



Handelshøyskolen BI - campus Bergen

BTH 14111

Bacheloroppgave - Human Resource
Management

Bacheloroppgave

Kunstig intelligens, men ikke kunstig moral? Menneskets
oppfatning av kunstig intelligens og dens moral.

Navn: Eivind Hundven Kjenes, Anders
Naustdal

Utlevering: 08.01.2018 09.00

Innlevering: 04.06.2018 12.00

Bacheloroppgave ved Handelshøskolen BI

Kunstig intelligens, men ikke kunstig moral? Menneskets oppfatning av kunstig intelligens og dens moral.

Eksamenskode og navn:

BTH 14111 – Human resource management

Innleveringsdato:

01.06.2018

Studiested:

BI Bergen

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	III
SAMMENDRAG	IV
TEORETISK RAMMEVERK OG HYPOTESEUTVIKLING	4
STORDATA (BIG DATA).....	4
RETTFERDIGHETSTEORI.....	4
RETTFERDIGHET OG PRIS	5
PRIS OG MORAL	6
DEN MENNESKELIGE AGENTEN.....	9
FORSKNINGSDESIGN OG METODE	11
VIGNETTMETODEN	11
BRUK AV DATA	12
VALIDITET OG RELIABILITET	12
REKRUTTERING OG UTVALG	13
EKSPERIMENTETS DESIGN – VIGNETTUNDERSØKELSEN.....	13
UTFALLSMÅL	15
<i>Opplevd rettferdighet:</i>	15
<i>Opplevd moralitet:</i>	15
<i>Ris og Ros:</i>	15
<i>Opplevd sensitivitet:</i>	16
RESULTATER	17
CHRONBACH’S ALPHA – RELIABILITETEN I SVARENE	17
DEMOGRAFISKE DATA	17
SAMMENHENG MELLOM VARIABLENE	19
<i>Korrelasjonstabell</i>	19
TEST AV MELLOMGRUPPEFORSKJELLER.....	20
DISKUSJON	22
OPPSUMMERING AV UNDERSØKELSEN	22
ANBEFALINGER TIL VIDERE FORSKNING	26
KONKLUSJON	27
REFERANSER	28

Forord

Denne oppgaven er skrevet som avsluttende del av utdanningen ved Handelshøyskolen BI Bergen. Ideen til oppgaven hentet fra vår veileder Mads Nordmo Arnestad, og er basert på en studie som er laget for å belyse menneskets oppfatning av moral innenfor forskningsfeltet kunstig intelligens. Det har vært en lang og krevende prosess, og med særdeles lite forskning på feltet tidligere har vi blitt utfordret til å bruke all vår kompetanse og kunnskap som er blitt opparbeidet gjennom vår tid på studiet. Det å kunne jobbe med et slikt dagsaktuelt tema har vært spennende, og vi har blitt mer og mer bevisst på viktigheten av å belyse denne “teknologiske bølgen”, som treffer oss med enorm hastighet. Vi ønsker å takke vår veileder Mads Nordmo Arnestad, som har gitt oss tilgang og mulighet til å jobbe med dette interessante studiet, og for god hjelp og oppfølging undervegs. Vi ønsker også å takke alle respondentene som var med i undersøkelsen. Vi har hatt stor hjelp fra familie og venner som har bidratt til rekruttering av deltakere. Å skrive denne oppgaven har vært ekstremt lærerikt. Vi føler oss privilegert som har fått muligheten til å kunne jobbe med et så stort og ikke minst nytt felt. Den stadige utviklingen av teknologi gir også muligheter til stadig nyere oppdagelser og videre forskning. Studien er skrevet av Eivind Hundven Kjenes og Anders Naustdal i samarbeid med forskere fra USA og Norge.

Sammendrag

I oppgaven tar vi for oss menneskets oppfatning av moral, og i hvilken grad vi tilegner en kunstig intelligent programvare evnen til å handle moralsk riktig. Det er gjort særdeles få studier innenfor dette området. For å forstå og forklare mer om emnet, har vi ved hjelp av et teoretisk rammeverk forankret i forskningsarbeid, prøvd å belyse dette temaet i enda større grad.

Studier viser at mennesket kan oppleve negative emosjoner i en urettferdig prissetting av en vare grunnet forskjellige årsaker, også på vegne av tredjepart. Ved å se på forskning gjort innen moral, og dens kompleksitet, oppleves det at situasjoner med et langt svakere offer vil bli ensidig sett på som moralsk forkastelig.

Det kommer tydelig fram at vi ikke skiller mellom urett gjort av en programvare utstyrt med kunstig intelligens, og den samme handling gjort av et menneske. Forskjellen i variablene blir anerkjent, men ikke nok til at vi kan si at det merker seg tydelig ut. Vår anbefaling til videre studier innen feltet, er å utføre forsøket i en større, internasjonal skala. Dette vil belyse problemstillingen på et mer globalt nivå, og ta høyde for kulturforskjeller og politisk ståsted.

Helt fra den første dampmaskinen ble patentert en gang rundt 1780, har vi mennesker utviklet maskiner som gjør det mulig å erstatte vår egen muskelkraft og intelligens. I kjølvannet av denne utviklingen har vi i løpet av de siste 10-20 årene kommet til et punkt hvor det ikke bare er fysiske maskiner som erstatter våre arbeidsoppgaver, men programvarer som er enda mer effektive og feilfrie enn oss mennesker. Klaus Schwab (2016) skriver at vi nå er i begynnelsen av den fjerde industrielle revolusjon, og at denne kommer til å fundamentalt endre alle aspekter ved livet vårt. Den vil også ha en hastighet og gjennomslagskraft som vi aldri har sett før.

I USA ble det publisert en forskningsrapport hvor de så på sannsynligheten for at ulike jobber blir erstattet av roboter i fremtiden. De kom frem til at rundt halvparten av jobbene vil være i faresonen for å bli erstattet av roboter (Frey & Osborne, 2013). Når det gjelder Norge, er digitaliseringen i arbeidslivet kommet så langt at det vil være naturlig å tro at dette vil være gjeldende også her. (Carlin et al., 2015)

SINTEF skriver i sin rapport om teknologitrender som påvirker transportnæringen, at vi innen kort tid kan forvente å måtte forholde oss til både selvkjørende tog, båter og fly (Bakken et al., 2017). Bilmerket Tesla har allerede selvstyrende biler på markedet, men det er i de fleste tilfeller en person bak rattet som må ha kontroll på trafikken og ta over kjøringen ved behov. Det er i dag 14 andre selskaper som er i gang med utviklingen av selvkjørende biler (Endsley, 2017).

En annen følge av utviklingen, er bruken av elektroniske spor, også kalt stordata. I en rapport fra Norsk regnesentral oppsummeres det at bruken av elektroniske spor bare vil øke i tiden fremover, og ikke kommer til å bremse med det første. De nevner også at de aller fleste systemer, som behandler elektroniske spor, har mye større muligheter enn det som gis rom for, eksempelvis i personopplysningsloven. (Danielsson, Groven, Kristoffersen, Rivertz & Skomedal, 2005, s. 11)

Uansett hvor man beveger seg, møter man på en eller annen form for kunstig intelligens som har en funksjon eller nytte for oss. Drar man på handletur vil man

for eksempel møte kunstig intelligens i mange forskjellige former. Eksempler på dette kan være lys som indikerer tilgjengelige parkeringsplasser i parkeringshuset, digitale parkeringsautomater, prissettings-systemer, selvbetjente kasser og automater for tobakksvarer. Bare for 10 år siden fantes disse forskjellige funksjonene i svært liten grad, eller var de ikke eksisterende.

Den store etableringen av kunstig intelligens vekker følelser i mennesket. I San Francisco har roboter og selvkjørende biler gjentatte ganger blitt angrepet av sinte mennesker (Wong, 2018). I mars 2018 ble en kvinne påkjørt og drept av en slik selvkjørende bil som var bare ett av selskapet Uber sine mange test-kjøretøy. Uber stanset all testing av sine førerløse biler etter ulykken. Det første dødsfallet forårsaket av selvkjørende biler, skjedde i 2016 da en Tesla i autopilot-modus, opplevde en systemsvikt (Wong & Levin, 2018).

