspunkter på moralsk ansvar. Det er derfor ikke grunnlag til tildele ansvar til en KI, da det ikke er vanlig å tenke at disse har fri vilje til å ta valg. KI er i dag å ses på som et verktøy, brukt av mennesker som besitter kontroll. I dagens robotassisterte operasjoner har ikke roboten autonomi til å utføre handlinger uten eksplisitt godkjennelse fra mennesker. Dette forventes å endre seg i fremtiden, i at robotene vil operere mer og mer selvstendig (Ahus, 2018; Anandan, 2018). Dette kan tenkes å komme på bekostning av menneskelig kontroll over robotene. EU har i dag kommet med egne etiske regler om KI innen ulike sektorer og deres bruk i fremtiden. De mener at KI burde overholde de grunnleggende etiske prinsippene om respekt for menneskets autonomi, forebygging av skade, rettferdighet og ansvarlighet (European Commission, 2019).

KI fungerer allerede i dag på nivå med mennesker innen flere felt, og forventes å overgå menneskelig kompetanse i fremtiden (Dietvorst, Simmons & Massey, 2016; Silver, et al, 2016; Grady, 2019; Hannun, Rajpurkar et. al, 2019). I mange tilfeller gjør KI mindre feil enn mennesker, og er derfor et statistisk bedre valg over menneskelig utførelse av diverse oppgaver. En potensiell hindring for at disse overlegne KI tas i bruk, er at mennesker viser mistillit til uperfekte KI, selv om de feiler sjeldnere enn mennesker (Dietvorst et al, 2015).

I vår oppgave baserer vi oss på tanken om at moralsk ansvars plassering er avhengig av at aktørens valg oppfattes som fritt (Greenfield, 2011; Savani &

Rattan, 2012). Ettersom kirurgi er et såpass alvorlig og uvanlig tema, kan emosjoner forventes å spille inn på om valgets oppfattelse (Nichols & Knobe, 2007). I tillegg vet vi at mennesker ikke alltid klarer å tenke rasjonelt rundt algoritmer og KI, i at de undervurderer evnene til KI og overvurderer evnene til menneskelige eksperter (Highhouse, 2008).

Forskning på algoritme-aversjon viser at personer får mer tillit til når de har evnen til å modifisere den på noen som helst måte. Dette kan potensielt kompensere for mistilliten. Dette ønsket for modifikasjon overkjører evnen til å rasjonelt forstå at det statistiske korrekte valget er å la algoritmen stå uendret (Dietvorst, Simmons, et al, 2016).

Helseorganisasjoner forventes å basere seg i større grad på statistikk og fakta over menneskelige emosjoner i sine pasientbehandlinger. Norske leger er pålagt å unngå metoder som medfører unødige fare (Den Norske Legeforening, 2015). Dersom KI utfører oppgaver med kategorisk mindre risiko enn mennesker, forventes KI å konsekvent velges over mennesker til å utføre oppgavene.

Vi ønsker altså kontroll i en så stor grad at vi lar det gå utover suksessraten til tingen vi kontrollerer. Roboter er i dag ikke helt selvstendige, men forventes å få mer autonomi i fremtiden. Når man utfører handlinger som er fullstendig i ens kontroll, vil det moralske ansvaret for handlingen ligge helt hos personen som utførte den. I en fremtid der kontrollen over KI er minimal, kan ansvars plasseringen bli mindre klar. Hvem vil ende opp med det moralske ansvaret i en situasjon der roboten utfører oppgaven uten menneskelig innblanding?

Dette bringer oss til vår problemstilling.

1.4 Valg av problemstilling og hypoteser

Hovedproblemstillingen i denne oppgaven er “hvordan oppfattes moralsk ansvar når en kunstig intelligens gjør feil i kirurgi”. Vi vil undersøke om det å ha en valgmulighet vil påvirke hvor mye moralsk ansvar man får, selv om det andre valgalternativet er dårligere. Kritikverdighet, straffverdighet og sinne tenkes å

være nært relatert moralsk ansvar, og vil derfor belyses. Vi ønsker å undersøke dette i en situasjon der en robot har operert selvstendig og begått en feil.

Vi har også fokus på hvordan KI og moralsk ansvar virker inn på organisasjoner, og hvilke muligheter og utfordringer det fører med seg for fremtidens arbeidsplass. Vi har presentert teori fra to hovedfelt, som begge er relevante for å belyse vårt eksperiment. Disse er kunstig intelligens og moralsk ansvar. Vårt forskningsspørsmål og hypoteser er dannet på grunnlag av disse teoriene.

Forskningsspørsmålet vårt er: Blir kirurgen sett på som mer moralsk ansvarlig for feilen dersom han har et valg?

H1) Kirurgen vil oppfattes som mer moralsk ansvarlig for feilen som inntreffer, dersom han har muligheten til å utføre inngrepet selv.

H2) Kirurgen vil oppfattes som mer kritikkverdig for feilen som inntreffer, dersom han har mulighet til å utføre inngrepet selv.

H3) Kirurgen ville oppfattes som mer straffverdig for feilen som inntreffer, dersom han har mulighet til å utføre inngrepet selv.

H4) Robotens feiltrinn vil vekke mer sinne hos respondentene, dersom kirurgen har muligheten til å utføre inngrepet selv.

H5) Flere vil se kirurgen som erstatningsansvarlig ovenfor de pårørende, dersom han har muligheten til å utføre inngrepet selv.

H6) Flere vil se sykehuset som erstatningsansvarlig for feilen som inntraff, dersom kirurgen *ikke* har muligheten til å utføre inngrepet selv.

H7) Flere vil se robotens produsent som erstatningsansvarlig for feilen som inntreffer, dersom kirurgen *ikke* har mulighet til å utføre inngrepet selv.

2.0 Metode

2.1 Forskningsdesign

Vi samlet inn våre primærdata til denne kvantitative studien ved bruk av et eksperiment, utviklet av Mads Nordmo Arnestad ved Handelshøyskolen BI Bergen. Eksperimenter involverer direkte manipulering og kontroll av variabler, for deretter å observere responsen (Cozby & Bates, 2012). Dataene ble samlet inn med en spørreundersøkelse som ble utviklet elektronisk med Qualtrics. Alle deltakere anonymiseres før videre bruk av data. Dataene fra spørreundersøkelsen viderebehandles med statistisk analyse i SPSS, og resultater av analysene benyttes for å støtte eller undertrykke våre hypoteser.

2.2 Seleksjon

Vi ønsket å si noe om nordmenns holdning til moralsk ansvar og KI. Vår populasjon er “nordmenn flest”, og med dette mener vi voksne i arbeid. Vi har derfor forsøkt å unngå enkelte undergrupper, slik som studenter og pensjonister, da disse tenkes å ikke ha overførbare holdninger til resten av befolkningen. Utvalget er et bekvemmelighetsutvalg der det har blitt tatt hensyn til arbeidssituasjon. Undersøkelsen ble sendt til personer i arbeid, og vi forsøkte til vår beste evne å unngå svar fra studenter uten jobb.

Eksperimentet ble distribuert online til et utvalg av mennesker på rundt 220 stykker, der 121 fullførte undersøkelsen. Vi fikk de fleste respondentene våre ved å ta kontakt med organisasjoner i ulike sektorer, og hadde fokus på å få respondenter med forskjellig bakgrunn. I hovedsak ble det brukt mail for å sende ut spørreundersøkelsen, men Facebook Messenger ble også brukt i innspurten for å få de siste svarene som trengtes.

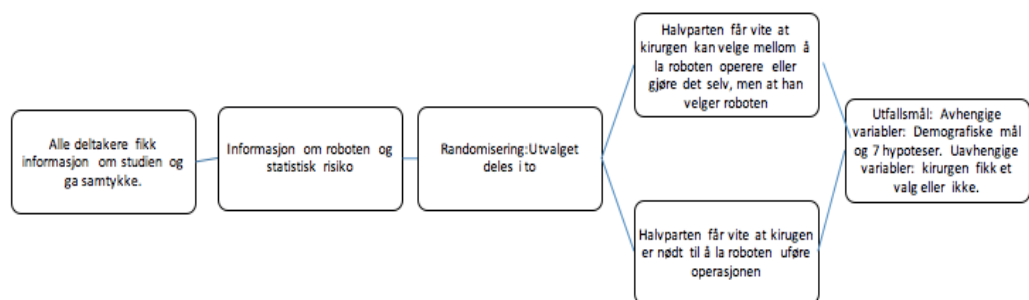
2.3 Eksperimentets design

Det finnes ingen standardiserte mål for moralsk ansvarsfølelse. I undersøkelsen ble det derfor utviklet egne mål på dette. Begrepsvaliditet består av to viktige

forhold: konvergent validitet og diskriminant validitet. Konvergent validitet går ut på at indikatorer som skal måle den samme teoretiske variabelen er høyt korrelert med hverandre (Gripsrud, Olsson, Silkoset, 2017). Det ble inkludert flere mål på samme fenomen i undersøkelsen, for å øke sannsynligheten for at vi får målt det vi faktisk ønsker å vite noe om, altså øke den konvergente validiteten. Intern konsistens på målene testet vi ved å bruke Cronbachs alfa.

Først i eksperimentet blir respondentene bedt om sitt samtykke. Deretter får de en kort beskrivelse av robotens funksjoner og sannsynlighet for at den begår feil. Det kommer tydelig frem at roboten gjør færre feil enn mennesker (roboten gjør feil i 1/30 tilfeller, mennesker i 1/10 tilfeller). Roboten opererer og gjør en fatal feil på pasienten. Deretter blir respondentene randomisert inn i to grupper. Den ene gruppen får vite at kirurgen Sigurd ikke kan utføre denne operasjonen selv, og må la roboten gjøre det. Den andre gruppen får vite at kirurgen kan velge mellom å utføre operasjonen selv, eller la roboten gjøre det. Her velger kirurgen å la roboten operere. Resten av undersøkelsen er likt for begge grupper. Det er inkludert et kontrollspørsmål for å sikre at respondenten har fått med seg teksten. Resten av spørsmålene er holdningsmål om situasjonen og demografiske spørsmål. Vi er interessert i om det at han hadde valgmuligheten vil påvirke hvor ansvarlig respondentene mener han er for feilen.

