LUNDS UNIVERSITET

Ekonomihögskolan

Institutionen för informatik

Demografiska egenskapers påverkan på åsikter angående AI-partiskhet

En kvantitativ studie om möjlig vidareutveckling av Moral Machine

Kandidatuppsats 15 hp, kurs SYSK16 i Informatik

Författare: Jesper Gerdin Olsson
Alexander Kvarnmark

Handledare: Blerim Emruli

Rättande lärare: Björn Svensson
Odd Steen

Demografiska egenskapers påverkan på åsikter angående AI-partiskhet: En kvantitativ studie om möjlig vidareutveckling av Moral Machine

ENGELSK TITEL: Demograpics' Impact on Opinions Regarding AI-bias: A Quantitative Study on a Possible Expansion of Moral Machine

FÖRFATTARE: Jesper Gerdin Olsson, Alexander Kvarnmark

UTGIVARE: Institutionen för informatik, Ekonomihögskolan, Lunds universitet

EXAMINATOR: Christina Keller, Professor

FRAMLAGD: maj, 2021

DOKUMENTTYP: Kandidatuppsats

ANTAL SIDOR: 46

NYCKELORD: Artificiell intelligens, etik, partiskhet, självkörande bilar, demografiska egenskaper, Moral Machine

SAMMANFATTNING (MAX. 200 ORD):
Samhället blir alltmer AI-baserat och till följd av denna vetenskap framkommer frågan kring etik upp. Studien siktar på att utforska utvecklingsmöjligheter hos verktyg Moral Machine som används för att undersöka människans åsikter kring AI-partiskhet. I linje med studiens syfte har brister hos verktyget identifierats och utifrån framställd forskningsfråga har undersökningen baserats på att se ifall demografiska egenskaperna ålder och kön bör inkluderas i verktyget. Litteratur har framställts med intentionen att belysa de områden som sedan kopplas med det empiriska materialet. En kvantitativ empiri som efterliknar den av Moral Machine har genomförts för att se om dessa åsiktsskillnader mellan demografiska egenskaperna existerar. Efter analysering av data framkom det att en responsskillnad går att se angående kön och därmed kan idéen lyftas att Moral Machine bör se till utvecklingsmöjligheter för att uppnå optimala data.

Innehåll

Figurer

Tabeller

1 Introduktion

Detta kapitel ämnar gå genom uppsatsens övergripande omfång, förklara grundläggande information till ämnet artificiell intelligens (AI) samt presentera författarnas inriktning och mål.

1.1 Bakgrund

Artificiell intelligens (AI) är idag väl etablerat i samhället och har ett brett omfång av användningsområden. AI är verksam inom flera sektorer av arbetsmarknaden, men även i människors privatliv genom exempelvis digitala assistenter i hemmet, riktad reklam online eller i självkörande bilar. Även i offentliga sammanhang används AI för administrativt arbete, smart infrastruktur och i kampen mot Covid-19 (Guillot, 2021b). Med ett så kraftfullt och applicerbart verktyg som AI hamnar fokuset lätt på möjligheterna och nyttan som kan skapas, men samtidigt diskuteras etiska dilemman kopplade till användandet av AI. Huruvida exempelvis en självkörande bil eller en AI-styrd domare kan agera i linje med mänsklig etik i kritiska situationer, samt partiskhet av AI mot människor, är frågor som debatteras och forskas på (Unesco, n.d.). Vilka andra komplikationer som kan dyka upp till följd av AI diskuteras också, däribland frågor som AI:ns hot mot arbetsmarknaden och ansvarsfördelningen för AI:ns agerande (Guillot, 2021a).

Partiskhet kan vara något positivt i samhället men är oftast uppfattat som negativt och problematiskt (Danks & London, 2017). Partiskhet är inte ett nytt fenomen som kommit med AI utan är något som kommer från mänskliga värderingar och är i stor utsträckning inetsat i samhället sedan långt tillbaka. Redan 1988 fällde en brittisk kommission ett brittiskt lärosäte för diskriminering då AI-systemet som användes för bestämmande av vilka sökande som skulle bli kallade till intervju var automatiskt partiskt emot kvinnor samt personer med utom-europeiska namn. Nämnvärt i detta fall var att trots att denna diskriminering identifierandes såg man att i jämförelse med övriga liknande lärosäte var det automatiska ansökningsprogrammet ändå mindre partiskt än ansökningsprocesser administrerade av människor (Silberg & Manyika, 2019).

För att se vad människor hade tagit för beslut samt hur de hade agerat i situationen AI-systemet befinner sig i, har tankeexperiment samt verktygsprogram utvecklats. Philippa Foot presenterade The Trolley Problem med syftet att se vilken handling som anses mest moralisk. Tankeexperimentet bestod av ett scenario där ett tåg oundvikligen kommer att köra på en eller flera personer, beroende på förarens val. Sedan dess har The Trolley Problem utvecklats och tagit flera former som man kan ta del av idag (Thomson, 1985).

1.2 Problemområde

"How will machines know what we value if we don't know ourselves?"

John Havens (2019) lyfter med denna fråga problematiken bakom att programmera maskiner och system utifrån mänsklig etik, när den mänskliga etiken i sig är svårdefinierad.

Etiska dilemman såsom The Trolley Problem används retoriskt för att exemplifiera problematiken i att ta beslut i en situation, och därmed bli involverad i ansvaret för utfallet av situationen. För att bygga vidare på denna teori och undersöka människors beslutsfattande i etiska dilemman, har experimentet Moral Machine konstruerats. Detta program, som samlar in data enligt enkätmetodik, har fokus på hur respondenterna tycker att en självkörande bil borde agera i olika etiska dilemman baserat på partiskhet (Awad, Dsouza & Chang, n.d.). Medan detta ger en bra grund för forskningen på självkörande bilars etiska agerande, menar vi att forskningen kan utökas till fler samhälleliga områden som AI är verksamt i samt till att inkludera andra aspekter såsom respondenternas demografiska tillhörighet.