Hvis man tenker seg et scenario hvor to menn i 60 årene har blitt diagnostisert med kreft. Den ene er en hardtarbeidende bidragsyter i samfunnet, og den andre er en narkoman med liten sannsynlighet for å overleve de neste fem årene, uavhengig av behandling eller ikke. Hvem av de ville stille først i køen, om det var en kunstig intelligent selekterings maskin som skulle ta avgjørelsen ved hjelp av data og medisinsk historikk? Maskiner tilegner seg fordommer og stereotypi av andre, men det er lite forskning i dag som tar for seg hvordan mennesker bedømmer maskinens evne til å ta moralske beslutninger.

I et intervju til NRK forteller professor i filosofi ved Universitetet i Agder, Einar Duenger Bøhn, at når vi i større grad gir over kontrollen til systemer, vil vi jo ikke at de skal være umoralske (Tolfsen, 2018). Vi vil at de skal ta de rette valgene. Sammen med CAIR jobber Bøhn for å gi kunstig intelligens et moralsk kompass. CAIR er et institutt for forskning av kunstig intelligens ved Universitetet i Agder. Bøhn mener det er to måter å forstå forskningen på. Den ene er om det er mulig å simulere moral, noe det er. Bøhn tar for seg «Iphone» sin kunstig intelligente assistent Siri, som vil kunne svare deg høflig uavhengig av hva du sier til henne. Dette er ikke fordi Siri har en evne til å handle moralsk, det er fordi hun er programmert til det. Den andre måten å forstå forskningen på er at de kunstig intelligente systemene tar moralske valg selv. Dette er enda ikke en realitet, og

Bøhn nevner at forskningen på kunstig moral ikke har kommet noe lengre siden 1950-tallet.

Vi ønsker derfor å se nærmere på bruken av kunstig intelligens. Vi vet at en maskin som automatisk justerer prisene i en butikk, bare er en ansamling av informasjon og data, og at den har «lært» å være sånn. Matthias Scheutz (2016) skriver i sin artikkel at dagens roboter er enkle med begrenset autonomi. Forskning viser at fremtiden vår peker mot mer autonome systemer med potensiale til å skade mennesket. Dagens algoritmer tar ikke hensyn til det moralske aspektet. De kan heller ikke på noen måte å oppdage om de begår et moralsk brudd, basert på de handlingene de velger å utføre. Etter hvert som utviklingen fortsetter, vil det haste med å ta hånd om de problemene som gjør at autonome systemer er i stand til å handle på en måte som vil være moralsk forkastelig. Autonome roboter sees å være fremtiden og vil ha stor nytteverdi for samfunnet i tiden som kommer. (Müller & Bostrom, 2016, s. 525-526)

Basert på disse teoriene, sitter vi igjen med spørsmålene: Vil vi som mennesker stole på at programvare setter de riktige prisene? Vil vi for eksempel synes det er riktig at den skruer opp prisen på is når det er fint vær, siden etterspørselen da går opp? Eller, noe som er enda mer alvorlig, vil vi at den skal skru opp prisene på vaksine under et epidemiutbrudd? På bakgrunn av vårt tema og forskningsmetode har vi valgt problemstillingen:

“I hvilken grad vi mennesker tilegner kunstig intelligente systemer og maskiner evnen til å handle moralsk”

Vi mener vi har et bredt grunnlag for å besvare problemstillingen med hjelp av vår spørreundersøkelse og det teoretiske rammeverket vi benytter oss av. Dette er et felt som svært få har skrevet om tidligere, noe som gjør det spennende og motiverende for oss som skriver oppgaven.

Teoretisk rammeverk og hypoteseutvikling

Stordata (Big data)

Mange har påstått at bruken av stordata kan være en ny vekstfaktor i økonomien. Over hele det globale markedet er det mange som ser potensialet i det å bruke stordata til å transformere selskaper, markedsstrategier og interaksjoner med kunder. I boken «Creating Value with Big Data Analytics» (2016), skriver de at stordata vil endre markedsanalyser og hvordan markedsføring vil bli utført i de kommende tiår. Vi har allerede sett flere forandringer den siste tiden, hvordan markedsførere bruker analyse i deres avgjørelser. Disse har blitt mer faktabaserte samtidig som markeds- og forbrukerinnsikt har blitt veldig viktig i hvordan vi utformer strategier. Hvor det for 30-40 år siden var forskere innen markedsføring som publiserte sine utviklede modeller, observerer vi nå i ledende organisasjoner en sterkere tilstedeværelse av analytiske verktøy. Disse søker etter innovative og effektive måter å analysere deres data og skape innsikt. Den økende tilgjengeligheten av store biter med data vil bidra til å sette enda mer fart på utviklingen. Det ultimate målet for markedsanalyse og spesielt stordata-analyser, er å skape verdier for kunden og organisasjonen (Verhoef et al., 2016)

Rettferdighetsteori

I følge Robert Folger og Russel Cropanzano (2001), legges det innenfor rettferdighetsteori til grunn, tre spesifikke kontrafaktiske utsagn som parallelle. Dette er “ville stille”, “kunne stille” og “burde stille”-spørsmål. Generelt sett viser rettferdighetsteori at en organisasjon vil bli holdt ansvarlig for sosialt uansvarlig atferd, dersom det foreligger tre forhold: Den nåværende tilstand er ikke like gunstig som et tenkt alternativ (ville), handlingen som forårsaket denne tilstanden kunne ha blitt unngått (kunne) og handlingen brøt en moralsk standard (burde). Dette vil gi en annen måte å tenke på. Vår oppfattelse og tolkning av informasjon ikke er en mekanisk prosess. Det er heller en prosess hvor vi hjelp av fantasi kan bruke egne tanker og erfaringer for å skape en mer kompleks tenkemåte. Videre skriver de at ved hjelp av fantasi, sammenligner vi hva som kunne skjedd i alternative scenarier under ulike forhold, handlinger og agenter. Det synes mest

sannsynlig at individer faktisk skiller urettferdigheter begått av enkeltpersoner fra de som blir begått av organisasjoner. Gitt at de begge kan holdes ansvarlige for urettferdighet, er det viktig å utforske den psykologiske dynamikken ved hvilke beslutninger om ansvar som er truffet (Cropanzano, Chrobot-Mason, Rupp & Prehar, 2004)

Det finnes lite forskning på området innen rettferdighetsoppfatning der mennesket skiller mellom urett begått av kunstig intelligens versus mennesker, men det vil være naturlig å kunne trekke paralleller mellom overnevnte rapport og vår studie.

Rettferdighet og pris

Prisrettferdighet kan defineres som forbrukerens vurdering og tilhørende følelser om forskjellen (eller mangelen på forskjell) mellom selgerens pris og prisen på en komparativ annen part. Det vurderes om prisen er rimelig, akseptabel eller forsvarlig. (Weisstein, Monroe & Kukar-Kinney, 2013; Xia, Monroe & Cox, 2004).