2.4 Utfallsmål



Figur 1: Forskningsdesign

Eksperimentets spørsmål er basert på å måle om det er kritikkverdig, om det skal straffes og om sinne eksisterer innen bruk av manual mode eller ingen valg. Disse holdningene har blitt målt ved bruk av en Likert-skala i SPSS som går fra 1-7, hvor 1 er helt uenig, 4 er nøytralt og 7 er helt enig. For å få en samlet holdning måtte noen av tallverdiene snus for at det skulle bli riktig. Likert-skalaen blir brukt for å teste om det er flere respondenter som svarer likt eller nesten likt, hvis det er slik så er de ikke egnet til å måle de forskjellene som finnes i holdningene og blir dermed tatt ut.

I eksperimentet er det blitt brukt Cronbachs alfa for å sjekke skalaens reliabilitet, dette for å sjekke at spørsmålene faktisk måler det vi er ute etter. Cronbachs alfa bør ligge på over 0,6 da dette er eksperimentelle data der to av gruppene har fått forskjellig beskjed. Reliabiliteten bør ikke være for nært 1 for å forhindre at indikatorene er for like hverandre og derfor ikke kan fange opp “hele” begrepet (Gripsrud et. al, 2016, s. 215).

I eksperimentets analyser opererer vi med to uavhengige variabler: “valg” og “ingen valg”, disse blir målt opp mot de avhengige variablene som er demografien og hypoteser.

Spørsmålene Q13, Q14 og Q15 måler H1: Moralsk ansvar og om kirurgen vil oppfattes som mer moralsk ansvarlig dersom han utfører inngrepet selv. Spørsmål Q8, Q10 og Q11 måler H2: Kirurgen oppfattes som mer kritikkverdig, dersom han utfører inngrepet selv.

Deretter har vi spørsmålene Q16, Q17 og Q18 som måler H3: Kirurgen oppfattes som mer straffverdig dersom han utfører inngrepet selv.

Spørsmålene Q25 og Q26 måler H4: Om det vekkes mer sinne dersom roboten gjør feil og kirurgen har mulighet til å utføre inngrepet selv.

Det ble også brukt enkeltstående variabler for å måle de tre siste hypotesene, dette for å finne ut fra respondentene hvem de mente var erstatningsansvarlig.

Spørsmålet Q21 ble brukt for å måle H5: Om det er kirurgen som er erstatningsansvarlig. Spørsmål Q22 ble brukt for å måle H6: Om det er sykehuset

som er erstatningsansvarlig. Og spørsmål Q23 ble brukt for å måle H7: Om det er produsenten av roboten som er erstatningsansvarlig.

Variabel (Hypotese)	Utvalg	Cronbach's Alpha (α -verdi)
Moralsk ansvar (H1)	Q13, Q15, Q14 (Q14 måtte snus)	.928
Kritikkverdigg (H2)	Q8, Q10, Q11	.908
Straffverdigg (H3)	Q16, Q17, Q18 (Q18 måtte snus)	.902
Sinne (H4)	Q25, Q26	.954
Kirurg erstatningsansvarlig (H5)	Q21	.
Sykehuset erstatningsansvarlig (H6)	Q22	.
Produsenten av roboten erstatningsansvarlig (H7)	Q23	.

Tabell 1: Holdningsmål

2.5 Statistiske analyser

Behandlingen av våre data har blitt gjort i SPSS, mens grafene og tabellene som er brukt er blitt konstruert i Excel. I SPSS har det blitt utført Cronbachs alfa-analyser for å sjekke skalaens reliabilitet til måleparameterne, dette er beskrevet over i utfallsmål. Det ble også utført uavhengige t-tester for å teste forskjellene mellom om “Valg” og “ingen valg” blant hypotesene. “Enkel hypotesetesting består i å teste én populasjonskoeffisient Beta om gangen, i motsetning til multipl hypotesetesting som består i å teste verdien til flere Betaer om gangen” (Sucarrat, 2017).

Flere t-tester ble utført på hele respondentgruppen, men de var delt inn i to grupper; En med “Valg” og en med “ingen valg”.

- Den første gikk ut på moralsk ansvar hos kirurgen.
- Den andre gikk ut på kritikkverdighet hos kirurgen.
- Den tredje gikk ut på straffeverdighet hos kirurgen.
- Den fjerde gikk ut på sinne rettet mot kirurgen.
- Den femte gikk ut på om de pårørende burde få erstatning fra kirurgen.
- Den sjette gikk ut på om de pårørende burde få erstatning fra sykehuset.
- Den syvende gikk ut på om de pårørende burde få erstatning fra produsenten av roboten.

Det ble også utført en Pearsons korrelasjonsanalyse for å se om det var en sammenheng mellom de syv hypotesene i seg selv og opp mot demografiske variabler som alder, utdanning og lønn. Valget av Pearsons korrelasjonsanalyse er på grunnlag av Likertskalaen. Dette er fordi med Likertskalaen så får vi dataene med like store intervaller fra 1-7, og når de da måles på ordinalnivå så er det beste å bruke en Pearsons korrelasjonsanalyse.

3.0 Resultat

I resultatdelen ønsker vi å drøfte de ulike resultatene fra eksperimentet opp mot vår problemstilling: “hvordan vi tenker om moralsk ansvar når kunstig intelligens gjør feil i kirurgi”.

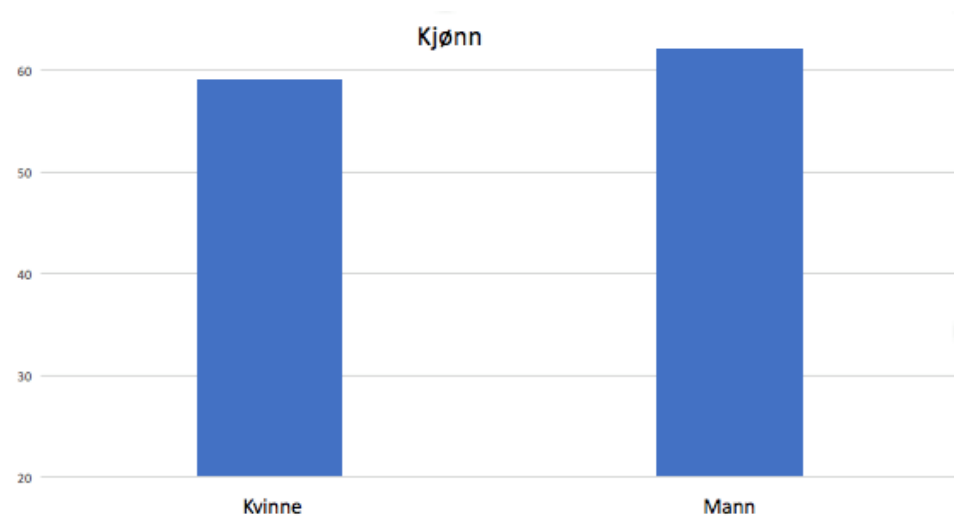
Vi utførte vårt eksperiment på et utvalg nordmenn i alderen 20-75 i arbeid, halvparten fikk en situasjon der kirurgen kunne velge å utføre operasjonen selv, den andre halvparten fikk vite at han ikke hadde noe valg. Dette var altså vår uavhengige variabel. Spørsmålene var ellers identiske for de to gruppene. Resultatene fra eksperimentet ble behandlet med SPSS, der vi utviklet variabler og verdier vi kunne bruke til å teste hypotesene våre. Disse analysene skal vi nå presentere.

3.1 Deskriptive data

For å gi en oppsummering på våre funn, har vi valgt å gjøre en deskriptiv analyse av data, med fokus på de demografiske målene. Vi har gjort en frekvensanalyse på kjønn, alder, om de er ledere på sin arbeidsplass, privat eller offentlig sektor, høyeste utførte utdanning, og hvor mye de tjener.

3.1.1 Kjønn og antall respondenter

Antall respondenter og kjønn. I vår undersøkelse var de demografiske variablene til slutt, så enkelte av holdningsmålene har flere svar enn de demografiske målene.

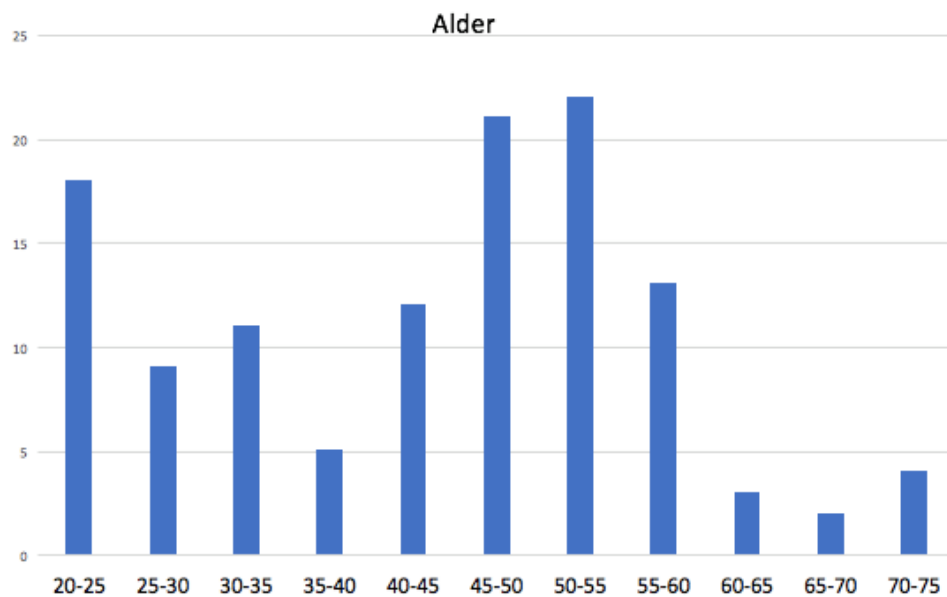


Figur 2: Kjønnfordeling

Vi sendte undersøkelsen ut til ca. 220 personer, og av disse fikk vi 151 respondenter. 30 av disse måtte fjernes grunnet signifikante missing values. Vi endte da opp med 121 respondenter, der 59 (49%) var kvinner og 62 (51%) var menn. Dette er en svært jevn kjønnsfordeling, der begge kjønn er godt representert.