Eftersom litteratur för enkätmetodik lyfter fram vikten av att samla in demografdata för sin undersökning (Eljertsson, 2019; Trost & Hultåker, 2016), har vi identifierat en brist i Moral Machine då undersökningen inte ställer in respondenternas fulla demografiska tillhörigheter, utan enbart vilket land de befinner sig i. Detta leder till att demograferna ålder och kön, som annars skulle kunna leda till värdefulla data, uteblir och undersökningens resultat enbart analyseras utifrån geografisk plats.

Baserat på tidigare litteratur och aktuella tillståndet hos etiska mätningsverktyg vill vi därmed lägga ett fokus att undersöka demograferna kön och ålders eventuella relevans vid undersökning av etiska beslutstagande. Detta då vi uppfattar det som att data uteblir och resultat från Moral Machine inte blir rättvisande eller kan nyttjas till sin fulla potential.

1.3 Forskningsfråga

Frågeställningen baseras som beskrivet i problemområdet på en utbyggnad av Moral Machine, med hänvisning till teorin att etiska beslut har kopplingar till demografisk tillhörighet. Med empirisk undersökning vill följande forskningsfråga besvaras:

- Hur skiljer sig åsikter mellan de demografiska egenskaperna ålder och kön angående frågan om partiskhet inom AI?

1.4 Syfte

Syftet med studien är att undersöka en utveckling av redan existerande verktyg för att mäta etiska beslutstagande. En avsaknad av krav på demografer hos redan aktuella verktyget Moral Machine har identifierats och undersökningen vill se om detta är rättfärdigat eller det resulterar i en avsaknad av väsentliga data. Studien vill lyfta frågan för vidare forskning på relevansen i att inkludera demografer som kön och ålder vid vidare utveckling och användande av verktyg för etiska beslutstagande.

1.5 Avgränsningar

Avgränsningar görs i denna studie. Den empiriska undersökning som kommer att genomföras är skriven på svenska och vänder sig därmed främst till svenskar. I en teknisk avgränsning kommer endast nödvändig teknisk beskrivning anföras i form av begreppsdefinitionen. Övriga demografiska avgränsningar kring ålder samt kön görs i överensstämmelse med forsknings-frågan.

2 Litteraturgenomgång

Litteraturgenomgången presenterar relevanta begrepp och teorier inom området AI och etik. Teorierna är en genomgång av det ramverk som sedan används för den empiriska undersökningen.

2.1 Demografiska skillnader i etik

Vid insamling av empiriska data tas respondenternas demograf ofta upp för att både belysa eventuella skillnader och för att kunna bearbeta resultaten därefter. Eljertsson menar att det ofta finns skillnader mellan demograferna ålder och kön, och att det är lämpligt att ta hänsyn till dessa om frågan är relevant för ämnet (Eljertsson, 2019). Trost och Hultåker (2016) stärker även att både variablerna kön och ålder skiljer sig sinsemellan och är därför relevanta att inkludera i empiriska undersökningar.

Utöver denna metodlitteratur stöds även teorin att demografer som ålder och kön har en påverkan på individers moral av Arutyunova, Alexandrov & Hauser (2016) som redogör för variationerna. Författarna som undersökt skillnader i moraliska beslut mellan kulturer, kön och ålder konstaterar att resultaten varierar till en grad där det är möjligt att se mönster av skillnader mellan grupperna (Arutyunova et al., 2016).

Vilken påverkan kön som demografisk variabel har på etik och moral beror på vilken typ av fråga som ställs. Det har identifierats att kvinnor tenderar att ha en mer strikt etisk standard än män, alltså att de tenderar att vara mer konsekventa med sina åsikter vilket i sin tur leder till att värderingar i olika frågor och scenarion påverkas av principen att vara konsekvent (Nikoomaram, Roodposhti, Ashlagh, Lofti & Taghipourian, 2013). Män har även identifierats ha en starkare koppling till utilitaristiska, mer praktiskt resonerade beslut, exempelvis i det etiska dilemmat att offra en individ för att rädda flera (Arutyunova et al., 2016),(Banerjee, Huebner & Hauser, 2010).

Ålder som demograf är en variabel som anses ha en implikation på det slutgiltiga empiriska resultatet vid mätning av etiska värderingar. Forskning visar att denna demograf har en koppling mellan ålder och styrkan i de etiska värderingarna, med en nedåtgående trend i takt med att man blir äldre (Arutyunova et al., 2016). Det förekommer studier som inte identifierat detta samband, och Nikoomaram et al., (2013) argumenterar att det finns moment i deras forskning som pekar åt att det inte finns betydliga samband. Dock dras slutsatsen från en helhet av forskningen, vilken menar att skillnaderna är betydliga nog för att kunna säga att skillnaderna existerar (Nikoomaram et al., 2013).

2.2 Artificiell intelligens som begrepp

Att definiera intelligens i sammanhanget av AI är en vetenskapligt omdiskuterad fråga som ännu inte blivit fullt besvarad (Wang, 2019). AI kan ses som en maskinell enskild entitet med intelligenta förmågor, eller som en hel bransch för forskning på intelligenta förmågor i maskiner (Merriam-Webster, n.d.-a), (McCarthy, 2007). Ordet intelligens i sig förtjänar förtydligande och taget utanför sammanhanget av AI, definieras som en förmåga att lära, förstå och applicera kunskap (Merriam-Webster, n.d.-b). Dessa definitioner ger en viss grund för avgränsning i hur ordet får användas, men räcker inte som förklaring i sammanhanget då det finns olika typer av intelligenskapacitet. Det är därför viktigt att klargöra för hur varje uppsats kommer att använda begreppet (Wang, 2019). Eftersom denna uppsats syftar på att undersöka fältet i sin helhet och AI som koncept snarare än en specifik typ, kommer begreppet här att syfta på det samhälleliga användandet av maskinella intelligenta förmågor.