Studier viser at forbrukeren dømmer hvorvidt prisen er urettferdig, avhengig av informasjonen som er tilgjengelig. Forbrukeren vil ikke automatisk bedømme økte kostnader som en rettferdig grunn til å øke prisene på en vare (Vaidyanathan & Aggarwal, 2003). Fra et forbrukerperspektiv ser det ut til at prisforskjeller på et produkt bare er rettferdig om det kan tilskrives kvalitetsforskjeller (Bolton, Warlop & Alba, 2003). Når forbrukeren vurderer kostnadene til bedriften i en slik sammenheng, vil de etter all sannsynlighet ignorere alle kostnader bedriften har, bortsett fra kostnaden for selve varen (Haws & Bearden, 2006)

Tidligere forskning demonstrerer at det er en rekke psykologiske faktorer som påvirker forbrukerens respons på pris, og hva de mener er rettferdig i forhold til prissetting. (Ferguson, Ellen & Bearden, 2014; Haws & Bearden, 2006; Maxwell, 2002; Xia et al., 2004). I en studie gjort på forbrukere, indikeres det at prisendringer gjort innenfor svært korte tidsperioder oppfattes som mer urettferdig enn endring over lengre tidsperioder. Det ble funnet bevis på at forbrukernes oppfatninger av rettferdighet kan påvirkes, både ved å oppleve et gunstig utfall, eller et mindre gunstig utfall (Haws & Bearden, 2006) . Hvis kjøperen oppfatter selgerens motiver som negative vil de evaluere organisasjonen og dens oppførsel i

negativ retning (Xia et al., 2004). Videre viser Sarah Maxwell (Maxwell, 2002) at forbrukere vil føle en økt urettferdighet om de får vite at en selger ikke opptrer etter ærlige prissettingsnormer, enn om de ikke vet noe angående prissettingen. Dette forteller oss at det er en fordel å la kundene få vite om grunnen til prisen. Også forskningen til Ferguson, Ellen og Bearden (2014) viser at den største urettferdigheten for forbrukeren oppstår ved at selger setter en urettferdig pris. Under naturkatastrofer og epidemier vil dette skape mulighetsområder for selgeren, hvor etterspørselen vil føre prisen til et relativt høyt nivå. Det er i slike tilfeller at man ser viktigheten av det å skape tillit gjennom «goodwill», da det øker oppfattelsen av rettferdighet. Kunder forventer at selgeren skal verdsette forholdet mellom kunde og selger gjennom det å tilby rettferdige priser. På den annen side kan kunden oppfatte selgerens handlinger som et svik og forverrer deres oppfattelse av urettferdighet. (Xia et al., 2004)

I kjølvannet av disse teoriene ønsker vi altså å gå enda dypere i den moralske oppfatningen av mennesket og maskin. Både når det kommer til oppfatningen og utførelsen av handlinger. I en situasjon der kunstig intelligens er programmert til å overvåke markedet og følge den grunnleggende regelen for pris og etterspørsel, vil den gjennom innsamling av enorme mengder data, forstå hvilke prisnivå som er aktuelt, uavhengig av vare. Vår første hypotese er derfor:

H1: Mennesker vil oppfatte prisen som mer rettferdig, om det er en robot som setter den.

Pris og Moral

I vår studie ønsker vi blant annet å undersøke forskjellen mellom hvordan vi oppfatter mennesker og kunstig intelligente systemers evne til å skille mellom rett og galt. I boken «The Mind Club» (Gray & Wegner, 2016), mener de at vi sorterer verden og dens enheter inn i to grupper, eller to dimensjoner. Disse kalles «experience» og «agency». «Experience» forklarer hvordan sinnet er fra innsiden. Hvordan føles det å være en person, en katt, hund eller flaggermus? Fordi dette eksisterer på innsiden av sinnet, kan det være vanskelig å beskrive eller forstå for andre. «Agency» er lettere å forstå fordi det omfatter hvordan sinnet er fra utsiden. Vi kan fastsette en enhets evne til å planlegge og tenke, bare gjennom å

observere dens handlinger og reaksjoner (Gray & Wegner, 2016). Dimensjonene beskrives også i forskningsartikkelen «Mind Perception Is The Essence Of Morality» (Gray, Young & Waytz, 2012). Her forteller de at «Experience» er den oppfattede kapasiteten for følelser som sult, frykt, smerte, glede og bevissthet. «Agency» beskrives som den oppfattede kapasitet til å ha intensjoner, og til å handle (Selvkontroll, bedømmelse av andre, evnen til å ta avgjørelser, kommunikasjon, refleksjon, tanker og minne). En enhet som for eksempel det voksne mennesket, kan score høyt i begge dimensjonene. Enheter som scorer lavt på experience og høyt på agency kan eksempelvis være Gud, Google eller Microsoft. Barn og dyr, kanskje spesielt kjæledyr, scorer høyt på «experience» og lavt på «agency». Til og med livløse og avdøde enheter er med i modellen. Disse skårer lavt på begge (Gray et al., 2012). Dette beskriver de også i boken «Dimensions of Mind Perception» (Gray, Gray & Wegner, 2007)

Innenfor etikkens verden, var det først Aristoteles som var inne på dette med moralske agenter og sårbare pasienter. Han skiller mellom «sårbare følere» og «tenkende handlere» (Stigen, 1973). Moralske agenter er de enhetene som scorer høyt på skalaen for agency. Dette er de enhetene som gjør umoralske eller moralske handlinger, heltene og skurkene som oppfører seg godt og vondt. Moralske pasienter er de som er på mottakersiden av disse moralske/umoralske handlingene. De drar fordelene av det gode og er offer for det onde. (Gray et al., 2007; Gray & Wegner, 2016) Linjene mellom menneske og maskin blir stadig mer utydelig, og oppfattelsen av kunstig intelligens som noe livløst uten evne til å tenke selv, begynner å bli mer og mer fjern.

Noen vil mene at en organisasjon eller bedrift ikke kan ha moral, og vi må trekke et skille mellom mennesker og organisasjoner. Men kan man egentlig det? En organisasjon er fullstendig klar over sine handlinger, og det samles kontinuerlig inn data om konsekvensene av disse. Arbeidsforholdet til de ansatte, deres effektivitet, produktivitet og arbeidsmiljø er ting som er under stadig overvåkning. Professor Kenneth E. Goodpaster og John B. Matthews, mener at med den mengde data og kunnskap man har i en organisasjon, er det ingen grunn for at organisasjonen kan utvise like stor rasjonalitet og respekt for personer som det individuelle menneske kan (Shaw & Barry, 2015, s. 207). Filosofen Peter French, mener at organisasjonen absorberer og utstråler intensjonene og

handlingene fra det enkelte mennesket, og derfor er det en organisatorisk beslutning. Han tror at organisasjonenes intensjoner er nok til å mene at bedriftens handlinger blir gjort «med vilje» og derfor er bedriften moralsk ansvarlig (Shaw & Barry, 2015).

Innenfor det som er naturlig, vil oppfattelsen av kunstig intelligens være at det er et livløst objekt som har blitt programmert til å tjene mest mulig penger, uten og ta hensyn til hvilke følger det vil få for andre. Vil man da skille mellom handlinger utført av en programvare kontra et menneske? Dette ønsker vi å finne ut i vår andre hypotese.

H2: Mennesker vil oppleve at det er mer moralsk riktig, om en robot setter prisen på en vare.

I artikkelen «Mind Perception is The Essence of Morality» (Gray et al., 2012) skriver de om et eksempel hvor man spiller tennis alene. Ser man for seg at man står og slår ballen i veggen gjentatte ganger, merker man fort at noe mangler. Uten en å spille mot vil det neppe være mulig å kalle det tennis. Det samme gjelder moral, hvor de i artikkelen beskriver forholdet mellom bevisste moralske agenter og lidende moralske pasienter. Hvis moralens kjerne består av kombinasjonen skadelig hensikt og smertefullt utfall, vil det si at handlinger begått av agenter med større bevissthet som òg resulterer i mer lidelse, bli vurdert som mer umoralsk.

Loven er relativt klar på at man får mer skyld for en bevisst handling enn for en handling som er utilsiktet. Som de skriver i (Gray et al., 2012), er intensjonene til en handling så sterkt knyttet til skylden av handlingen at selv irrelevante intensjoner kan øke skyldspørsmålet. Dette kommer også frem i et eksperiment utført av (Woolfolk, Doris & Darley, 2006), hvor de avdekker at folk som blir tvunget til å skyte noen, oppfattes som mer umoralske når de faktisk ønsker vedkommende død. Basert på disse teoriene ønsker vi å se nærmere på hvordan individer skiller mellom menneske og kunstig intelligens, når de tilegner “agenten” disse intensjonene, og i hvilken grad man bør straffes for dette. Vår tredje problemstilling blir derfor:

H3: Man vil gi en kunstig intelligent programvare mindre kritikk, enn et menneske hadde fått for samme beslutning.