3.1.2 Alder

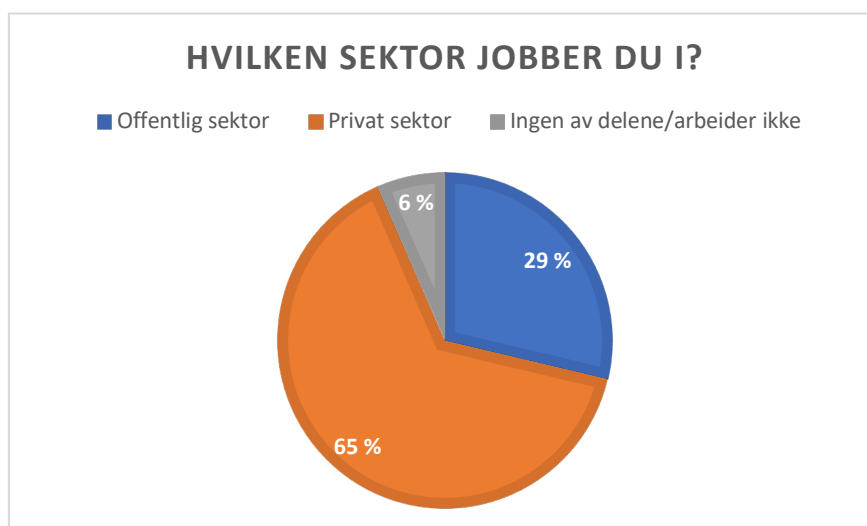
Gjennomsnittlig alder for vårt utvalg er 43,033 år, og standardavviket er på 13,588, som viser en høy spredning av alder. Vår yngste representant var 20 år, vår eldste 74. Dette viser en høy spredning, noe vi ønsket for at utvalget skulle være mest mulig representativt.



Figur 3: Aldersfordeling

3.1.3 Offentlig eller privat sektor

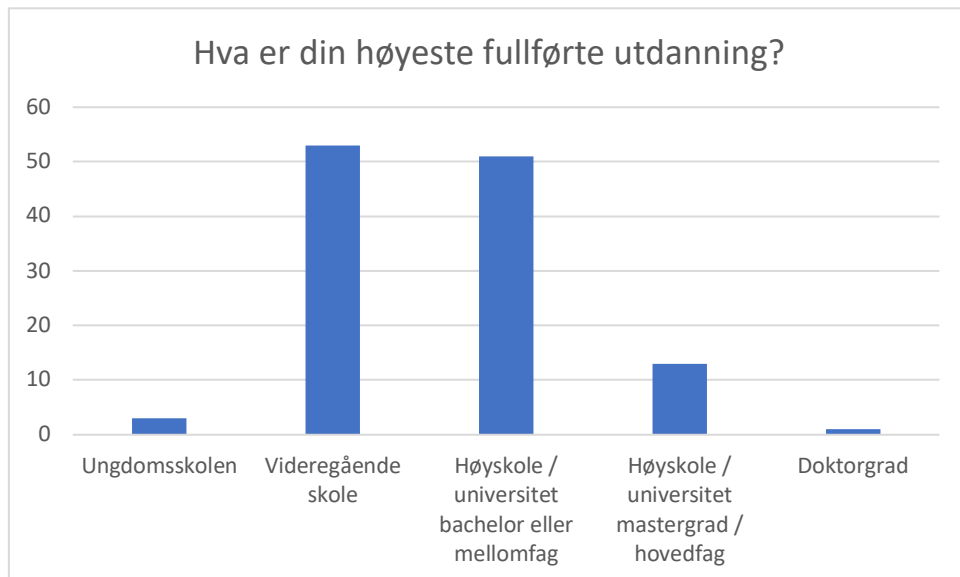
79 (65%) av våre respondenter jobber i privat sektor, 35 (29%) jobber i offentlig. 8 (7%) svarte at de ikke jobber i verken offentlig eller privat, eller at de ikke arbeidet. Vår fordeling er relativt lik fordelingen i hos nordmenn flest.



Figur 4: Privat eller offentlig sektor

3.1.4 Høyeste fullførte utdanning

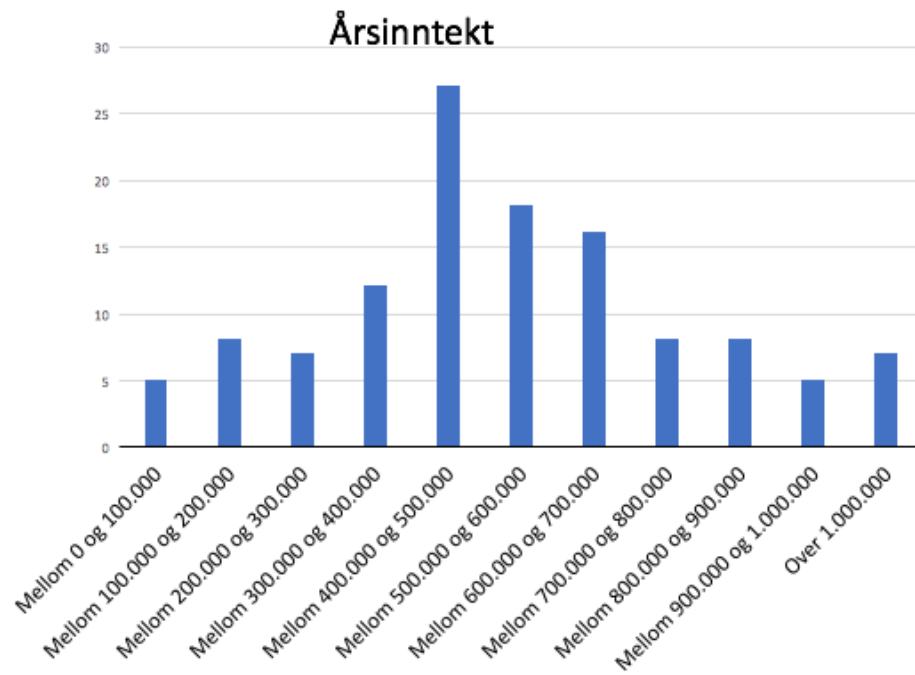
De fleste respondentene svarer at deres høyeste fullførte utdanning er videregående skole (43%) eller bachelor/mellomfag (42%). 11% av respondentene har masterutdanning/hovedfag, og 1 har doktorgrad. 53,28% av respondentene har altså høyere utdanning.



Figur 5: Utdanning

3.1.5 Inntekt

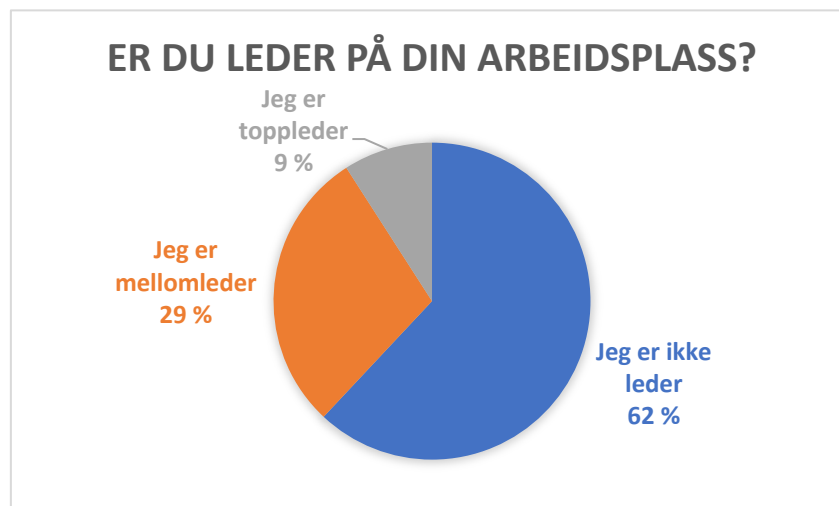
Inntekten ble målt i 11 intervaller, der den laveste var “0-100.000” og den høyeste “over 1.000.000”. Det vanligste intervallet er årsinntekt på mellom 400.000 og 500.000, da 27 av respondentene svarte dette. Dette reflekterer norsk arbeidsliv godt.



Figur 6: Årsinntekt

3.1.6 Leder

75 av våre respondenter sier at de ikke er leder på sin arbeidsplass. Av de gjenværende, er 35 mellomleder, og de resterende 11 er toppledere. Dette viser til bra variasjon i utvalget.



Figur 7: Lederfordeling

3.2 Cronbachs alfa

Cronbachs alfa måler reliabilitet mellom indikatorer på samme variabel. Dette bygges på den gjennomsnittlige korrelasjonen mellom indikatorene (Gripsrud et.al, 2016). Har indikatorene en høy Cronbachs alfa, kan de kombineres til en variabel, da vi da vet at de måler det samme fenomenet. Våre data er eksperimentelle, og vi ønsker derfor en Cronbachs alfa på over .6, men under 1.

Variabel	Cronbachs alfa
Kritikkverdig (Q8, Q10, Q11)	.908
Moralsk ansvar (Q13, Q14, Q15)	.928
Straff (Q16, Q17, Q18)	.902
Sinne (Q25, Q26)	.954

Tabell 2: Cronbachs alfa

3.2.1 Kritikkverdig

Som nevnt i 2.4 *Utfallsmål* tidligere i oppgaven, så må enkelte av holdningsmålene reverseres for at dataene kan brukes i videre analyse. Vi reverserer Q11, og regner så ut Cronbachs alfa på Q11, Q8 og Q10. Dette er for å se om indikatorene tenkt å måle samme det samme, faktisk har høy korrelasjon. Cronbachs alfa er på .908, og vi kan derfor si at måleinstrumentet på kirurgens kritikkverdighet er svært høy. Vi slår indikatorene sammen til et felles mål på denne variabelen.

3.2.2 Moralsk ansvar

Vi reverserer Q14, og regner ut Cronbachs alfa på dette spørsmålet sammen med Q13 og Q15. Cronbachs alfa er på .928, som er tilstrekkelig høyt til å slå sammen indikatorene for å lage et felles mål på moralsk ansvar.

3.2.3 *Straff*

Vi reverserer Q18. Vi finner Cronbachs alfa av Q16, Q17 og Q18. Den er på .902, som også er mer enn høyt nok til å slå indikatorene sammen til et samlet mål på om kirurgen fortjener straff.

3.2.4 *Sinne*

Vi regner ut Cronbachs alfa av Q25 og Q26 og får at denne er .954. Dette viser at måleinstrumentet på sinne er svært reliabelt, og dermed kan vi samle indikatorene til et felles mål på sinne.