2.3 Etiska dilemman inom AI

Ett etiskt dilemma är just ett dilemma för att det inte på något enkelt sätt går att lösa. För fältet av AI syftar etiska dilemman på hur systemen ska programmeras för att följa en viss etik som fastställs av mänskliga värderingar. Hur man än väljer att designa systemets etik kommer det att uppstå situationer där beslut måste tas som kommer att gå emot somligas uppfattning av etik, men även gå i linje med andras (Kompella, 2020).

Före sin tid tog författaren Isaac Asimov fram tre lagar om robotiken, med en nollte lag tillagd i efterhand, för syftet av science-fiction. Dessa lagar har blivit omdiskuterade och vidare använda i både science-fiction såväl som i akademiska sammanhang. Vidare påbyggnad på lagarna menar att det finns utrymme för en femte och sjätte lag för att komplettera för de etiska beslut autonoma system behöver ställas inför (Nagler, Van den Hoven & Helbing, 2018). De ursprungliga lagarna från Asimov lyder som följande:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.
0. A robot may not harm humanity or, by inaction, allow humanity to come to harm.

Eftersom ingen av dessa lagar berör de etiska dilemman som uppstår när en AI i kritiska situationer blir tvungen att strida mot den första regeln, påpekar Nagler et al., (2018) behovet av nya tillägg som följande:

5. Humanity and robots must do everything possible to reduce the occurrence of ethical dilemma situations.
6. If the application of rules 1-4 leads to ethical dilemmas, which could not be avoided as required by rule 5, decisions should be randomized, giving each person the same weight.

För den föreslagna sjätte regeln argumenteras att det går i linje med principen om jämlikhet, men om det är rätt sätt att hantera etiska dilemman för AI råder det skilda meningar om. Att lösa de etiska besluten genom att inte ta några beslut alls för syftet att undvika partiskhet är inte tillräckligt, då all partiskhet inte nödvändigtvis är dålig eller etiskt fel (Danks & London, 2017). Denna synvinkel tas även upp och problematiseras i filmen ”I, Robot” (Proyas, 2004), som också inspirerats av lagarna från Asimov. I filmen visas en AI-styrd robot i ett kritiskt dilemma att avväga vilket människoliv att rädda mellan en vuxen man och en tolvårig flicka på väg att drunkna. Filmen syftar på att påpeka vikten av att inkludera mänsklig etik i AI:ns beslutsfattande, då AI:ns beräkning i detta scenario rent logiskt resulterar i att den vuxna mannen hade större chans att klara sig och därför är den som blir räddad. Problemet syftar på att framhäva att en människa normalt sett hade ignorerat detta och räddad den unga flickan, eftersom detta menas är den mer etiskt korrekta partiskheten.

2.3.1 Partiskhet

För vidare definition av partiskhet kan det förklaras som att ha en vinkling eller preferens för det ena över det andra. I kontexten av AI brukar det användas för att beskriva diskriminering utifrån algoritmiskt inlärd partiskhet, men ofta är partiskhet också nödvändigt för att ge kontext till data (Ferrer, van Nuenen, Such, Coté & Criado, 2020) Vidare är det värt att understryka att partiskhet är mänskligt och subjektivt och det som vissa menar är positiv partiskhet som stärker etiska principer, menar andra är negativ partiskhet som diskriminerar (Danks & London, 2017).

Den negativa sidan av partiskhet, som ofta leder till diskriminering, är det som vill motverkas och ofta är riktat mot individer eller grupper baserat på deras yttre eller inre egenskaper. En högst aktuell form av negativ partiskhet är diskriminering och preferenstagande utifrån personliga och fysiska attribut så som kön, etnicitet, funktionshinder, sexuell läggning och ekonomisk status. Sådan partiskhet är påvisad i människor och på ett samhälleligt plan då exempelvis domares beslut omedvetet kan påverkas av deras egna personliga egenskaper och åsikter, samt att rekryterare tar partiska beslut när de ser till identiska meriter men olika etniska bakgrunder. Människor saknar tidvis förmågan och självinsikten att definiera faktorerna som var avgörande vid ett visst beslut och att ett beslut kan vara influerad av en omedveten partiskhet (Silberg & Manyika, 2019).

Partiskhet i AI kommer från dess inlärning eller användning men är på något plan kopplat till det mänskliga medverkandet i processen. Partiskheten är algoritmisk och kan förklaras genom följande uppdelning i tre kategorier med tillhörande subkategorier (Ferrer et al., 2020; Danks & London, 2017).

1. Ett sätt AI kan komma att bli partiskhet är ofta ifrån träningsdata som benämns *träningspartiskhet* och ifall träningsdata är partisk i sig kommer agerandet och beslutstagande därefter även vara partiskt.
2. Två typer av orsaker till partiskhet förekommer inom orsaken *partiskhet genom modellering. Algoritmisk processpartiskhet* - kan introduceras när medvetna åtgärder genomförs i form av parametrar för att lindra partiskhet i träningsdata. *Algoritmisk fokuspartiskhet* – där man försöker framställa subjektiva bedömningar utifrån objektiva data, exempelvis när slutsatser dras baserat på data utan hela kontexten i åtanke.
3. Två typer av orsaker bakom partiskhet; *Sammanhangspartiskhet* – innebär att algoritmer som är skapade med att agera utifrån ett visst sammanhang eller en viss population kan orsaka partiskhet ifall den ställs inför en annan population än den tilltänkta.

Tolkningspartiskhet - när en misstolkning av ett algoritmiskt resultat leder till partiska handling.

Det finns två perspektiv man kan se på partiskhet inom AI. Det första är; där mänsklig omedveten eller medveten partiskhet förekommer ska AI tillämpas och kunna bidra med att minska den mänskliga subjektiva synen vid ett beslutsfattande. Detta fungerar då systemets algoritm för maskininlärning endast tar relevanta variabler med i beslutfattandet och därmed kan framställa ett korrekt beslut baserat på tidigare data. Bevis finns som visar på att detta ser till ett bättre och mer rättvist beslutsfattande. Man kan även efter ett beslut enklare examinera beslutet och varför det togs (Silberg & Manyika, 2019). Perspektivet styrks med tankesättet att AI inte endast ska ses som orsaken till potentiella problem utan även som möjligheten att kunna identifiera och göra förbättringar. Detta då AI kan hjälpa till med identifiering av partiskhet vilket ger upphov till idén att AI kan vara medveten om den. I detta fall kan då AI assistera med att behandla och reducera aktuell negativ partiskhet (Ferrer et al., 2020).