Den menneskelige Agenten

Som tidligere nevnt i introduksjonen er utviklingen av kunstig intelligens i dag kommet langt. Dagens maskiner er i stand til å tilegne seg enorme mengder med kunnskap og evner ved hjelp av datainnsamling, og de er i stadig utvikling. En maskin er i et menneskets sinn ikke i stand til å føle eller tenke selv. En mobiltelefon kan for eksempel ikke noe for at den får systemsvikt, eller at den ikke har dekning på enkelte steder, men i de fleste tilfeller er det mobiltelefonen som må ta støyten. Vi liker å ha noen å skyld på, selv om de fleste vet at det er brukeren selv, eller produsenten som har gjort noe feil i prosessen. I slike tilfeller kan vi kalle mobilen for en moralsk agent uten experience, siden personen som kaster mobiltelefonen i veggen, ikke syns synd på den, men heller blir oppgitt over sin egen mangel på kontroll (Gray & Wegner, 2016

). Deborah G. Johnson (2006) argumenterer for at datasystemer ikke møter kravene for å kunne kalles en moralsk agent. De kan ikke handle med mening fordi de ikke har friheten til et menneske.

I en studie om moralske signaler, ble kvinner forespeilet to scenarioer. I det ene scenarioet har kvinnen vært utro, og mannen velger å utføre vold mot kvinnen. I det andre scenarioet, var kvinnen utro, men mannen valgte å utføre vold mot kvinnens katt istedenfor. Resultatet fra studien var at vedkommende som utførte vold mot katten ble sett på som en person med dårligere moralsk karakter.

Studiene viste altså at en handling som kan bli sett som mindre alvorlig enn en annen, kan samtidig signalisere mer alvorlige mangler i egenskaper som empati (Tannenbaum, Uhlmann & Diermeier, 2011).

En amerikansk studie ønsket å finne svar på om mennesket ville tilegne offerløse situasjoner et offer. De fant de bevis på at deltagerne ofte nominerte individet som offeret, spesielt når det kom til å bruke narkotika, incest, prostitusjon og selvmord. Det vil si at de tolket det samme individet som både forbryteren og offeret.

Deltagere nominerte også lik som et offer. (DeScioli, Gilbert & Kurzban, 2012). Forskningen til Gray, Schein og Ward (2014) støtter opp under dette, og viser at

en umoralsk handling uten et offer, kan knyttes opp mot fysisk smerte. Dette også selv i helt harmløse situasjoner.

Med utgangspunkt i disse teoriene blir vår 4 og 5 hypotese utformet slik:

H4: De anerkjenner mer at Peter kan kommunisere og tenke enn en datamaskin.

H5: I hvilken grad synes forsøkspersonene at Peter er i stand til å føle (lykke, smerte, sult etc.).

Forskningsdesign og metode

Vi vil i dette kapittelet presentere hvilke metoder vi benyttet oss av, for å samle inn datamaterialet som skulle bygge opp under vårt tema og problemstilling. I følge (Johannessen, Tuft & Christoffersen, 2016) omhandler samfunnsvitenskapelig metode, om hvordan vi skal gå frem for å få informasjon om den sosiale virkeligheten, og hvordan situasjonen kan analyseres. Man skiller mellom kvantitativ og kvalitativ metode, hvor kvalitativ metode viser til egenskaper rundt et fenomen, mens kvantitativ viser til mengde eller antall. Ottar Hellevik (Hellevik, 1999) definerer kvantitativ metode som en fremgangsmåte der forskeren først systematisk skaffer seg sammenlignbare opplysninger om flere undersøkelsesobjekter av et visst slag, så uttrykker disse opplysningene i form av tall, og til slutt foretar en analyse av mønsteret i dette tallmaterialet. Det er naturlig at problemstillingen styrer metodevalget til en viss grad. Siden vi ville utføre et eksperiment, var det mest nærliggende for oss å velge kvantitativ metode. Dette ga oss muligheten til å hente ut informasjon som kan tallfestes, eller uttrykkes i form av tall.

Vignettmetoden

Datamaterialet vi har hentet inn til denne bacheloroppgaven er basert på vignettmetoden. Metoden ble utviklet på 1950-tallet av sosiologen Peter H. Rossi. Dette var en lite brukt metode i Skandinavia frem til 1980-tallet, da den ble mer vanlig innenfor forskning i sosialt arbeid (Jergeby, 1999). En vignettundersøkelse kan beskrives som korte historier (vignetter) om hypotetiske karakterer i spesifiserte omstendigheter, som respondentene er invitert til å gjøre seg opp en mening om (Finch, 1987). I en slik undersøkelse bør vignettene være realistiske for respondentene. Situasjonene som beskrives bør være realistiske og man bør unngå uvanlige figurer og hendelser. De bør være skrevet slik at deltakerne lett forstår alle detaljer i historien. I tillegg bør vignettene gi nok kontekstuell informasjon til respondentene for at de klart skal forstå situasjonen som blir illustrert, men være tvetydige nok til å sikre at flere løsninger finnes.

Man baserer seg på en grunnleggende forutsetning ved bruk av denne forskningsmetoden; teorien om at mennesket er rasjonelt og bastant i sine meninger, slik at det gjenspeiles en viss "lovmessighet" i de vurderingene vi gjør

oss og som gjør at vi vil vurdere to lignende situasjoner ganske likt (Jergerby, 1999). Det er derfor viktig i en slik undersøkelse at man ser på problemene ved det faktum at man i utgangspunktet kun får svar på hva mennesker mener og tror, og at de kan handle annerledes om situasjonen faktisk oppstår i deres virkelige liv. Det finnes lite forskning som sier noe om den eksterne validiteten til disse metodene, og som undersøker om determinantene som forklares av hypotetiske valg fra undersøkende respondenter, samsvarer med de som forklarer hvilke emner som gjør at man foretar lignende valg i virkelige situasjoner (Hainmueller, Hangartner & Yamamoto, 2015) Livsfaser og emosjonelle, relasjonelle og materielle forhold påvirker vår dømmekraft, hvordan man forstår og handler i situasjoner. Dette erkjennes av forskere som bruker vignettmetoden, men de hevder at metoden likevel er i stand til å få frem typiske trekk og tendenser (Jergerby, 1999).

Bruk av data

Dataene som tas i bruk når man driver med denne type forskning, kalles primær og sekundærdata. Man kan velge å bruke en av dem, eller en kombinasjon av begge.

Data som samles inn til eget formål, altså for å få svar på en konkret problemstilling kalles primærdata. Disse regnes som de «beste» data man samler inn for å utføre forskning. For å finne de data man trenger, går man gjerne ut i felten for å samle disse inn, enten fysisk eller digitalt. Derfor kalles primærdata også felldata. Data som allerede eksisterer, og som blir brukt til å støtte opp under problemstillingen kalles sekundærdata. Dette kan være lignende undersøkelser gjort til andre formål, eller teoretisk rammeverk rundt temaet man skriver om. (Silkose, Olsson & Gripsrud, 2010)

Validitet og reliabilitet

Kristen Ringdal skriver “Validitet går grovt sett på om man faktisk har målt det man vil måle” (Ringdal, 2006

s. 221) Med andre ord vil validitet si gyldigheten av noe. Det er fullt mulig å måle med stor del av nøyaktighet og pålitelighet og oppnå svært valide resultater flere ganger på rad. Likevel er det mulig å gjøre feil. Måler man noe annet enn det som var hensikten å måle, har man gjort en avgjørende feil, noe som ikke er

ønskelig. Reliabilitet forklares som påliteligheten i resultatene. Det vil si at om man gjentar undersøkelsen flere ganger vil man få de samme resultatene (Ringdal, 2006 s. 86)

Rekruttering og utvalg

Gjennomføring av eksperimentet ble gjort elektronisk via programvaren Qualtrics. For å skaffe deltakere, sendte vi rundt en invitasjon til å delta, hvor man fikk presentert studien, og henvist til en link som bar videre til spørreundersøkelsen. Invitasjonen var formulert slik:

“Vil du delta i studie? Klikk på denne linken, og delta i et psykologisk eksperiment om oppfattelse av pris og prisendringer. Eksperimentet tar under fem minutt, og du kan delta når og hvor du vil, fra mobil, nettbrett eller PC. Følg linken her:”

Invitasjonen ble sendt til venner, familie og bekjente av oss, som også har fått tillatelse til å sende linken videre. Vi har forbeholdt oss mot at studenter skal ta del i denne studien, og har holdt studien vekke fra sosiale medier, slik at validiteten i svarene skal styrkes. Studien ble sendt ut til rundt 150 personer, i hovedsak via e-post og tekstmeldinger. Antallet respondenter vi satt igjen med etter studien var gjennomført, var 121 stk.