3.3 Korrelasjonsanalyser

Ved bruk av Pearsons korrelasjonskoeffisient måler vi en lineær sammenheng mellom to variabler. Korrelasjonsanalysen gir oss en korrelasjonskoeffisient som kan variere mellom -1 til +1, som er et kvantitativt uttrykk for reliabiliteten av undersøkelsen (Gripsrud et al., 2016, s.211). Korrelasjonsverdier mellom 0-.3 gir ingen korrelasjon mellom variablene. Verdier mellom .3-.6 gir en middels korrelasjon og verdier mellom .6-1 gir en sterk korrelasjon mellom variablene.

Av våre syv variabler ligger de på middels til sterk korrelasjon mellom variablene, utenom ved sykehuset og produsenten som erstatningsansvarlig. Her er det ingen korrelasjon som kan måles. Vi kan blant annet se at mer straffverdig korrelerer kraftig med følelse av mer moralsk ansvar [$r=.779$]. Mer straffverdig og mer kritikkverdig korrelerer også kraftig [$r=.657$]. Så har vi også mer kritikkverdig som korrelerer kraftig med moralsk ansvar [$r=.635$].

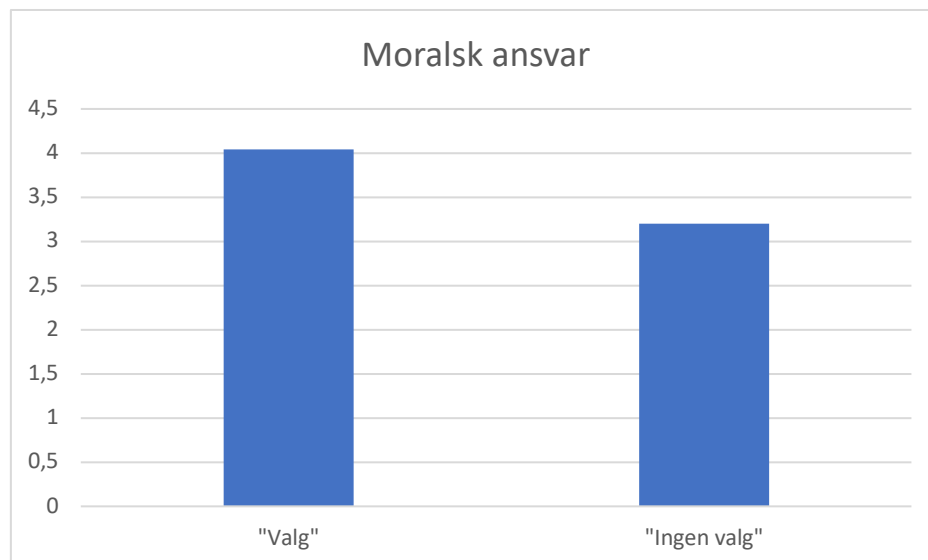
Produsent som erstatningsansvarlig og følelse av moralsk ansvar har en negativ korrelasjon [$r=-.041$] som tilsier at respondentene så ingen sammenheng mellom dette, dette ser vi også i produsent som erstatningsansvarlig med mer kritikkverdig [$r=-.027$].

Variabel	M	SD	1	2	3	4	5	6	7
1. Følelse av mer moralsk ansvar	3.65	1.73	---						
2. Mer kritikkverdig for feilen som inntraff	3.06	1.57	.635	---					
3. Mer straffverdig	2.58	1.40	.779	.657	---				
4. Vekker mer sinne	2.54	1.41	.522	.314	.508	---			
5. Kirurgen som erstatningsansvarlig	2.56	1.60	.387	.311	.482	.408	---		
6. Sykehuset som erstatningsansvarlig	5.18	1.74	.095	.106	.130	.129	.110	---	
7. Produsent som erstatningsansvarlig	4.48	1.83	-.041	-.027	.070	.266	.133	.486	---

Tabell 3: Korrelasjonsmatrise

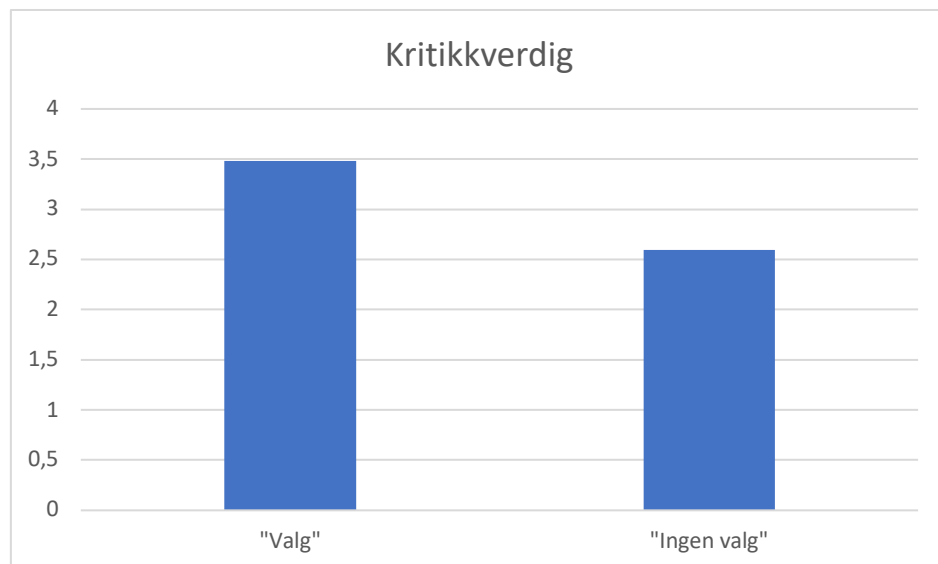
3.4 Mellomgruppeskjeller

Vår første hypotese var at kirurgen ville oppfattes som mer moralsk ansvarlig for feilen som inntraff, dersom han hadde muligheten til å utføre inngrepet selv. Dette til tross for at han visste at roboten var en sikrere løsning med tanke på risiko for feil. For å teste denne hypotesen utførte vi et uavhengig utvalg t-test for å studere forskjeller i gjennomsnitt mellom gruppene. Gruppen som leste at kirurgen kunne valgt å operere selv følte at han hadde noe moralsk ansvar for det inntrufne [M=4.04, SD=1.62]. Gruppen som leste at kirurgen ikke hadde noe valg følte at han hadde betydelig mindre moralsk ansvar for det inntrufne [M=3.20, SD=1.74]. Denne forskjellen viste seg å være statistisk signifikant [$t[124]=2.81, p=.006$]. Vi finner dermed støtte for vår første hypotese. Kirurgens tilgjengelige alternativ forårsaker et større opplevd moralsk ansvar, enda dette valgalternativet er suboptimalt.



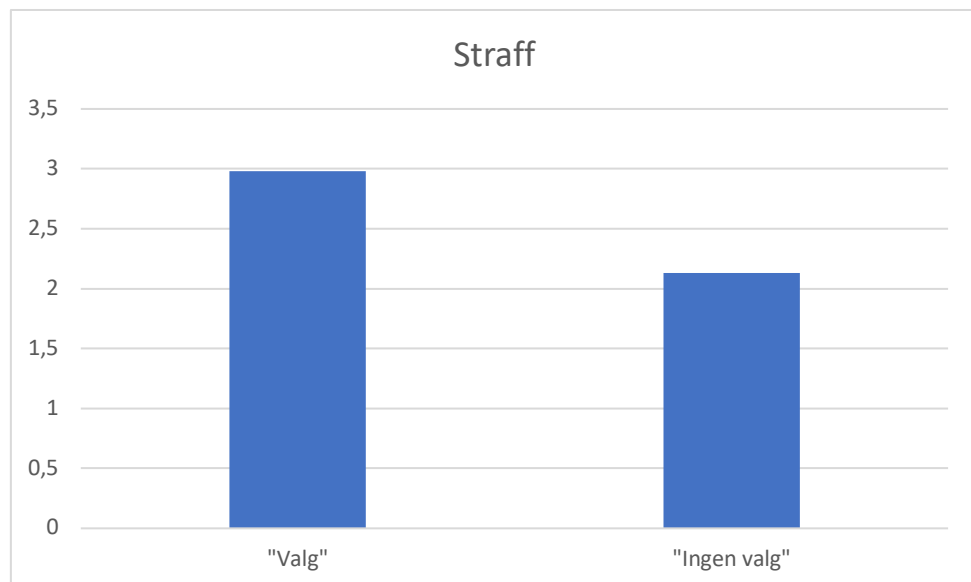
Figur 8: Gruffeforskjell moralsk ansvar

Vår andre hypotese var at kirurgen ville oppfattes som mer kritikkverdig for feilen som inntraff, dersom han hadde muligheten til å utføre inngrepet selv. Dette til tross for at han visste at roboten var en sikrere løsning med tanke på risiko for feil. For å teste denne hypotesen utførte vi et uavhengig utvalg t-test for å studere forskjeller i gjennomsnitt mellom gruppene. Gruppen som leste at kirurgen kunne valgt å operere selv følte at han var noe mer kritikkverdig for det inntrufne [M=3.48, SD=1.54]. Gruppen som leste at kirurgen ikke hadde noe valg følte at han var betydelig mindre kritikkverdig for det inntrufne [M=2.60, SD=1.47]. Denne forskjellen viste seg å være statistisk signifikant [$t[126]=3.31, p=.001$]. Vi finner dermed støtte for vår andre hypotese. Kirurgens tilgjengelige alternativ forårsaker en større opplevd grad av kritikk, enda dette valgalternativet er suboptimalt.