Det andra perspektiv ser till risker och anledningarna till att partiskhet inom AI förekommer. Forskning har sett en förekomst av partiskhet och flertal exempel finns. Detta påvisar hur AI i syfte till något positivt kan resultera i obemärkt, icke kontrollerad partiskhet (Silberg & Manyika, 2019). Att lindra existerande partiskhet är även till följd sett som en stor utmaning utifrån flertalet aspekter men inte omöjligt. Exempel på dessa är; att människan inte alltid identifierar sin egen partiskhet, det är svårt att se hur stor påverkan det har på ens beslutstagande, att minska varenda kognitiv partiskhet (ens subjektiva tolkning av verkligheten) kan komma till att vara stor utmaning då de existerar i så stort antal samt att lokalisera gömd partiskhet i en algoritms träningsdata anses som stor utmaning (Violago & Quevada, 2018).

2.4 Moral Machine

Inom beslutsfattande AI tas ofta The Trolley Problem upp för att exemplifiera problematiken i de beslut som en AI, i de flesta fallen självkörande bilar, behöver ta i olika kritiska situationer. Etiken för att ta sådana beslut behöver programmeras in redan från grunden, men kan inte antas vara helt representerad av programmerarna själva (Kompella, 2020). För att få en mer rättvisande bild av etiken som ska följas behöver den algoritmiska träningsdata vara opartisk genom ett rättvisande urval och baserad på ett större sammanhang (Danks & London, 2017). Mänskligheten behöver vara inkluderad i processen för att ge en nyanserad grund för träningsdata som ska användas (Kompella, 2020).

Moral Machine är ett datainsamlande online-experiment som argumenteras av skaparna Edmond Awad, Sohan Dsouza, Paiju Chang, Danny Tang, Iyad Rahwan, Jean-Francois Bonnefon och Azim Shariff för att kunna användas i syftet att fylla behovet att inkludera mänskligheten i processen för att generera träningsdata i etik. Experimentet testar användarens preferenser för olika personliga egenskaper såsom ålder, kön, social status och kroppsfysik i scenarion som utgår från att en självkörande bil måste ta ett beslut mellan två grupper av människor där en av grupperna kommer bli påkörd och den andra skonad (Awad, Dsouza, Shariff, Rahwan & Bonnefon, 2018).

Anledning till att Moral Machine uppstod var då svagheterna kring The Trolley Problem identifierades då det inte är förankrat till ett mer aktuellt scenario. Självkörande bilar var även ett

mer aktuellt faktum där en riktig skillnad på liv och död faktiskt kan göras. Självkörande bilar medför en så stor mängd fördelar så att de frågetecken som väl finns vill kunna klargöras och optimeras. Att därmed kunna inhämta stora mängder data från ett tidigare filosofiskt tankeexperiment var något givande som var värt att vidareutveckla (Awad, Dsouza, Bonnefon, Shariff & Rahwan, 2020a).

Moral Machine införskaffar en stor mängd data som leder till diskussion och forskningsslutsatser. Med den geografiska demografen som verktyget spårar har det lett till vidare forskning för personer som står bakom experimentet. År 2020 färdigställde utvecklare och skaparna av Moral Machine en undersökning vars fokus låg på att lokalisera och analysera differenser i responsen utifrån deras geografiska plats. Utifrån en stor mängd data som kunde hämtas framställdes ett resultat utifrån analyseringen som visade på att det fanns moraliska skillnader beroende på geografisk plats. Däremot medges faktumet att svaren som inkom från länder utanför västvärlden var betydligt färre än västvärlden själv vilket bör uppmärksammas och tas i hänsyn. Skillnaden som uppmärksammades till högst grad var att det generellt fanns en mindre tendens till acceptans att offra ett liv för att rädda flera utanför västvärlden. Detta menar författarna förmodligen beror på kognitiva olikheter mellan respondenter (Awad, Dsouza, Shariff, Rahwan & Bonnefon, 2020b).

När man ser på vad Moral Machine har nått för slutsatser sedan den först släpptes och gick viralt år 2016 så har den år 2020 uppnått 40 miljoner svar från 233 länder och territorium. Det innebär att det är den största datainhämtningen någonsin kring ämnet. De största generella skillnaderna i vad människor tenderar att svara är följande: att rädda människor framför djur, att rädda fler än färre människor samt att rädda yngre framför äldre. Detta är en generaliserad syn och man kan som tidigare nämnts se tendensskillnader och skillnader i data geografiska platser emellan (Awad et al., 2020a). Resultatet visar även i den globala sammanställningen att det finns preferenser för att skona gravida, kvinnor, läkare, vältränade och yrkesmässigt högt uppsatta personer gentemot deras demografiska motsatser. Faktorer som att hålla kursen mot en grupp eller väja mot den andra gruppen beräknas också, och visar en mindre tendens till att hålla kursen över att väja. Vidare visar resultaten att det finns en viss tendens att skona de som i scenariot går lagligt på övergångsställen över de som har gått mot rött (Awad et al., 2018).

Forskarna förklarar även i vilken mån en sådan datainhämtning från människor ska kunna tillämpas och bidra till något givande vid utvecklingen av självkörande bilar. Ett ramverk har därför framställs där tanken är att kunna placera verktyget Moral Machine och dess data som en del av databasen utvecklare inom branschen för självkörande bilar använder sig av. Skaparna av Moral Machine understryker hur undersökningar som dessa ger AI sektorn en nytta genom att förstärka och inkludera människans makt och röst i AI utvecklingen (Awad et al., 2020a). Forskarna stärker studien med den stora mängd data som hämtats men medger även de stora brister som finns, vilket är att urvalet anses vara väldigt snävt i hänsyn till demografer då de inte inkluderar ålder, kön och utbildning (Awad et al., 2020b).