Eksperimentets design – Vignettundersøkelsen

Vår undersøkelse er basert på en studie utviklet av professorer fra Norge og USA. Ansvarlig for undersøkelsen er vår veileder Mads Nordmo Arnestad. Siden vår intensjon var å belyse moralske aspekter i forhold til bruk av kunstig intelligens, var det nok med en vignett hvor den manipulererte variabelen var: menneske vs. kunstig intelligens. Deltakerne ble tilfeldig allokert til en av de to utfallene.



Som man ser i modellen over, kan undersøkelsen fordeles i tre deler. I første del ble man presentert selve vignetten, hvor problemstillingen kommer frem. Vignetten lyder som følger, her presentert med Programvaren PriceComp som uavhengig variabel:

“Det kan være vanskelig å sette riktig pris på varer og tjenester.

Prescriptions.com bruker derfor et selvstyrt, intelligent datasystem, PriceComp, til å fastsette pris på sine produkter. Selv om PriceComp opprinnelig ble programmert av mennesker, opererer den nå helt uavhengig og selvstendig, ved bruk av en dynamisk Bayesiansk algoritme. Den justerer priser basert på endringer i tilbud og etterspørsel.

Et av produktene som PriceComp fastsetter pris for er en influensavaksine.

En vinter oppstår det en spredning av et influensavirus som kan være dødelig for babyer. PriceComp registrerer at etterspørselen etter influensavaksinen øker dramatisk, ettersom mange foreldre vil ønske å vaksinere barna sine. Den øker dermed prisen på en vaksinedose fra 20 til 50 \$.”.

I andre del skulle respondentene gjøre sine vurderinger og anbefalinger i forhold til problemstillingen i vignetten. De fikk både spørsmål og påstander hvor de

skulle svare alt i fra “svært uenig” til “svært enig”. Vi vil beskrive spørsmålene og skalaene mer grundig når vi kommer til Utfallsmål. Den tredje delen av undersøkelsen var for å skaffe oss et bilde av hvilke respondenter som hadde deltatt. Der skulle deltakerne taste inn blant annet kjønn, alder, politisk ståsted og utdanning. Denne delen var helt lik for alle respondenter.

Utfallsmål

Opplevd rettferdighet:

Denne variabelen består av tre ledd. Målet med variabelen var å finne om respondentene opplever det som mer rettferdig om en robot satte prisen på vaksinen eller om det var et menneske. Eksempler på leddene vi benyttet er: “Det var riktig å la Peter/PriceComp sette prisen på produktet” og “Det burde være forbudt og la Peter/PriceComp sette prisen på produktet”. Det var syv svaralternativ som gikk fra “svært uenig” (skåre 1) til “svært enig” (skåre 7). På sistnevnte spørsmål måtte vi reversere svarkategoriene for å få samsvar. (7=1, 1=7 etc.)

Opplevd moralitet.

Denne variabelen består av 4 ledd som skal måle i hvilken grad utvalget mener at Peter eller PriceComp handlet moralsk. Eksempler på leddene er “Hvor moralsk var Peter/PriceComps handling” og “Hvor rett eller gal var Peter/PriceComps beslutning”. Her var også svarkategoriene fra 1 til 7. (1=svært umoralsk/gal, 7=svært moralsk/rett).

Ris og Ros:

Målet for denne variabelen er å fange opp hvorvidt det menes at Peter/PriceComp bør sanksjoneres for sine beslutninger. I denne variabelen er det to ledd: “Hvor mye kritikk eller ros bør Peter/PriceComp få?” og “Hvor mye straff eller belønning bør Peter/PriceComp få?” Svaralternativene har en skala fra 1 (svært mye kritikk/Sterk straff) til 7 (svært mye ros/sterk belønning)

Agens: Med denne variabelen ønsker vi å måle hvordan utvalget tilegner Peter/PriceComp egenskaper eller evne til å handle. Denne variabelen består også av 2 ledd: “Peter/PriceComp er i stand til å kommunisere” og “Peter/PriceComp kan tenke”. Her gikk skalaen fra 1 (helt uenig) til 7 (helt enig).

Opplevd sensitivitet:

Vi ønsket at denne variabelen skulle si noe om hvorvidt respondentene opplever at Peter/PriceComp har følelser eller er sensitiv. Spørsmålene “Peter/PriceComp er følsom for smerte” og “Peter/PriceComp kan oppleve smerte” hadde en skala fra 1 til 7, hvor 1 var helt uenig og 7 var helt enig.

Resultater

Chronbach's alpha – Reliabiliteten i svarene

Variabel	Utvalg	Cronbach's Alpha (α -verdi)
1. Rettferdighet	Q26, Q28, Q29	0,890
2. Moralitet	Q37, Q38, Q40, Q41	0,896
3. Ris eller ros	Q36, Q39	0,813
4. Agens	Q59, Q60	0, 874
5. Sensitivitet	Q62, Q63	0,946

For å kunne si noe om konsistensen i svarene, har vi målt «Cronbach's alpha» for de avhengige variablene. Er det liten grad av konsistens beveger alpha-verdien mot null. Er det høy grad av konsistens, beveger den seg mot 1. Den optimale verdien ligger på rundt 0,7. Som vi kan se i tabellen over, er det høy konsistens i alle variablene vi har benyttet i studien, noe som forteller oss at reliabiliteten er god.

Demografiske data

Som vi vil vise i tabellen på neste side, er det en relativt lik kjønnsfordeling. Det er ca. 16% større andel kvinner enn menn. Flesteparten respondentene er mellom 20-30 år (ca 54%). Dette kan forklares ved at undersøkelsen ble også ble sendt rundt til vår omgangskrets, og at det var mest sannsynlig at de ville svare på undersøkelsen.

Ser vi på hvilken sektor deltakerne jobber i, er det der også en ganske jevn fordeling. 51% jobber i privat sektor, mens resterende 49% jobber i offentlig sektor. 56% av respondentene har en bachelorgrad som høyeste utførte utdanning, og de fleste respondentene tjener mellom 400 000 og 500 000kr i året. Det vi også kan se er at det er ingen personer over 70 år som har svart på undersøkelsen.

<u>Kjønn</u>		
Mann	50	42,32
Kvinne	71	58,67
<u>Alder</u>		
20-29	66	54,54
30-39	32	26,45
40-49	9	7,44
50-59	9	7,44
60-69	4	3,31
Over 70	0	0
<u>Høyeste fullførte utdanning</u>		
Videregående	35	28,93
Bachelorgrad	68	56,20
Mastergrad	17	14,05
Doktorgrad	1	0,83
<u>Sektor</u>		
Privat	60	49,59
Offentlig	59	48,76
<u>Lønn per år (i NOK)</u>		
0 – 100 000	3	2,48
100 000 – 200 000	8	6,61
200 000 – 300 000	6	4,96
300 000 – 400 000	24	19,83
400 000 – 500 000	36	29,75
500 000 – 600 000	24	19,83
600 000 – 700 000	8	6,61
700 000 – 800 000	10	8,26
Over 1000 000	1	1,65

Sammenheng mellom variablene

For å skaffe oss et bilde av hvor godt de ulike variablene i undersøkelsen samhandler, har vi kjørt en korrelasjonsanalyse. Alle verdier under 0,3 viser at det er liten sammenheng mellom variablene. Ut i fra tabellen ser vi at det er ingen systematisk forskjell på alder og hvor rettferdig de mener prissettingen er, da denne verdien er -0,11. Vi ser også at respondentenes syn på prisrettferdighet er heller ikke nevneverdig påvirket av utdanning eller lønn, med verdiene -0,04 og 0,05. Variablene Alder, utdanning og lønn satt opp mot moralitet, har verdiene -0,02 (alder), -0,00 (utdanning) og -0,03 (lønn). Dette viser oss at respondentene fant både Peter og PriceComps handlinger i forhold til prissetting umoralsk kritikkverdig uavhengig av hvem som utførte de. Vi finner ingen sammenheng mellom opplevd sensitivitet og de tre variablene alder, utdanning og lønn. Variabelen agens, forklarer i hvilken grad deltakerne mener Peter og PriceComp er i stand til å kommunisere med andre og tenke selv. Her finner vi at det er en sammenheng med alder (0,63). Dette vil si at jo eldre respondentene er, jo mer anerkjenner de at Peter og PriceComp er i stand til å tenke og kommunisere.