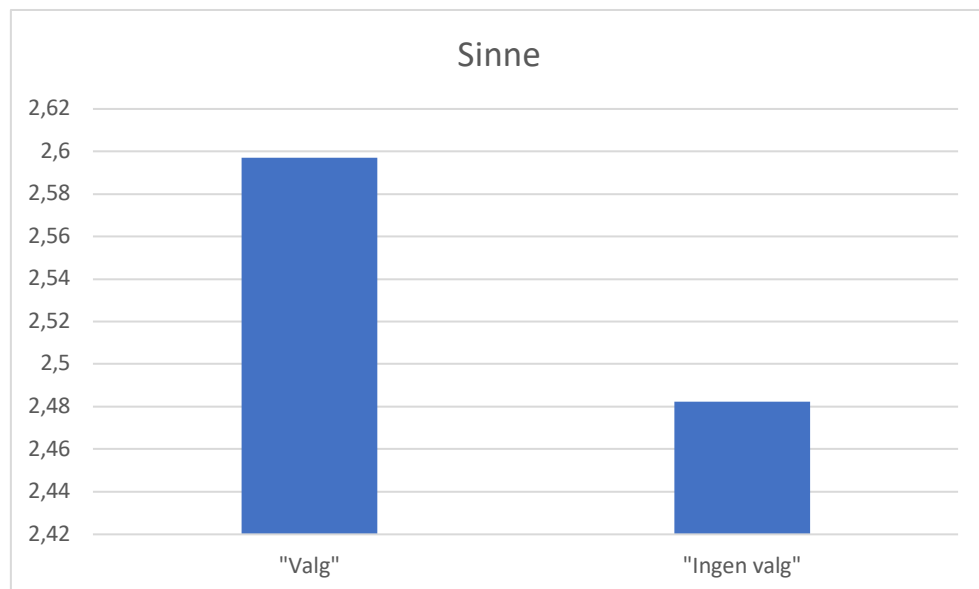
Figur 9: Gruppeforskjell kritikkverdige

Vår tredje hypotese var at kirurgen ville oppfattes som mer straffverdige for feilen som inntraff, dersom han hadde muligheten til å utføre inngrepet selv. Dette til tross for at han visste at roboten var en sikrere løsning med tanke på risiko for feil. For å teste denne hypotesen utførte vi et uavhengig utvalg t-test for å studere forskjeller i gjennomsnitt mellom gruppene. Gruppen som leste at kirurgen kunne valgt å operere selv følte at han var noe mer kritikkverdige for det inntrufne [$M=2.97$, $SD=1.44$]. Gruppen som leste at kirurgen ikke hadde noe valg følte at han var betydelig mindre kritikkverdige for det inntrufne [$M=2.12$, $SD=1.19$]. Denne forskjellen viste seg å være statistisk signifikant [$t[121]=3.53$, $p=.001$]. Vi finner dermed støtte for vår tredje hypotese. Kirurgens tilgjengelige alternativ forårsaker en større opplevd grad av straffverdighet, enda dette valgalternativet er suboptimalt.



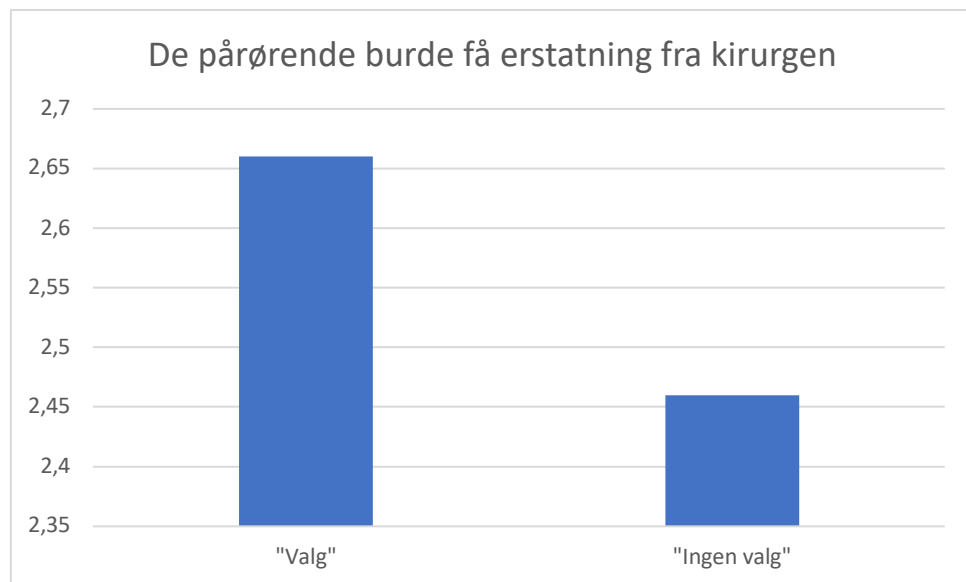
Figur 10: Gruppeforskjell straff

Vår fjerde hypotese var at robotens feiltrinn ville vekke mer sinne hos respondentene, dersom kirurgen hadde muligheten til å utføre inngrepet selv. Dette til tross for at han visste at roboten var en sikrere løsning med tanke på risiko for feil. For å teste denne hypotesen utførte vi et uavhengig utvalg t-test for å studere forskjeller i gjennomsnitt mellom gruppene. Gruppen som leste at kirurgen kunne valgt å operere selv følte at han var noe mer kritikkverdig for det inntrufne [$M=2.60$, $SD=1.32$]. Gruppen som leste at kirurgen ikke hadde noe valg følte at han var betydelig mindre kritikkverdig for det inntrufne [$M=2.482$, $SD=1.521$]. Denne forskjellen viste seg å ikke være statistisk signifikant [$t[122]=.449$, $p=.655$]. Vi finner dermed ingen støtte for vår tredje hypotese. Kirurgens tilgjengelige alternativ vekker ikke mer sinne.



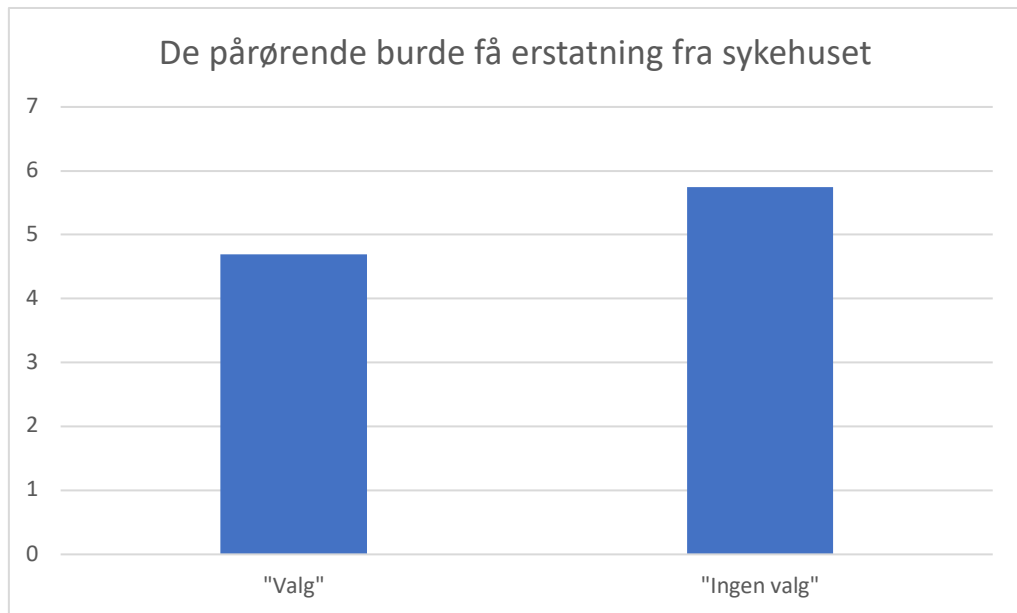
Figur 11: Gruppeforskjell sinne

Vår femte hypotese var at flere ville se kirurgen som erstatningsansvarlig til pårørende, dersom han hadde muligheten til å utføre inngrepet selv. Dette til tross for at han visste at roboten var en sikrere løsning med tanke på risiko for feil. For å teste denne hypotesen utførte vi et uavhengig utvalg t-test for å studere forskjeller i gjennomsnitt mellom gruppene. Gruppen som leste at kirurgen kunne valgt å operere selv følte at han var noe mer kritikkverdig for det inntrufne [M=2.66, SD=1.513]. Gruppen som leste at kirurgen ikke hadde noe valg følte ikke at han var betydelig mer erstatningsansvarlig for det inntrufne [M=2.46, SD=1.702]. Denne forskjellen viste seg ikke å være statistisk signifikant [$t[122]=.695, p=.489$]. Vi finner dermed ingen støtte for vår femte hypotese. Kirurgens tilgjengelige alternativ forårsaker en større opplevd grad av erstatningsansvar.



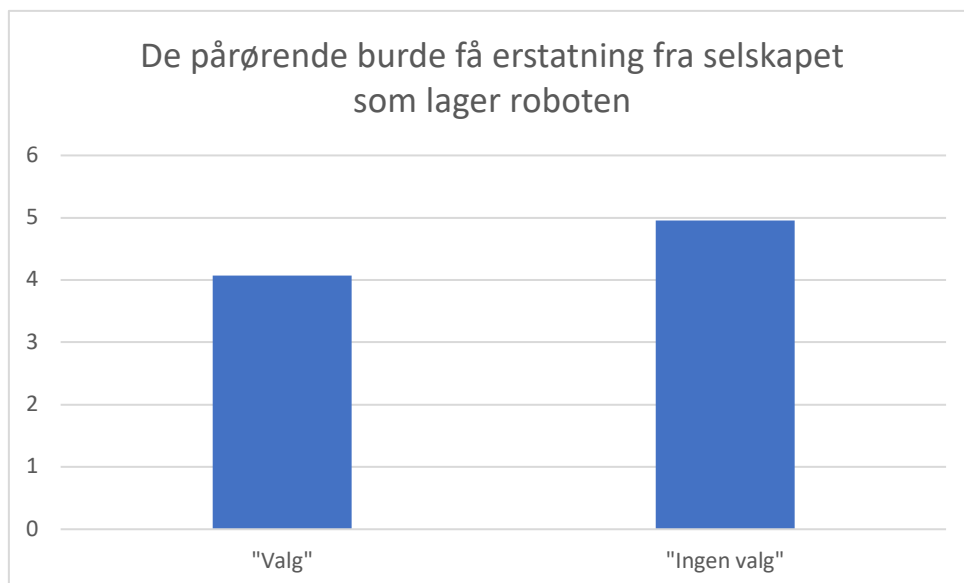
Figur 12: Gruffeforskjell erstatning fra kirurg

Vår sjette hypotese var at flere ville se sykehuset som erstatningsansvarlig for feilen som inntraff, dersom kirurgen ikke hadde muligheten til å utføre inngrepet selv. Dette på grunn av at han visste at roboten var en sikrere løsning med tanke på risiko for feil. For å teste denne hypotesen utførte vi et uavhengig utvalg t-test for å studere forskjeller i gjennomsnitt mellom gruppene. Gruppen som leste at kirurgen kunne valgt å operere selv følte i mindre grad at sykehuset var erstatningsansvarlige for det inntrufne [M=4.69, SD=1.877]. Gruppen som leste at kirurgen ikke hadde noe valg følte at han var betydelig mindre kritikkverdig for det inntrufne [M=5.75, SD=1.379]. Denne forskjellen viste seg å være statistisk signifikant [$t[122]=-3.555, p=.001$]. Vi finner dermed støtte for vår sjette hypotese. Kirurgens mangel på alternativ forårsaker at flere ser sykehuset som erstatningsansvarlige.