2.5 AI och partiskhet i samhället

Detta avsnitt presenterar områden i samhället var partiskhet förekommer och AI är verksam, med hänvisning litteratur och journalistik, som en grund för vilka områden som kan inkluderas i den empiriska undersökningen.

2.5.1 Självkörande fordon

AI som framför fordon i trafik med andra fordon har ett ansvar för att kunna göra detta på ett säkert sätt för att undvika faror. Korrekt körning enligt trafiklagar är dock inte alltid det mest säkra, då mänskligt beteende är en lika stor faktor i trafiken. Träning av autonoma fordon inkluderar således att lära sig förstå och imitera mänskligt beteende till viss grad (Li, Ota & Dong, 2018). Denna träning ges av en specifik bas av data, ofta framtagen av företaget som utvecklar AI:n eller hämtad från andra redan tillgängliga baser. Redan från grunden ligger en problematik i partiskhet från den träningsdata som används, då denna kan vara anpassad för en specifik, geografisk del av världen. Vad som anses vara normala trafiksituationer där AI:n utvecklas eller där data hämtas från, är inte naturligtvis en normal trafiksituation där fordonet sedan kommer att framföras. På samma vis kan även somliga trafiknormer som den lokala befolkningen är vana vid och respekterar som kultur, på en annan geografisk plats tolkas som direkt farliga (Danks & London, 2017). Skillnader i kultur kan variera kraftigt, som attityder kring alkohol i trafiken eller till vilken grad fortkörning anses vara okej. Exempelvis i två skilda kulturer som den i Sverige respektive Sydafrika finns skillnader i en högre tolerans för alkohol i trafiken och lättsammare inställning till fortkörning i Sydafrika (Sinclair, 2013). Ett annat specifikt fenomen som beskrivs av Sinclair (2013) är en ny norm i Sydafrika som menar att rödljus är mer av en riktlinje under natten.

Vidare diskussion kring partiskhet i autonoma fordon handlar även om etiska beslut som tas i kritiska situationer, exempelvis hur en AI ska förväntas kunna avväga mellan två människoliv. Huruvida den populära åsikten bland den generella befolkningen överensstämmer med detta problem utreds i The Moral Machine experiment (Awad et al., 2018).

2.5.2 Anställning

Användning av AI vid intervjuprocessen leder till stor effektivisering av anställningstiden men för även med nackdelar. Nackdelar kan vara partiskhet gällande exempelvis kön och etnisk tillhörighet, ofta på grund av den specifika träningsdata som system agerar utifrån (Fernández & Fernández, 2019). Ett exempel på detta kommer från ett verktyg som utvecklats inom Amazon för att automatisera anställning med en AI som granskar CV:n. Verktyget visade sig ha en partiskhet för att föredra män över kvinnor, på grund av den träningsdata som hade getts (BBC, 2018).

2.5.3 Försäkring

Försäkringssektorn är på gränsen till ett teknologiskt skifte och en marknad som i sitt nuvarande tillstånd baseras på att upptäcka och återställa skador och fel kommer att gå mer mot att förutspå och förhindra (Balasubramanian, Libarikian & McElhaney, 2021). AI bidrar med en ersättning och effektivisering av tidigare manuella uppgifter till att bli snabbare samt att färre

misstag sker. Forskare förutspår att till år 2025 kommer man potentiellt se en automatisering upp mot 25% av försäkringssektorn tack vare AI (Johansson & Vogelgesang, 2016). Denna upptrappning av sektorns användande av AI medför även, med logiken att AI standardiserar användarnas partiskhet (Silberg & Manyika, 2019), att sektorns partiskhet trappas upp i samma takt. Sektorn är idag i exempelvis USA redan partisk på ett vis som negativt påverkar minoriteter som får betala mer för att de bor i ett område som är överrepresenterat av en minoritetsgrupp (Luthi, 2020).

2.5.4 Sjukvård

AI-baserade system i olika sektorer står inför liknande utmaningar gällande partiskhet, och diskussionerna kring ämnet handlar till stor del om diskriminering. Inom sjukvård används AI för att identifiera sjukdomar och diagnostisera patienter. Utan en bred bas av träningsdata hämtad från olika demografer såsom kön eller etnicitet, kommer en AI att fungera sämre för sitt syfte för den gruppen av människor som är minst representerad i dess träningsdata. En sådan partiskhet missgynnar de som för sjukdomen är underrepresenterade eller som avviker från den överrepresenterade demografen i området där databasen framtogs (Kaushal, Altman & Langlotz, 2020).

Under extraordinära förhållanden som bristande resurser är nationella principer för prioritering inom intensivvård framtaget av Socialstyrelsen i Sverige. Som en fundamental regel framställs tydligt att innan någon prioritering görs ska bör alla möjligheter kring resursbehov uttömmas. Det återfinns sedan en prioriteringslista som är utformad utifrån faktorer som väger in. Listan är uppdelad i tre olika prioriteringar som grundar sig utifrån patientens förväntade överlevnadstid. Som ett tillägg till listan förklaras även att fortlöpande utvärderingar av patients tillstånd görs för att kunna omprioritera vid behov (Socialstyrelsen, 2020).

2.5.5 Rättsväsende

Amerikanska rättsväsendet använder sig av en prediktiv algoritm i hopp om att uppnå resultat av domar mot potentiella brottslingar med mindre partiskhet. En mjukvara vid namn COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) används med ändamålet att bidra med assistans till domaren för att avgöra ifall en person ska stanna i fängelset eller släppas fri. COMPAS baserar sin uträkning och framställda rekommendation utifrån mängder av historiska träningsdata inom rättsväsendet där den lokaliserar korrelationer mellan olika relevanta faktorer. Detta ska resultera i en riskbedömning som sedan kan användas som rekommendation till domaren som tar det slutgiltiga beslutet. Faktorer som inte ska få tas med i bedömning av risk för brottsligt återfall är hudfärg, dock påstås detta ske (Hao & Stray, 2019).