Korrelasjonstabell

Variabel	1.	2.	3.	4.	5.	6.	7.	8.
1. Rettferdighet	___							
2. Moralitet	0,71**	___						
3. Anerkjennelse	0,73**	0,93**	___					
4. Sensitivitet	0,35**	0,52**	0,53**	___				
5. Agens	0,49**	0,71**	0,73**	0,85**	___			
6. Alder	-0,11	-0,02	-0,05	0,02	0,63	___		
7. Utdanning	-0,04	-0,00	0,09	-0,11	-0,04	0,08	___	
8. Lønn	0,05	-0,03	-0,02	0,06	-0,07	0,41**	0,38**	___

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Test av mellomgruppeforskjeller

Vår hypotese H1 tilsa at deltakerne som leste om en robot, opplevde prissettingen som mer rettferdig enn de som leste om et menneske. Ved bruk av en uavhengig t-test fant vi at det var ingen signifikant forskjell i resultatene for PriceComp ($M=1,28$, $SD=0,67$) og Peter ($M=0,27$, $SD=0,71$). Forhold; $t(130) = 0,013$, $p = 0,990$. Vi ser at tendensene heller i riktig retning, men finner ikke støtte for H1.

Med den andre hypotesen vår; H2, ønsket vi å finne ut hvorvidt forsøkspersonene mente det var mer moralsk riktig når det var en prisrobot som satte prisen på vaksinen, enn når det var et menneske. T-test resultatene ble som følgende for PriceComp ($M=1,33$, $SD=0,47$) og for Peter ($M=1,24$, $SD=0,51$). Forhold; $t(130) = 1,04$, $p = 0,297$. Resultatet tenderer i samme retning men er ikke statistisk signifikant, og H2 støttes dermed ikke.

I vår tredje hypotese; H3 ville vi se i hvilken grad respondentene ville sanksjonere beslutningen som ble tatt, avhengig om det var PriceComp eller Peter som tok beslutningen. T-testen viste at for PriceComp ble resultatet ($M=1,43$, $SD=0,44$) og for Peter ble det ($M=1,34$, $SD=0,52$) Forhold; $t(130) = 0,99$, $p = 0,322$. Undersøkelsen viser at det skal være mer ros til roboten, men 1,4 på en skala fra 1-7, viser at det er mye kritikk til begge, men igjen, fortsatt en liten forskjell. Tendensene går igjen hele veien, men merker seg ikke tydelig ut. Vi finner derfor ikke støtte for H3.

I den fjerde hypotesen; H4 ville vi se om respondentene anerkjenner Peters evne til å tenke og kommunisere mer enn de anerkjenner PriceComps evner til det samme. Resultatet fra t-testen er for PriceComp ($M=1,70$, $SD=0,68$) og for Peter ($M=2,42$, $SD=0,94$). Forhold; $t(130) = -4,99$, $p = 0,00$. Utvalget er ikke imponert over Peters evne til å tenke selv. Det vi ser er at de mener PriceComp sine evner er enda dårligere. Dette forteller oss at det er betydelig hvordan beslutningstakeren fremstår, og H4 støttes.

Den femte og siste hypotesen; H5 går ut på i hvilken grad respondentene mener at Peter er i stand til å kjenne følelser som lykke og smerte. Resultatene fra testen var for PriceComp ($M=0,93$, $SD=0,58$) og for Peter ($M=2,30$, $SD=0,99$). Forhold;

$t(130) = -9,5, p = 0,00$. Dette forteller oss at det er en forskjell, men den er veldig liten. Deltakerne mener at Peter i veldig liten grad er i stand til å kjenne følelser, men fortsatt mer enn en robot. H5 støttes.

Diskusjon

I denne delen av oppgaven vil vi se nærmere på og drøfte resultatene som har kommet fram i studien vi har gjort. Målet var å finne ut om mennesket skiller mellom moralske ugjerninger begått av et individ kontra en kunstig intelligent programvare. Det ville ifølge tidligere forskning være forskjell på bedømmelse av handlinger begått av en organisasjon og et menneske (Cropanzano et al., 2004) Med relativt lite forskning innenfor dette feltet, vil det være trolig at en kunstig intelligens hadde fått rollen som organisasjonen, og dømmes på samme måte. Ut i fra disse forventningene om reaksjoner ble resultatet noe annerledes enn forventet.

Oppsummering av undersøkelsen

Resultatet fra undersøkelsen viser at forsøkspersonene ikke skiller mellom PriceComp og Peter i graden av urettferdighet og umoralskhet knyttet til deres handlinger, selv om vi kan se at tendensene peker i forventet retning ble det ikke funnet signifikant støtte. Der vi finner forskjellene, er i deltakernes anerkjennelse av Peters menneskelige sider. De opplever at hans evne til å tenke, føle og handle på egenhånd som signifikant høyere enn alternativet, men verdiene er såpass lave at det må kunne sies at de ikke er særlig imponert over hans egenskaper til å tenke, føle eller handle. Noe som viser at vi ikke skiller mellom handlingene som er begått, selv om det anerkjennes at Peter er et menneske.

Rettferdighet er et diffust begrep. For de fleste kan rettferdighet beskrives av enten en eller to definisjoner: Den typen forhold der mennesker behandles i overensstemmelse med moralske prinsipp, eller Den type forhold der alle mottar det de fortjener. Aristoteles var blant de første til å beskrive hva rettferdighet er. Han beskrev det som evnen til å behandle like tilfeller likt, handle lovmessig, og til å ikke ta mer enn sin rimelige andel (Stigen, 1973). Vi ønsker også å beskytte de som er svakere enn oss, de som ikke har evnen til å forsvare seg selv.

Umoralske gjerninger blir da gjerne enda mer umoralsk om de går ut over de svakere i samfunnet, slik som det ble beskrevet i teoridelen med bevisste moralske agenter og sårbare moralske pasienter (Gray & Wegner, 2016).

Den første hypotesen ønsket å teste om deltakerne synes prisen satt av en robot var mer rettferdig. Gjennom programmering følger programvaren den grunnleggende regelen: Hva er riktig pris for en vare? For programvaren vil dette være høyest mulig. Med økt etterspørsel på det nødvendige gode, vil det være naturlig å øke prisen på varen. Resultatet fra testen av variabelen «opplevd rettferdighet» tilsier at de som leste om «PriceComp», ikke opplevde det som mer rettferdig enn de som leste om Peter, selv om «PriceComp» ikke har evnene til å ta moralske beslutninger på lik linje. Det var tendenser i resultatet som pekte i riktig retning men resultatet var ikke betydelig, og vi kan derfor ikke konstatere at forsøkspersonene skilte mellom de to alternativene. Overraskende nok, er dette motstridende resultater fra forskningen vi henviser til i innledningen hvor det indikeres at mennesket skiller mellom urett gjort av organisasjoner kontra mennesker (Cropanzano et al., 2004). Forklaringen på funnene vi gjorde, kan være at respondentene har forventninger til at bedriften som selger vaksinen skal vise «goodwill» under et epidemiutbrudd, og at hvem som setter prisen på vaksinen er uavhengig til deres oppfattelse av urettferdighet (Xia et al., 2004) I eksperimentet blir deltakerne satt inn i en situasjon som de fleste kan kjenne seg igjen i. Det er ikke ukjent at prisen justeres etter etterspørsel slik at mulighetene for fortjeneste maksimeres. I dette forsøket blir man tvunget til å tenke på de rundt seg og konsekvensene dette kan få. Hvem som gjør en slik ugjerning er muligens uvesentlig når det går på bekostning av seg og sine egne.