Figur 13: Gruffeforskjell erstatning fra sykehus

Vår syvende hypotese var at flere ville se robotens produsent som erstatningsansvarlig for feilen som inntraff, dersom kirurgen ikke hadde muligheten til å utføre inngrepet selv. Dette på grunn av at han visste at roboten var en sikrere løsning med tanke på risiko for feil. For å teste denne hypotesen utførte vi et uavhengig utvalg t-test for å studere om robotens produsent er mer erstatningsansvarlige for det inntrufne [$M=4.96$, $SD=1.569$]. Denne forskjellen viste seg å være statistisk signifikant [$t[122]=-2.761$, $p=.007$]. Vi finner dermed støtte for vår syvende hypotese. Kirurgens mangel på alternativ forårsaker at flere ser robotens produsent som erstatningsansvarlige.



Figur 14: Gruffeforskjell erstatning fra selskap

4.0 Diskusjon

4.1 Hovedfunn

Hensikten med denne studien var å avdekke om valgalternativ om å benytte KI vil overføre mer moralsk ansvar til den som valgte, enn dersom han ikke hadde noe valg. Analysene våre har grunnlag i resultatene i mellomgruppeskjellene.

I resultatene vises betydelige forskjeller i de avhengige variablene, i de to forskjellige eksperimentelle gruppene. Eksperimentet avdekket at kirurgen ble sett på som mer moralsk ansvarlig for robotens feil, dersom han fikk muligheten til å velge om den skulle brukes, i tråd med vårt forskningsspørsmål.

4.2 Teoretiske implikasjoner

Hypotese 1 målte henholdsvis om kirurgen ble sett på som mer moralsk ansvarlig dersom han kunne velge om roboten skulle benyttes. Våre respondenter var informert om at roboten var et betydelig sikrere alternativ for pasienten, og dermed at kirurgens valg var riktig. Kirurgen kunne ikke velge å operere selv, uten å bryte med Legeforeningens etiske regler (Den Norske Legeforening, 2015). Til tross for dette, mener gruppen der han fikk velge, at han er noe moralsk ansvarlig for feilen som skjer. Dette tyder på at denne gruppen ser kirurgens valg som fritt, ettersom tvungne valg sjeldent fører med moralsk ansvar (Greenfield, 2011).

Vi mener at valget i realiteten ikke er fritt, men tvunget av den betydelige forskjellen i risiko, som hadde gjort det uetisk å velge å ta operasjonen selv.

Algoritme-aversjon viser at vår evne til å ta beslutninger basert på statistiske mål, lett kan svekkes. I dette tilfellet kan det tenkes at forskjellen i risiko ikke tas i full betraktning, fordi troen ekspertise-intuisjon overkjører rasjonell tenkning (Highhouse, 2008). At scenarioet som fremstilles er såpass alvorlig, kan også tenkes å ha innvirket på at valget oppfattes fritt. Jo mer emosjoner som vekkes, desto vanligere er det å se valg som frie (Nichols & Knobe, 2007).

Hypotese 2 og 3 målte om kirurgen ble sett på som mer kritikkverdig og mer straffverdig dersom han kunne velge å operere pasienten selv. Våre respondenter som fikk vite at kirurgen hadde et valg føler at han er mer kritikkverdig og straffverdig ovenfor det som skjedde, i motsetning til respondentene som fikk vite at han ikke hadde noe valg. De som fikk vite at han hadde et valg, bebreider kirurgen for feilen og mener at det er han som er skyld i det som har skjedd. Dette tenkes å henge nært sammen med oppfattelsen av valget hans som fritt, slik som i hypotese 1.

Forskere har kommet frem til et merittbasert og et konsekvensetisk syn på ros eller skyld plassering, dette kan bare brukes dersom det er en passende reaksjon og hvis dette kunne ha ført til en forandring (Eshleman, 2014). Dette virker å stride mot resultatene i hypotese 1, 2 og 3 da forandring av valget i dette tilfellet, ville ført til et betydelig suboptimalt valg.

Ved at respondentene bebreider kirurgen og føler at han fortjener straff for det som har skjedd kan tenkes å gå utover trivselen og psyken til kirurgen da det går på han personlig. Vanlige følger av pasientdød er skyldfølelse og følelse av inkompetanse hos innblandet helsepersonell (*Scott, Hirschinger et al, 2009*). Dette kan også øke følelsen av jobbusikkerhet, dersom kritikken og ønsket om straff blir for stort. Vi vet at mennesker ser andre som ansvarlige for de jobbene de velger å ta (Greenfield, 2011). Dette kan være med på å påvirke mennesker til å unngå å ta jobber for å ikke bli holdt personlig ansvarlig for de konsekvensene som måtte komme ut av dette valget.

Det er viktig at arbeidsplasser forstår konsekvensene dette kan føre med seg for ansatte. Studien vår viser at enkeltpersonens tilgjengelige valgmuligheter i bruken av KI, vil gi personen ansvar og skyld dersom noe går galt. Uten dette valgalternativet blir man ikke sett på som ansvarlig. Ledelsen i organisasjoner som benytter seg av KI, må ta hensyn til denne ansvarsrisikoen, og ta vare på sine ansatte. Risikoen kan minimeres ved å fjerne muligheten til å velge, som ved at organisasjonen innfører faste prosedyrer for bruken av KI. Ved å også gjøre produsentene til KI klar over risikoen, kan man jobbe sammen for å utvikle bedre løsninger. De enkeltansatte bør i så stor grad som mulig, unngå å implementeres i

risikofylte situasjoner utenfor deres kontroll, slik som kirurgen ble. Slik kan unødig påkjenning unngås.

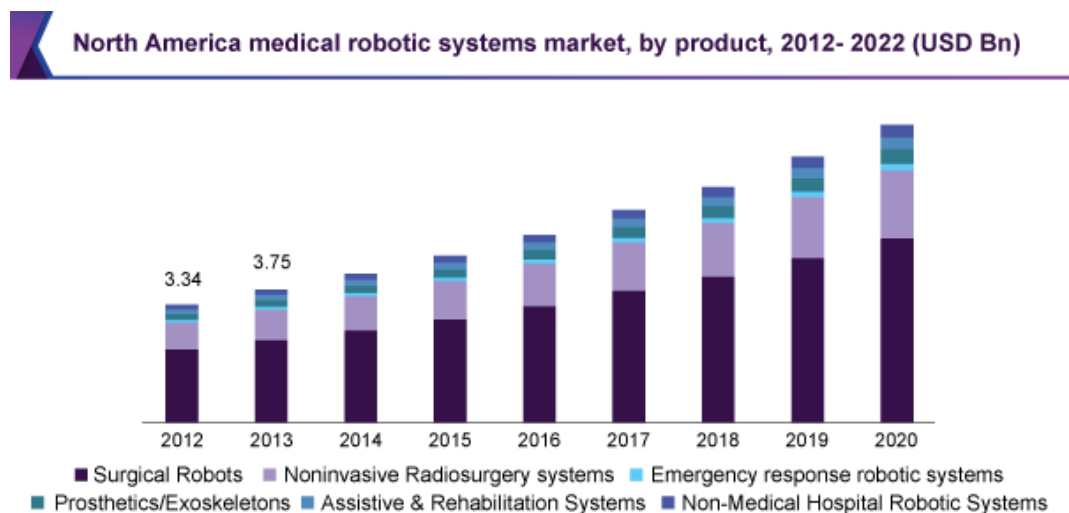
Hypotese 4 var at gruppen som leste om at kirurgen hadde et valg, ville oppleve mer sinne, enn gruppen der han ikke hadde noe valg. Vi fant ikke støtte for denne hypotesen, til tross for at gruppen med valg ser kirurgen som mer moralsk ansvarlig, kritikkverdig og straffeverdig. Dette kan tenkes å skyldes at scenarioet var fiktivt, og at den emosjonelle responsen kanskje ville vært annerledes i en mer virkelighetsnær situasjon. Det kan også tenkes at fordi respondentene ble informert om forskjellene i risiko mellom mennesket og roboten, visste de at valget hans var riktig. En implikasjon av dette er at jo bedre informert de involverte er om prosessen ved bruk av KI, jo mindre feilplassering av ansvar, kritikk og straff vil skje.

Hypotese 5 var at flere ville se kirurgen som erstatningsansvarlig dersom han hadde mulighet til å utføre inngrepet selv. Vi fant heller ikke støtte for denne hypotesen, til tross for at kirurgen ble sett på som moralsk ansvarlig i tilfellet der han kunne velge. Vi mener dette kan skyldes norske samfunnsnormer, det er ikke vanlig at enkeltansatte i helsevesenet skal betale erstatning til pårørende.

Hypotese 6 var at flere ville se sykehuset som erstatningsansvarlig, når kirurgen ikke kunne velge. Hypotese 7 var at flere ville se robotens produsent som erstatningsansvarlig, når kirurgen ikke kunne velge. Vi fant støtte for begge disse hypotesene. Dette tyder på at når kirurgen ikke kan velge, overføres noe av skylden fra kirurgen til sykehuset og robotens produsent. I Norge er det vanligst å søke erstatning fra sykehuset dersom feiltrinn begås, men ettersom roboten opererer selvstendig, er det ikke overraskende at mange mener produsenten er skyldig. Dersom dødsfallet skjedde som følge av en produktfeil hos roboten, vil naturlig nok erstatningsansvaret tilegnes produsenten, da de er ansvarlige for programmering og testing.

4.3 Implikasjoner for arbeidsliv

KI vil implementeres i et betydelig større antall arbeidsplasser fremover (Russel & Norvig, 2016). Innen helseorganisasjoner kan KI forbedre prosesser betydelig. KI brukes allerede for å gjøre behandlinger av pasienter smartere og mer målrettet mot ulike livstruende sykdommer, og i takt med Europas aldrende befolkning kan KI være et verdifullt verktøy for å være med på å redde enda flere liv i fremtiden (European Commission, 2019). Virtuell KI kan brukes i diagnostisering, behandling og oppfølging, ved å foreslå løsninger ut fra relevant informasjon den har tilgang til. Innen kirurgi tenkes kunstig intelligente roboter å gjøre fremskritt. En robot kan utføre samme oppgave med samme presisjon i timesvis uten å bli sliten. Dette er en betydelig fordel i kirurgi der menneskelige risikofaktorer som tretthet, ustødighet og lignende kan spille inn (Anandan, 2018).