7000 riskbedömningar från Broward County, Florida analyserades och en väldigt låg nivå av korrekt bedömning av COMPAS kunde identifieras där endast 20% av de som förutspåddes utföra kriminella handling verkligen gjorde det. Vad som var mer anmärkningsvärt är den identifierade AI-partiskheten. Detta då afro-amerikaner blev missbedömda som ”hög-risk” att återfalla i brottslighet på en betydande högre grad än vita personer (Angwin, Larson, Mattu & Kirchner, 2016).

2.5.6 Ansvar

AI-systemen som finns idag kan agera och ta beslut helt utan mänsklig interaktion. Med hänsyn till konceptet maskininlärning följer system inte en specifik standard utan de kan anpassas och förändras med tiden. Med AI finns det däremot inte alltid någon med tillräckligt tydlig kontroll för att ansvaret rättvist ska kunna läggas på någon. Detta kan benämnas som *responsibility gap.* Parter som programmerare, observatör, företag, politiker eller systemet själv kan alla vara potentiella huvudsakliga ansvarstagare (Matthias, 2004). Till följd försöker samhället till detta faktum anpassa sig, med syftet att undvika dilemman över vem som hålls ansvarig att uppstå, detta kan till exempel utforma sig i principiella riktlinjer eller ett sammanfattande regelverk angående AI som EU framställt (EuropeiskaKommissionen, 2018).

3 Metod

3.1 Val av metod

För att givande resultat skulle kunna framställas inom forskningsområdet i slutet av denna studie, så är insamling av människans åsikt kring ämnet viktigt. I samband med att en grundad teoribildning kring ämnet lärts in krävdes det en väl utformad empirisk undersökning för att slutresultatet skulle kunna bli relevant för frågeställningen och vidare forskning. Här fanns det främst två vägar att ta i form av vilken typ av undersökning som skulle göras. I enlighet med Jacobsen & Andersson (2017) valdes den typ av metod som ansågs lämpligast och givande för studiens testande problemställning angående demografers påverkan på åsikter om partiskhet inom AI. (Jacobsen & Andersson, 2017). Därmed valdes en kvantitativ undersökning i form av digital enkät.

För att forskningsfrågan i denna studie skulle kunna besvaras genomfördes en kvantitativ undersökning. Anledningen grundade sig först och främst i forskningsfrågan där vi ville kunna lokalisera data för specifika utvalda demografer kön och ålder inom undersökningen. Kopplat till det var även en anledning till valet av denna typ av metod, då ett stort mål med den empiriska undersökningen var att tillhandahålla givande data ifrån en vid mängd människor med olika demografer. Därmed fanns också målet med en större mängd inkomna svar att en mer grundad slutsats kunde göras. Att genomföra den kvantitativa delen i form av enkät nådde ut till en större mängd respondenter på ett snabbare tillvägagångsätt som sedan kunde analyseras. Detta kunde sedan bidra med framställning av slutsats (Oates, 2006).

3.2 Insamling av data

3.2.1 Enkätundersökning – planering och genomförande

Vid skapandet av enkäten har planeringen och utförandet utgått ifrån sex subkategorier som är följande; datakrav, datagenerationsmetod, urvalsramen, urvalsteknik, urvalsstorlek, svars- och icke-svarsfrekvens (Oates, 2006). Nedan beskrivs varje subkategori vid planering och genomförandet:

Datakrav: De demografiska egenskaper som ansågs relevanta för denna studie var endast ålder samt kön i enlighet med frågeställningen. Andra egenskaper som yrke, fritidsaktiviteter och andra personliga egenskaper ansågs överflödiga för att den data som skulle utvinnas av enkäten skulle vara relevant för frågeställningen. Den generella data som ville utvinnas av enkäten är en bild över folkets etiska värderingar i situationer där AI kan vara inkluderad i samhället, utifrån respondenters ålder och kön.

Datagenerationsmetod: Enkäten var formulerad i form av ett frågeformulär men som är skrivet och baserat på scenarion. Det underströks väldigt tydligt i introduktionen till enkäten att

enkäten var anonym så respondenter skulle svara ärligt i varje scenario och på de frågor som ställdes. Frågorna var utformade så att de även skulle kunna vara grund till intervjufrågor men detta ansågs inte vara nödvändigt med hänvisning att endast kvantitativ metod var vald. Svaren som inväntades kom därmed inte heller att kräva någon förkunskap inom AI sektorn och det tekniska området utan det var ett frågeformulär som in princip alla kunde ge respons på. Svar som differentierade stort från andra var inte heller därför något som exkluderades då varje enskildas åsikt om scenariona och frågor skulle respekteras och tas i hänsyn.

Urvalsramen: Denna undersökning var ute efter svar som inte nödvändigtvis vände sig mot en specifik grupp av människor. Kunskap inom tekniska området var inte en nödvändighet utan det som behövde förstås för att besvara frågan stod utförligt beskrivet innan varje del av formuläret. Därmed kunde enkäten spridas via sociala medier och vilka det var som besvarade var inte av största vikt förutom att vi satte en åldersgräns på att respondenterna skulle vara minst 18 år gamla. Med frågeställningen i åtanke ville vi se en jämn spridning av demograferna som besvarade enkäten. Svar från den allra äldsta gruppen människor var inte att förväntas då de på ett generellt plan inte tenderade att använda sig av sociala medier på samma sätt som yngre.

Urvalsteknik: Valet av urvalsteknik var inte något av högsta prioritet då enkäten hade ett öppet urval. Däremot skrevs en beskrivning av innebörden av enkäten vid publikation vilket troligen skulle leda till att respondenter som besvarade gjorde det av självintresse. Trots detta hade vi kommit till att använda sig av ett bekvämlighetsurval för att på ett så enkelt och så effektivt sätt som möjligt tillhandahålla mycket data (Trost & Hultåker, 2016). En svaghet med denna metod var att man inte kan ha kontroll på vilka grupper man missar, och det är därför inte möjligt att generalisera resultaten på ett sätt som kan accepteras som fakta (Jacobsen & Andersson, 2017). Denna urvalsmetod gjordes då den ansågs passande för få grundande data till en studie av detta slag.