I en deskriptiv betydning brukes moral om det sett av moralske normer, verdier og holdninger som aksepteres og forsøkes praktisert av et individ eller en gruppe. I en normativ betydning brukes moral om det sett av moralske normer som innehar gyldighet, og som dermed representerer hvordan vi faktisk bør handle. «Moral må skilles fra etikette, fra loven og fra uformelle yrkes regler og etikk» (Shaw & Barry, 2015) . Variabelen «opplevd moral» viste oss at det ikke var tydelig at deltakerne skilte mellom moralen til «Peter» og moralen til «PriceComp». Vi kan se de samme tendensene der som i «opplevd rettferdighet» hvor det er en ørliten forskjell men på langt nær nok til at vi kan si at den er signifikant. Vår hypotese om at forsøkspersonene kom til å bedømme Peters handlinger som mer umoralsk enn «PriceComp» støttes derfor ikke. Det kan virke til at folk også er enig med filosofen Peter French i påstandene om at bedriftene har et moralsk ansvar og uavhengig av om intensjonen av handlingen reflekterte selve handlingen som ble

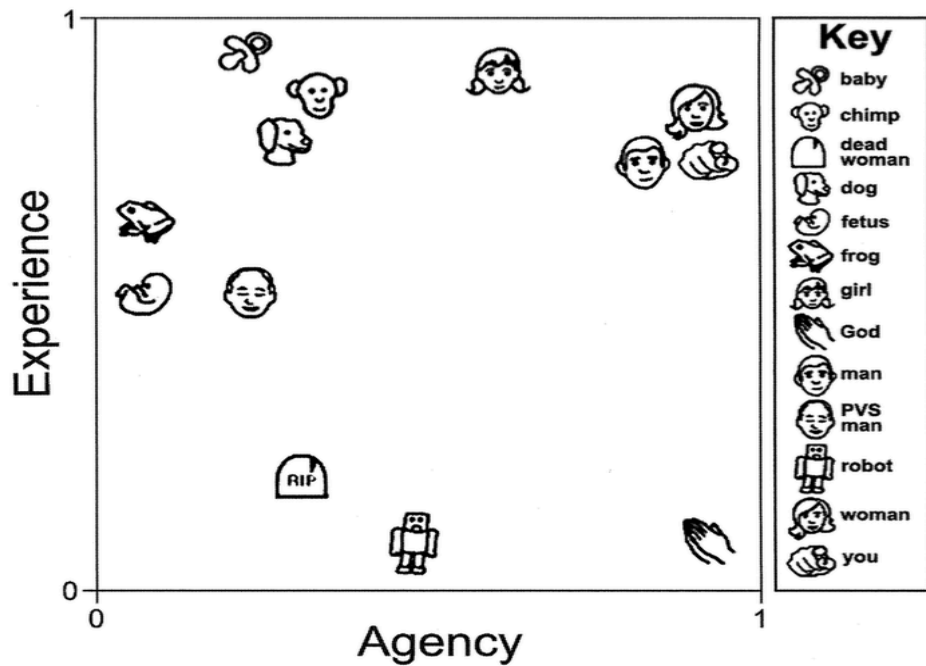
utført (Shaw & Barry, 2015, s. 207). Her kan man også trekke paralleller til (Gray & Wegner, 2016

), da både Peter og PriceComp blir dømt som moralske agenter uten noe form for experience. Grunnen til at vi får dette resultatet kan være at deltakerne ikke ser på PriceComp som en moralsk enhet, men heller bakmennene som har valgt å bruke enheten. Dette da vil stille PriceComp på noenlunde lik linje med Peter.

Under kategorien «Opplevd ris og ros» var det forventet at vi skulle finne en signifikant forskjell. Målet med denne variabelen var å finne ut om deltakerne ville straffe Peter mer enn PriceComp for samme beslutning, men resultatet ble på lik linje med de tidligere hypotesene forkastet. Vi ser de samme tendensene som tidligere nevnt, at roboten får litt mindre straff enn mennesket. Det ville også her, vært forventet at resultatet helte mot teoriene nevnt i artikkelen (Cropanzano et al., 2004) , men det var ingen betydningsfulle funn her heller.

Videre i undersøkelsen ser vi på Peters menneskelige egenskaper opp mot maskin. Forsøkspersonene bedømte Peters evne til å tenke selv, og vi satt igjen med variabelen «opplevd agens». I denne delen av undersøkelsen ser vi forskjell på mennesket og maskin, der de anerkjenner at Peter er et individ med evnen til å tenke på egenhånd. Vi finner her en signifikant forskjell, men det er absolutt verdt å nevne at deltakerne mener at Peters evne til å tenke selv er særdeles lav.

Anerkjennelse av Peters evner til å føle ble også satt på prøve. Resultatet fra variabelen «opplevd sensitivitet» viser oss at de til anerkjenner hans menneskelige evne til å føle, og resultatet av målingen er signifikant. Det som er interessant i dette tilfellet er at Peter dømt som en ganske kald og følelsesløs person selv om det var en forskjell sammenlignet med alternativet PriceComp.



For å kunne forklare resultatet av de to siste hypotesene har vi valgt å ta i bruk figuren over. Denne er hentet fra artikkelen «Dimensions of mind perception» (Gray et al., 2007). Mennesket og Kunstig intelligens blir dømt nokså likt, men Peter vil score høyere på dimensjonen for experience og derfor være nærmere det som i grafen kalles «voksne mennesker», mens PriceComp blir dømt på lik linje med Gud og alle de store «corporations». Dette var som nevnt over i forventet retning, men fortsatt er det noe overraskende at skillet er såpass lite som det kom fram i undersøkelsen.

Vår teori var bygget på forventningene om at Peter som agent ville bli dømt langt hardere enn PriceComp for sine handlinger. Peter er et tenkende vesen med samme bakgrunn for å bruke moral og skjønn som folk flest. Peters samhandling med mennesker gjennom livet, ville gitt han gode forutsetninger for å kunne ta moralsk gode avgjørelser og utvise skjønn i spesielt kritiske situasjoner, på vegne av sin bedrift. Dette i motsetning til PriceComp som kun baserer seg på graden av tilbud og etterspørsel i markedet. Selv med en enorm database av kunnskap, statistikk og tall er den fortsatt ikke programmert til å utvise samme skjønn eller ta de moralske hensyn som er forventet av et menneske. Evnen til å reflektere over rett og galt ligger i programmeringen til maskinen, og PriceComps moral vil trolig være tilsvarende moralen til den som har programmert den.

Alvorlighetsgraden og eventuelt konsekvensene av handlingene som blir utført påvirker også i hvorvidt folk vil føle en avsky overfor beslutningstakeren. Som det også kommer frem i resultatdelen blir Peters evne til å tenke selv, anerkjent. Begge to blir sett på som følelsesløse agenter, men maskinen oppfattes slik at handler uten skjønn og blir i den forstand bare en brikke i det store pengespillet.

Milton Friedman beskriver at det eneste samfunnsansvaret en bedrift har er å maksimere profitt, så lenge de følger spillets regler. Han mener at så lenge markedet får operere fritt med bare minimalistiske restriksjoner for å forhindre svindel, maktmisbruk, vil samfunnet maksimere sin økonomiske velferd. Det er opp til staten å ta ansvar for sosial målsetting og ha det offentlige ansvar (Shaw & Barry, 2015, s. 210-211). Kritikerne av denne teorien mener at bedriftene har et samfunnsansvar overfor alle parter den berører. Dette bringer oss videre til stakeholder teorien som forteller oss at bedriften ikke bare har økonomisk ansvar ovenfor andels-holderne, men for alle menneskene som har en påvirkning eller blir påvirket av bedriftens gjerninger. Sitter bedriften på muligheten til å redde liv, vil den få et moralsk ansvar overfor de påvirkede menneskene. (Shaw & Barry, 2015, s. 212)

Anbefalinger til videre forskning

Velferdssystemet i Norge, med gratis helsetjenester, trygd og pensjon er godt innarbeidet blant det norske folk. Dette kan bidra til at våre holdninger ovenfor medisinske spørsmål trolig vil bli annerledes enn i land som ikke besitter de samme velferdsordningene. Undersøkelsen ble ikke gjort i kontrollerte former, og deltakerne kunne besvare spørsmålene hvor de ville. Mye støy i omgivelsene, tretthet etter arbeid og fokus på andre oppgaver kan også ha spilt en rolle i deltakernes vilje eller tilstand til å reflektere over de spørsmål som ble stilt. For å få et bredere svar på vår problemstilling kunne vi ha gjennomført undersøkelsen i en større global skala for å jevne ut det politiske ståstedet til den typiske nordmann, hvor helsesektoren skal ta regningen. Det er også tenkelig at man ville finne større forskjeller der konsekvensene av ugjerningen som er begått hadde vært noe mindre.