Figur 15: Markedsprediksjon over medisinske roboter i Nord-Amerika

(Grand View Research, 2018)

Den virtuelle grenen av KI er mer utbredt når det kommer til utvikling, testing og implementering, enn den fysiske grenen (Christiansen, 2017; Jiang F, Jiang Y, Zhi H, et al, 2017; Hannun, Rajpurkar et. al, 2019; Grady, 2019; Yala et. al, 2019). Dette ser ut til å kunne endre seg i fremtiden, da interessen for investering i KI-roboter nå er svært høy (Allidina, 2016). Til tross for dette har vi funnet lite interesse for fullstendig selvstendige roboter innen kirurgi. Dagens fokus ligger på *robotassisterte* løsninger, der menneskelig helsepersonell tar alle avgjørelser om bevegelse og inngrep (Grand View Research, 2018; Sajeev & Chandra, 2018). Det

er usikkert om fullstendig selvstendige robotkirurger vil utvikles i det hele tatt, da dette vil avhenge av interessen for å ta dem i bruk.

4.4 utfordringer

Vi står ovenfor flere utfordringer når det kommer til bruk av KI i arbeidslivet. Den første tenkes å være å overbevise personer om å bruke KI, spesielt i tilfeller der den utfører oppgaver betydelig bedre enn mennesker kan. Vi vet allerede at mennesker er kritiske til bruken av KI dersom de ikke føler noe kontroll over den, og spesielt i tilfeller der de har vært vitne til at den ikke er perfekt (Dietvorst, Simmons, Massey, 2016).

Ved å gi brukere mulighet til å modifisere KI selv, øker villigheten til å benytte teknologien betraktelig (Dietvorst, Simmons, et al, 2015). Et relevant spørsmål blir derfor om vi kommer bedre ut av å la folk svekke suksessraten, så lenge dette betyr at KI tas i bruk? Er fordelene ved KI så store at det veier opp for den reduserte suksessraten? Her vil typen KI og bransjen måtte ha en innvirkning. En KI utviklet for å matche mennesker på en datingside, vil ikke sette liv i fare dersom den modifiseres litt og suksessraten svekkes. I helsesektoren vil potensielt skadelig menneskelig innblanding ha betydelig mer alvorlige konsekvenser. Det blir viktig for ledere av slike organisasjoner å basere seg på sikkerheten til ansatte og pasienter, noe som mest sannsynlig vil redusere muligheter for modifisering.

Det kan tenkes at suksessraten må ofres noe, for å gi helhetlig forbedring. Samfunnet i sin helhet tenkes å forbedres ved økt implementering av KI i tilfeller der KI viser seg å være bedre enn mennesker. Likevel viser våre resultater at ønsket om modifikasjon kan føre med seg uforutsette konsekvenser, da dersom vår kirurg hadde vært mer innblandet i robotens feiltrinn, tenkes ennå mer av det moralske ansvaret å ha blitt overført til han.

Produsentene til KI vil i stor grad kunne bestemme hvor mye modifikasjonsfrihet de vil gi til brukerne av sine produkter. Dersom modifikasjoner på deres produkter fører til flere feil, kan dette tenkes å reflektere negativt på produsenten. Det kan derfor tenkes at de ikke vil være villige til å tilby denne muligheten. Til tross for dette, viser vårt eksperiment at ansvaret ved feiltrinn kan bli overført til brukeren av KI-en, og dermed vil ikke produsenten ha noe interesse av å fjerne muligheten

for modifikasjon. Vi tenker at det mest sannsynlige er at produsentene vil fokusere på å sikre suksessraten til produktene sine, da dette har høy påvirkning på at de vil klare å selge produktene på markedet.

Innen helseorganisasjoner ser vi følgende utfordringer som svært relevante:

KI som tas i bruk må være tilstrekkelig sikkert og stabilt (European Commission 2019). God infrastruktur i form av stabile servere og opplæring av ansatte er essensielt. Tilstrekkelige alternative prosedyrer må også være på plass i tilfeller der systemene ikke fungerer som forventet.

Et potensielt problem med at KI blir dypt integrert på arbeidsplasser, er bruken av tvunget valg, altså at man ikke lenger kan velge vekk en KI når den er blitt implementert i den daglige bruken. Når man fremdeles har KI som et valg så kan ikke en handling bli tvunget på noen, da valget om å utføre handlingen må komme fra en person først og fremst, og all informasjon ved valget må være tilgjengelig. Man mister altså muligheten til å ikke bruke KI, noe som kan føre til problemer i situasjoner der denne ikke er perfekt.

Andre utfordringer vil bli at maskinlæringsystemer må fortsette å oppdateres etter at de tas i bruk, for å unngå skjevhet i data. Tendenser i befolkningen endres over tid, og dersom systemene ikke inkluderer ny data, vil de fort kunne bli utdatert. En siste utfordring er mangel på innsyn i beslutningstakingsprosessene i KI. Det er fortsatt svært vanskelig å forstå hvordan KI “tenker”, og kommer frem til beslutningene den tar (Olden & Jackson, 2002). I tilfeller der KI gjør feil, er dette et stort problem, da det kan være umulig å forstå hvordan feilen oppsto, og dermed hvordan den skal unngås i fremtiden.

4.4.1 Implikasjoner for ledere

Kunstig intelligens kan være et utrolig hjelpsomt verktøy, dersom det brukes på en produktiv måte. Dagens ledere bør være klar over dette, da det blir lett å falle fra dersom man ikke tar hensyn til utviklingen av ny teknologi. KI kan brukes til mye forskjellig, så potensialet for effektivisering er stort. Innen helseorganisasjoner går mye tid med på rutineoppgaver slik som å sette seg inn i en pasients sykdomshistorikk. Dersom KI kan ta over slike oppgaver, kan helsepersonell få mer tid til interaksjon med pasientene. KI er også god til å

forutse ting før de skjer, og kan være et svært nyttig hjelpemiddel i lesing av CT, EKG, mammografier og andre bilder, for å oppdage sykdommer på ennå tidligere stadier.

For å implementere KI på en arbeidsplass på en produktiv måte, må alle ansatte bli godt informert om den nye teknologien og dens begrensninger. KI kommer kanskje aldri til å bli perfekt, så ledere bør rutiner for å sikre sikkerheten til pasienter og ansatte. For å unngå problemer med ansvars plassering slik som vi belyser i vår studie, bør klare prosedyrer implementeres for bruk av KI. I nærmeste fremtid bør KI kun sees som et hjelpemiddel, der kyndige personer tolker beslutningene teknologien foreslår.

4.5 Begrensninger og videre forskning

Det er lite tilgjengelig forskning på vårt tema, KI i kirurgi, og bruk av KI i helseorganisasjoner. Kunstig intelligens i dag er ikke kommet langt nok til å ha blitt implementert på måten vårt eksperiment fremstiller, og eksisterer i dag kun som et verktøy for menneskelig assistanse. Dette gjør det vanskelig å sammenligne vårt eksperiment med tidligere forskning, og vi har derfor tatt egne forutsetninger rundt temaene der lite forskning finnes.

Vårt utvalg respondenter er begrenset av hvem vi hadde tilgjengelig. Et slikt bekvemmelighetsutvalg kan være mindre overførbart for den generelle befolkningen, da en rekke potensielle respondenter blir utelukket (Gripsrud, et al, 2016). Kun de som hadde tilgang til internett via PC eller mobil kunne ta undersøkelsen, og alle respondenter måtte være knyttet til oss for at vi kunne nå ut til dem. Vi hadde mange missing values i våre data, noe som kan ha ført til en skjev fordeling.

Typiske utfordringer for spørreundersøkelser er tilfellet for oss også.

Utenforstående variabler kan ha spilt inn på våre respondenters svar. For å unngå at respondentene bare trykket på svar uten å tenke, ble enkelte spørsmål snudd, og kontrollspørsmål inkludert. Likevel er det en mulighet for at respondentene ikke svarte ærlig på spørsmålene. Det kan også tenkes at menn får en sterkere reaksjon til vårt presenterte scenario, da prostataoperasjoner er mer aktuelt for dem. Vi vet

at sterke emosjoner påvirker evnen til å tenke rasjonelt (Sunstein, 2002), noe som det usikkert om våre resultater er overførbare til en reell situasjon.

Ved videre forskning kan det være relevant å utvide populasjonen til andre kulturer for å få et større perspektiv og se om man kan finne en sammenheng. Et eksempel kan være i USA hvor det er vanlig i kulturen å kunne saksøke hvis en feil skjer, her vil man tro at det kan eksistere mer sinne mot kirurgen og meninger om større erstatning. Å replikere studien i Norge med en større populasjon, vil også kunne gjøre resultatene mer reelt overførbare. Vi synes også det hadde vært svært interessant å se kontrasten mellom ansvars plassering i vårt fiktive scenario, og et scenario der kirurgen velger å ta operasjonen selv, men begår tilsvarende feil. Slik kunne man lettere å ha sammenlignet graden av moralsk ansvar som overføres til kirurgen.

5.0 Konklusjon

Problemstillingen vår gikk ut på å besvare hvordan mennesker oppfatter moralsk ansvar i en situasjon der en kunstig intelligent robot gjør feil i kirurgi. Det ble utviklet syv hypoteser for å hjelpe oss med å besvare dette. Funnene fra vårt eksperiment indikerer at fem av de syv hypotesene støttes, og betraktes som korrekte. Sett i sammenheng med teori, viser hypotesene det moralske ansvaret for situasjoner der KI gjør feil, vil overføres til den nærmeste menneskelige beslutningstakeren.

Dette viser at kunstig intelligens vil føre med seg viktige etiske problemstillinger i fremtiden. Disse etiske spørsmålene om ansvars plassering bør tas hensyn til før KI implementeres i stor grad i samfunnet. Det jobbes allerede med å utvikle fastsatte retningslinjer for bruk av KI, og dette tenkes å komme ennå mer i fokus i nær fremtid.

Vi besvarer forskningsspørsmålet vårt, med at kirurgens mulige valg, gjorde at han fikk tildelt mer ansvar for feilen. Basert på dette kan vi konkludere med at dersom det oppfattes at man har et valg, vil dette føre med seg moralsk ansvar for hendelser som følger.