Urvalsstorlek: För att utvinna ett så givande och grundande empiriskt underlag till en studie krävdes det att enkäten medförde en stor mängd svar. Ju fler svar som kunde utvinnas, desto bättre. I målet mot att kunna bidra med att lyfta demografiska aspekten hos The Moral Machine som annars fattas, krävdes så mycket respons som möjligt.

Svarsfrekvens: Vid utlämning av enkäten tilldelades den till potentiellt flera hundra möjliga respondenter. För att göra enkäten mer attraktiv att besvara var enkäten rubricerad ”Testa din partiskhet” samt en kort förklaring vad den handlade om samt att den var rolig och tänkvärd att besvara. Sedan spreds den även med indikationen att respondenterna gärna fick sprida vidare enkäten för att ännu fler svar ska utvinnas.

3.2.2 Kvantitativ undersökning

En större undersökning genomfördes med en internetbaserad enkät som delades av författarna med våra sociala kretsar med förhoppningen att nå ut till fler individer. Enkäten presenterade scenarion från dels självkörande bilar, AI-baserade beslut och andra fall av AI-partiskhet. Svarsalternativen gav den svarande möjlighet att välja ett av två eller flera utfall av scenariot utifrån deras personliga etik. I fallet då den svarande inte kunde eller ville besluta, hade vissa scenario svarsalternativet att låta den hypotetiska slumpen avgöra utfallet hur den självkörande bilen skulle agera. Eftersom det för undersökningen var relevant att kartlägga människors etiska beslut i scenarion, var det lika relevant för undersökningen att kartlägga när människor upplevde att ett sådant beslut inte var möjligt att ta i specifika situationer.

Enkäten byggde på att undersöka respondenters etiska värderingar och ställningstagande i olika situationer där AI även kunde befinna sig i, utifrån respondenternas demografiska egenskaper ålder och kön. För att AI skulle förstå vad som anses etiskt korrekt i situationer behövde folket främst undersökas. Detta återspeglades i problemformuleringen och var en stor del av undersökningen inom AI-partiskhet. I enlighet med Oates (2006) analyserades sedan responsen med mål att finna mönster som kunde bidra till en sedan tagen slutsats.

För att på ett effektivt sätt som möjligt kunna analysera och identifiera mönster hos respondenternas svar användes hjälpmedel i form av kalkylbladsprogram. Att ta hjälp av mer komplicerade statistiska hjälpmedelsprogram ansåg vi inte vara nödvändigt och inte heller optimalt då vi ville förstå data och skapa korrekta uppfattningar kring det. Det fanns annars en risk att förståelsen för presenterade data uteblev vilket varken är bra för oss eller undersökningen i sin helhet (Oates, 2006).

3.2.3 Motivering av enkätstruktur

Enkäten inleddes med frågor angående respondentens demografiska egenskaper och var obligatorisk att besvara. Sedan ställdes respondenten inför olika scenarion som liknades mycket vid Moral Machine. Problematiken kring programmet var det som utreddes och därmed skulle frågorna som ställdes efterlikna programmet. Vidare fortsatte enkäten utforska respondentens åsikter angående andra scenarion där AI-partiskhet kan förekomma, samt slutligen utreddes var ansvaret ansågs ligga.

DEL I: Inledande delen av enkätundersökningen ställde in respondentens demograf utefter ålder och kön. Då frågorna i enkäten var av en något känslig karaktär och utgick från att respondenten hade en viss grad av världslig kunskap, var en minimumåldersgräns satt på 18 år. För att kunna svara på frågor om trafiksituationer och andra samhällssituationer bör respondenten vara i myndig ålder och ha haft tid att erfara dessa för att de skulle kunna ge mogna svar. Vidare var ålder och kön det som var av största intresse för undersökningen så att eventuella kopplingar kunde göras mellan demografer och svar.

DEL II: I denna del presenterades olika scenarion i trafiken och ställde frågan utifrån vad respondenten skulle rekommendera att en självkörande bil borde göra. Scenariona var inspirerade av Moral Machine (Awad et al., n.d.). I varje scenario fick respondenten välja mellan två utfall, med möjligheten i denna undersökning för ett tredje alternativ att inte ta ett beslut utan låta slumpen avgöra utfallet i scenariot, dvs. att inget aktivt val kunde eller ville göras. För undersökningens skull var det av lika stort intresse att se i vilka scenarion respondenterna tog beslut, som i vilka de valde att avstå och låta slumpen avgöra.

DEL III: Tredje delen presenterade scenarion från olika samhälleliga funktioner där beslutsfattande AI används. De inkluderade områdena sjukvård, försäkringar, jobbansökningar och rättsväsende valdes med hänsyn till identifierade användningsområden för AI samt vilka områden som var vardagliga nog att den genomsnittliga respondenten skulle kunna sätta sig in i scenariot. För dessa frågor var ställningstagandet inte mellan två utfall, utan mellan att hålla med om scenariots etiska grund eller inte. Syftet var att undersöka respondenternas ställningstaganden, men det var även av intresse att undersöka graden av enighet eller splittring i svaren.

DEL IV: Den avslutande delen i denna enkät ämnade undersöka respondenternas tycke kring ansvarsfördelning inom AI. Dessa scenarion var utformade för att återkoppla till föregående

delar där respondenten redan fått ta ställning, men med fokus på ett scenario som gått fel och med frågan riktad på vem som bar ansvaret. Respondenten fick uttrycka i flervalsalternativ vilka de ansåg var huvudsakligen ansvariga för händelsen inom varje scenario. För varje område identifierades olika involverade parter och svarsalternativen har således skiftat efter relevans till frågan. Enkäten avslutades med en fråga om ett ultimat ansvar för partiskhet inom AI, som en sammanfattning på undersökningens helhet, med svarsalternativ för olika involverade parter med makt över situationen.