Konklusjon

Resultatet av forskningen viste at deltakerne i eksperimentet følte en eller annen form for avsky over handlingene som ble presentert. Dette viser at uavhengig av alder, yrkesgruppe og sektor, var forsøkspersonene samstemte i sin oppfattelse av moral. Med andre ord viser våre funn at mennesket ikke skiller mellom de umoralske handlinger begått av en kunstig intelligent programvare og handlinger utført av et menneske. Våre forventninger til forsøket, var å bevise at en robot ikke ville bli dømt like hardt som et menneske for sine umoralske handlinger. Disse forventningene ble ikke innfridd. Det finnes svake tendenser til at det heller i den retningen, men som nevnt tidligere i oppgaven tenderte det ikke nok, til at vi kan kalle resultatet signifikant. Dette resultatet oppsto selv om det ble funnet tydelig støtte for at forsøkspersonene forstod at variabelen Peter, var et menneske med evnen til å handle og tenke på egenhånd. Selv om det i teorien skulle tilsi at mennesket skiller mellom urett gjort av organisasjoner og mennesket, slik vi kunne lese i artikkelen til (Cropanzano et al., 2004), kan vi ikke trekke den samme konklusjonen om forskjellen på handlinger utført av mennesket og kunstig intelligens. Utvalget kan sies å være representativt for det norske folk, da det er et bredden av forskjellige samfunnsklasser, både offentlig og privat sektor. Bruken av kunstig intelligens blir stadig mer utbredt, og vår forståelse av menneskets holdninger og oppfattelse av avanserte teknologiske programmer og hjelpemidler blir viktige. Spesielt gjelder dette i tilfeller der etikk og moral står sentralt. Da det er mer utbredt klasseskille og andre kjerneverdier som preger mennesket i forskjellige deler av verden er det å inkludere et større globalt mangfold i undersøkelsen. Det er da mulig at vi vil se signifikante forskjeller i den tenderte retningen.

Referanser

- Bakken, T., Carlin, M. S., Bjerkan, K. Y., Westerheim, H., Nordlander, T., Bahr, R., . . . Transeth, A. (2017). Teknologitrender som påvirker transportsektoren. *SINTEF Rapport*.
- Bolton, L. E., Warlop, L., & Alba, J. W. (2003). Consumer perceptions of price (un) fairness. *Journal of consumer research*, 29(4), 474-491.
- Carlin, M. S., Skjellaug, B., Nygaard, S., Vermesan, O., Svagård, I. S., Andreassen, T. W., . . . Røhne, M. (2015). Effekter av teknologiske endringer på norsk nærings-og arbeidsliv.
- Cropanzano, R., Chrobot-Mason, D., Rupp, D. E., & Prehar, C. A. (2004). Accountability for corporate injustice. *Human Resource Management Review*, 14(1), 107-133.
- Danielsson, F. J., Groven, A.-K., Kristoffersen, T., Rivertz, H. J., & Skomedal, Å. (2005). Elektroniske spor. *NR Report 1008, ISBN: 978-82-53-90516, 7*.
- DeScioli, P., Gilbert, S. S., & Kurzban, R. (2012). Indelible victims and persistent punishers in moral cognition. *Psychological Inquiry*, 23(2), 143-149.
- Endsley, M. R. (2017). Autonomous driving systems: A preliminary naturalistic study of the Tesla model S. *Journal of Cognitive Engineering and Decision Making*, 11(3), 225-238.
- Ferguson, J. L., Ellen, P. S., & Bearden, W. O. (2014). Procedural and distributive fairness: Determinants of overall price fairness. *Journal of business ethics*, 121(2), 217-231.
- Finch, J. (1987). The vignette technique in survey research. *Sociology*, 21(1), 105-114.
- Folger, R., & Cropanzano, R. (2001). Fairness theory: Justice as accountability. *Advances in organizational justice*, 1, 1-55.
- Frey, C. B., & Osborne, M. (2013). The future of employment. *How susceptible are jobs to computerisation*.
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *science*, 315(5812), 619-619.
- Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, 143(4), 1600.
- Gray, K., & Wegner, D. M. (2016). *The Mind Club*. USA Penguin books.
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological inquiry*, 23(2), 101-124.
- Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences*, 112(8), 2395-2400.
- Haws, K. L., & Bearden, W. O. (2006). Dynamic pricing and consumer fairness perceptions. *Journal of Consumer Research*, 33(3), 304-311.
- Hellevik, O. (1999). *Forskningsmetode i sosiologi og statsvitenskap*: Universitetsforlaget.
- Jergeby, U. (1999). Att bedöma en social situation: tillämpning av vinjettmetoden.

- Johannessen, A., Tufte, P. A., & Christoffersen, L. (2016). Introduksjon til samfunnsvitenskapelig metode (Vol. 5, s. 458): Abstrakt Forlag AS.
- Johnson, D. G. (2006). Computer systems: Moral entities but not moral agents. *Ethics and information technology*, 8(4), 195-204.
- Maxwell, S. (2002). Rule-based price fairness and its effect on willingness to purchase. *Journal of economic psychology*, 23(2), 191-212.
- Müller, V. C., & Bostrom, N. (2016). *Fundamental issues of artificial intelligence* (Vol. 376): Springer.
- Ringdal, K. (2006). Enhet og mangfold (Vol. 1). Bergen: Fagbokforlaget.
- Scheutz, M. (2016). The need for moral competency in autonomous agent architectures *Fundamental Issues of Artificial Intelligence* s. 517-527: Springer.
- Schwab, K. (2016). *The Fourth Industrial Revolution*. Geneva, Switzerland.
- Shaw, W. H., & Barry, V. (2015). *Moral issues in business*: Cengage Learning.
- Silkoset, R., Olsson, U. H., & Gripsrud, G. (2010). Metode og dataanalyse (Vol. 2). Bergen: Cappelen Damm.
- Stigen, A. (1973). Etikk. I Aristoteles (Red.), (Vol. 3, s. 148). Oslo: Gyldendal Norsk Forlag ASA.
- Tannenbaum, D., Uhlmann, E. L., & Diermeier, D. (2011). Moral signals, public outrage, and immaterial harms. *Journal of Experimental Social Psychology*, 47(6), 1249-1254.
- Tolfsen, C. (2018.). Når dataprogrammene får moral. Hentet fra <https://www.nrk.no/kultur/xl/nar-dataprogrammene-far-moral-1.13867044>
- Vaidyanathan, R., & Aggarwal, P. (2003). Who is the fairest of them all? An attributional approach to price fairness perceptions. *Journal of Business Research*, 56(6), 453-463.
- Verhoef, P. C., Kooge, E., & Walk, N. (2016). *Creating value with big data analytics: Making smarter marketing decisions*: Routledge.
- Weisstein, F. L., Monroe, K. B., & Kukar-Kinney, M. (2013). Effects of price framing on consumers' perceptions of online dynamic pricing practices. *Journal of the Academy of Marketing Science*, 41(5), 501-514.
- Wong, J. C. (2018.). Rage against the machine: self-driving cars attacked by angry Californians. Hentet fra <https://www.theguardian.com/technology/2018/mar/06/california-self-driving-cars-attacked>
- Wong, J. C., & Levin, S. (2018.). Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian. Hentet fra <https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe>
- Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition*, 100(2), 283-301.
- Xia, L., Monroe, K. B., & Cox, J. L. (2004). The Price Is Unfair! A Conceptual Framework of Price Fairness Perceptions. *Journal of Marketing*, 68(4), 1-15.