6.0 Bibliografi

- Akershus Universitetssykehus. (2018, September 04). *ahus.no*. Hentet fra Oppstart av robotkirurgi: <https://www.ahus.no/nyheter/oppstart-av-robotkirurgi>
- Allidina, S. (2016, August 22). *Raconteur*. Hentet fra The rise of artificial intelligence in 6 charts: <https://www.raconteur.net/business-innovation/the-rise-of-artificial-intelligence-in-6-charts>
- Anandan, T. M. (2018). Robots and AI in the OR. *Robotics Industry Insights*.
- ARTAS - Robotic Hair Restoration. (2019). *Artas.com*. Hentet fra <https://artas.com/about-the-artas-procedure/>
- Attia, Z. I., Kapa, S., Lopez-Jimenez, F., McKie, P. M., Ladewig, D. J., Satam, G., . . . Carter, R. (2019, Januar 07). Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nature Medicine*(25, 70-74 (2019)).
- Benestad, H. B. (2019, Mars 22). Tolkning av EKG med kunstig intelligens. *Tidsskriftet*.
- Cappelen, A., Fest, S., Sørensen, E., & Tungodden, B. (2016, Desember 21). Choice and personal responsibility: What is a morally relevant choice? *Statsøkonomen*.
- Christiansen, A. (2017, Mars 06). *forskning.no*. Hentet fra Kunstig intelligens kan hjelpe leger: <https://forskning.no/helse-partner-universitetet-i-agder/kunstig-intelligens-kan-hjelpe-leger/361054>
- Connor-Simmons, A., & Gordon, R. (2019, Mai 7). *MIT News*. Hentet fra Using AI to predict breast cancer and personalize care: http://news.mit.edu/2019/using-ai-predict-breast-cancer-and-personalize-care-0507?fbclid=IwAR2mQkDCPe9LeW0f_SXY_n9Hvf__w88hkS2OooUJSi sGXAzhO5lfGqRm69c
- Cozby, P. C., & Bates, S. (2012). *Methods in Behavioral Research*. McGraw Hill Higher Education.
- Den norske legeforening. (2015). *Den norske legeforening*. Hentet fra Etske regler for leger: <https://beta.legeforening.no/om-oss/Styrende-dokumenter/legeforeningens-lover-og-andre-organisatoriske-regler/etske-regler-for-leger/>

- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016, August 15). Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science*.
- Dietvorst, B., Simmons, J. P., & Cade, M. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*. Hentet fra <https://psycnet.apa.org/record/2014-48748-001>
- Dolonen, K. A. (2018, Oktober 23). *Sykepleien.no*. Hentet fra 110 000 pasientskader oppstod under sykehusopphold i fjor: <https://sykepleien.no/2018/10/110-000-pasientskader-oppsto-under-sykehusopphold-i-fjor>
- Eshleman, A. (2014). Moral Responsibility. *Stanford Encyclopedia of Philosophy Archive*.
- European Commission. (2019). *Ethics guidelines for trustworthy AI*. European Commission.
- Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*(521:452-459).
- Grady, D. (2019, Mai 20). *The New York Times*. Hentet fra A.I. Took a Test to Detect Lung Cancer. It Got an A.: <https://www.nytimes.com/2019/05/20/health/cancer-artificial-intelligence-ct-scans.html>
- Grand View Research. (2018, Mars). *Grand View Research*. Hentet fra Medical Robotic Systems Market Size, Share & Trends Analysis Report By Product (Surgical, Orthopedic, Laparoscopy, Neurological, Rehabilitation, Assistive, Prosthetics, Orthotics, Steerable) And Segment Forecasts, 2012 - 2022: <https://www.grandviewresearch.com/industry-analysis/medical-robotic-systems-market>
- Greenfield, K. (2011). *The Myth of Choice: Personal Responsibility in a World of Limits*. Yale University Press.
- Gripsrud, G. O., & Silkoset, R. (2017). *Metode og dataanalyse* (Vol. 3). Cappelen Damm Akademisk.
- Hamet, P., & Tremblay, J. (2017). *Artificial intelligence in medicine*.
- Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019, Januar 07). Cardiologist - lever arrhythmia

- detection and classification in ambulatory electrocardiograms using a deep neural network. (25, 65-69 (2019)).
- Hanson, W. C., & Marshall, B. E. (2001). Artificial Intelligence applications in the intensive care unit. *Crit Care Med*.
- Helo, S., & Moulton, C.-A. E. (2017). Complications: Acknowledging, managing, and coping with human error. *Translational Andology and Urology*.
- Helsedirektoratet. (2018). *Pasientskader i Norge 2017*. Helsedirektoratet.
- Highhouse, S. (2008). Stubborn Reliance on Intuition and Subjectivity in Employee Selection. *Industrial and Organizational Psychology*.
- Horvitz, E. J., Breese, J. S., & Henrion, M. (1988, Juli). Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning*, ss. 247-302.
- INTUITIVE. (2019). *INTUITIV*. Hentet fra Da Vinci XI:
<https://www.intuitive.com/en-us/products-and-services/da-vinci/systems>
- Jain, A. K., Mao, J., & Mohiuddin, K. (1996, Mars). *Artificial Neural Networks: A Tutorial*. Hentet fra
http://metalab.uniten.edu.my/~abdrahim/mitm613/Jain1996_ANN%20-%20A%20Tutorial.pdf
- Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., . . . Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*.
- Makary, M. A., & Daniel, M. (2016, Mai 03). *thebmj*. Hentet fra Medical error - the third leading cause of death in the US:
<https://www.bmj.com/content/353/bmj.i2139>
- Nichols, S., & Knobe, J. (2007, Oktober 25). *Wiley Online Library*. Hentet fra Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions: <https://onlinelibrary-wiley-com.ezproxy.library.bi.no/doi/full/10.1111/j.1468-0068.2007.00666.x>
- Olden, J. D., & Jackson, D. A. (2002, April 18). *ScienceDirect*. Hentet fra Illuminating the "black bok": a randomization approach for understanding variable contributions in artificial neural networks:
<https://www.sciencedirect.com/science/article/pii/S0304380002000649>
- Papadatou-Pastou, M. (2011). Are connectionist models neurally plausible? A critical appraisal. *Encephalos*, ss. 48(1):5-12.

- Ramesh, A. N., Kambhampati, C., Monson, J. T., & Drew, P. J. (2004). *Artificial Intelligence in medicine*. the Royal College of Surgeons of England.
- Roy, A. (2000, Januar). Artificial Neural Networks - A Science in Trouble. (2), ss. 33-38.
- Russel, S. J., & Norvig, P. (2016). *Artificial Intelligence: a modern approach* (Vol. 3). Pearson.
- Sacks, G., & Ettner, S. L. (2016). *eScholarship*. Hentet fra Surgeon Perception of Risk and Benefit in the Decision to Operate Under Uncertainty: <https://escholarship.org/uc/item/5r91k4sc>
- Sajeev, S., & Chandra, G. (2018, April). *Allied Market Research*. Hentet fra Surgical Robotics Market by Component (systems, accessories and services), by surgery type (gynecology surgery, urology surgery, neurosurgery, orthopedis surgery, general surgery and other surgeries): Global opportunity analysis and industry forecast: <https://www.alliedmarketresearch.com/surgical-robotics-market>
- Savani, K., & Rattan, A. (2012, Juni 14). A choice mind-set increases the acceptance and maintenance of wealth inequality. *Psychological science*.
- Scott, S. D., Hirschinger, L. E., Cox, K. R., McCoig, M., Brandt, J., & Hall, L. W. (2009, Oktober 18). The natural history of recovery for the healthcare provider "second victim" after adverse patient events. *US National Library of Medicine National Institutes of Health*, ss. 18(5):325-30.
- Shapiro, S. C. (1992). *Encyclopedia of Artificial Intelligence* . (2).
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Driessche, G. v., . . . Sutsk. (2016, Januar 28). Mastering the game of Go with deep neural networks and tree search. (doi:10.1038/nature16961).
- Strawson, G. (1993, September 15). The impossibility of moral responsibility. *Philosophical Studies: An international journal for philosophy in the analytic tradition.*, ss. 5-24.
- Sucarrat, G. (2017). *Metode og økonometri: En moderne innføring*. Fagbokforlaget.
- Sunstein, C. R. (2002). *Probability neglect: Emotions, worst cases, and law*. Yale Law Journal.
- The Economist. (2016). *The return of the machinery question*. The Economist.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (Vol. 3). elsevier Science.

Yala, A., Lehman, C., Schuster, T., Portnoi, T., & Barzilay, R. (2019). *Radiology*.
Hentet fra A Deep Learning Mammography-based Model for Improved
Breast Cancer Risk Prediction:
<https://pubs.rsna.org/doi/10.1148/radiol.2019182716>

7.0 Vedlegg

Presenter scenario over de to forskjellige gruppene:



Kirurgene ved Sykehuset Østlandet har tatt i bruk den nyeste og mest avanserte kirurgiroboten på markedet, da Vinci Xi. Roboten er utviklet for å utføre operasjoner på pasienter med prostatakreft. Den benytter seg av en rekke sensorer, fintfølende presisjonsinstrumenter, og kunstig intelligens, for å utføre nøkkelhullsoperasjoner.

Forskning har vist at da Vinci Xi er langt mer nøyaktig enn menneskelige kirurger, og gjør langt færre feil. Det er altså langt mer risikabelt å bli operert av en menneskelig kirurg enn av denne roboten. Undersøkelser har vist at roboten gjør feil i bare 1 av 30 tilfeller, mens menneskelige kirurger gjør feil i 1 av 10 tilfeller. Kirurgene på Sykehuset Østlandet er kjent med disse undersøkelsene.

En dag skal en pasient opereres for prostatakreft. Kirurg Sigurd Rosfjord tar imot pasienten, og ser til at da Vinci Xi utfører operasjonen.

Roboten utfører dessverre en fatal feil i inngrepet. Den misoppfatter en stor frisk blodåre til å være en del av svulsten. Roboten forsøker derfor å fjerne deler av blodåren. Pasienten lider store indre blødninger, og dør på operasjonsbordet.

Kirurg Sigurd Rosfjord var ansvarlig behandler for pasienten som døde. Sykehuset sine prosedyrer stadfester at kirurgen bestemmer hvorvidt operasjonen skal utføres av kirurgen eller av roboten. Det var Rosfjord som valgte å la pasienten bli operert av da Vinci Xi. Rosfjord hadde anledning til å operere pasienten selv, men foretrakk å la roboten utføre inngrepet.