3.3 Validitet och reliabilitet

Metodologiska överväganden var något man behövde ta i hänsyn gällande enkätundersökningen i fråga. Valet av design samt att det skulle vara frågor formulerade på ett lättförstått sätt hade en stor påverkan på enkätens validitet och reliabilitet. Att respondenter skulle ha en full förståelse för vad de svarade på och varför de svarade på det var något som var av stor vikt. Dels för att man skulle vara medveten om att alla svarade med samma uppfattning om frågan och därmed att svaren i sin helhet var relevanta. Skulle frågorna vara otydliga eller dåligt formulerade påverkades validiteten. Detta kunde leda till att respondenten inte förstod eller missförstod, vilket kunde medföra felaktiga data och lägre svarsfrekvens (Trost & Hultåker, 2016). Därmed hade enkätens varje del en beskrivning av kommande del för att man på ett så enkelt sätt som möjligt skulle förstå innebörden på det man skulle svara på. Respondenten skulle inte under enkätens gång behöva undra något. Reliabiliteten hanterades genom att undvika onödigt krångliga ord, satser eller andra missvisande meningsuppbyggnader, allt för att undvika att respondenter missförstod frågor vilket hade lett till en låg nivå av reliabilitet (Trost & Hultåker, 2016). Mycket av detta var saker vi som skapare av enkäten haft i åtanke och det var även den primära anledningen till en pilotversion användes.

3.3.1 Pilotversion

Vid färdigställning av enkäten tilldelades först en version ut till ett fåtal utvalda personer som fick besvara enkäten. Sedan kunde dessa personer ge feedback och tips på vad som kunde förbättras till när vi sedan publicerade enkäten till allmänheten. Anledningen till detta var då vi ville få andras uppfattning om enkäten, så att därmed alla frågor var korrekt formulerade och tillräckligt förståeliga för att därmed minska risken till missförstånd, som rekommenderat av Eljertsson (Eljertsson, 2019).

3.4 Etik

Vid denna empiriska undersökning var det viktigt att ta hänsyn till etik. Vi arbetade utefter följande tre krav vid genomförande av enkäten, vilket bör tillgodose detta (Jacobsen & Andersson, 2017);

1. Informerat samtycke.
   Den undersökta, i detta fall respondenten som gör enkäten, gör så på egen vilja och med förståelse för innebörden i deltagandet.

2. Rätten till privatliv.
   Undersökningen respekterar att respondenten har rätt till att inte uppge viss information som är privat, känslig eller kan användas för att identifiera enskilda individer.
3. Krav på korrekt presentation av data.
   Resultat analyseras i sitt fullständiga sammanhang så att respondenternas svar presenteras korrekt. Dataförfalskning och manipulation är inte tillåtet och för att visa detta bör forskningsprocessen och undersökningens svar presenteras med öppenhet.

I denna digitala enkätundersökning har respondenten behövt uppge privata värderingar och information som vissa kunde anse vara känslig. Då detta var det faktiska syftet med undersökningen gick det inte att undvika att respondenterna eventuellt reagerade känslosamt på frågorna. För att respondenten inte skulle pressas till att svara på känsliga frågor var introduktionen av enkäten tydlig med att det var okej att avbryta genom hela undersökningen. Frivillighet, anonymitet och enkätens syfte presenterades också tydligt i introduktionen. I enighet med kravet om Rätten till privatliv, samlades ingen data in som direkt kunde knytas till en specifik individ. Enkäten krävde inte heller någon inloggning eller registrering av e-postadress, så respondenten kunde därför känna sig trygg i att uppge ärliga svar med full anonymitet. Resultaten presenterades i sitt fulla sammanhang med undersökningens råa data tillgänglig för läsaren att granska.

3.5 Bearbetning av data

Innan analys och bearbetning kunde genomföras behövde bortfallet identifieras för att se hur det påverkade analyserat resultat (Trost & Hultåker, 2016). Något som tydligt identifierades var att respondenters ålder var i majoritet mellan 18–24. Resterande åldersspann stod för ännu färre delar av responsen vilket inte var optimalt. Detta omöjliggjorde möjligheten till analys inom denna demograf. Detta då det inte hade blivit ett tillförlitligt resultat (Jacobsen & Andersson, 2017). Trots forskningsfrågan inkluderade vi därför inte demografen ålder i analysen. Cirkeldiagrammen som presenterar den obearbetade data är dock framtagen utan hänsyn till någon demograf och är därför skev på grund av åldersfördelningen i undersökningen. Fördelning mellan kön blev däremot mer hanterbar och kunde analyseras till ett mer tillförlitligt resultat.

Då fördelningen av könstillhörighet landade i att ett av alternativen, annat, blev grovt underrepresenterat med bara en respondent, har denna demograf inte räknats med då en person inte räckte för att föra en talan. Respondentens svar har dock räknats med i frågorna som helhet, det var alltså bara analyser för demografen kön som inte kom att räkna med könstillhörighetsalternativet annat.

Då enkäten genomfördes i Google Forms, extraherade vi data till Google Sheets för att använda som primärt verktyg för bearbetning och för att skapa cirkel- och stapeldiagram. Dessa diagram krävde ingen bearbetning för att generera, men för att sedan utvinna information om frågornas demograffördelning och respondenters tendenser i svar använde vi formler för att räkna i programmet och ställa upp tabeller. Dessa tabeller var dels för att omformulera svarsfördelningen till procent av demografens totala deltagande för att ge ett tydligare resultat som visade skillnader mellan demografer mer rättvist. En annan tabell skapades för att kunna identifiera eventuella mönster i individsvar. Detta gjordes helt utan hänsyn till demograf och var enbart för att kunna användas som ett verktyg för att se om det fanns någon tendens bland respondenterna att hålla sig till ett av svarsalternativen av ren princip. Om ett resultat från denna

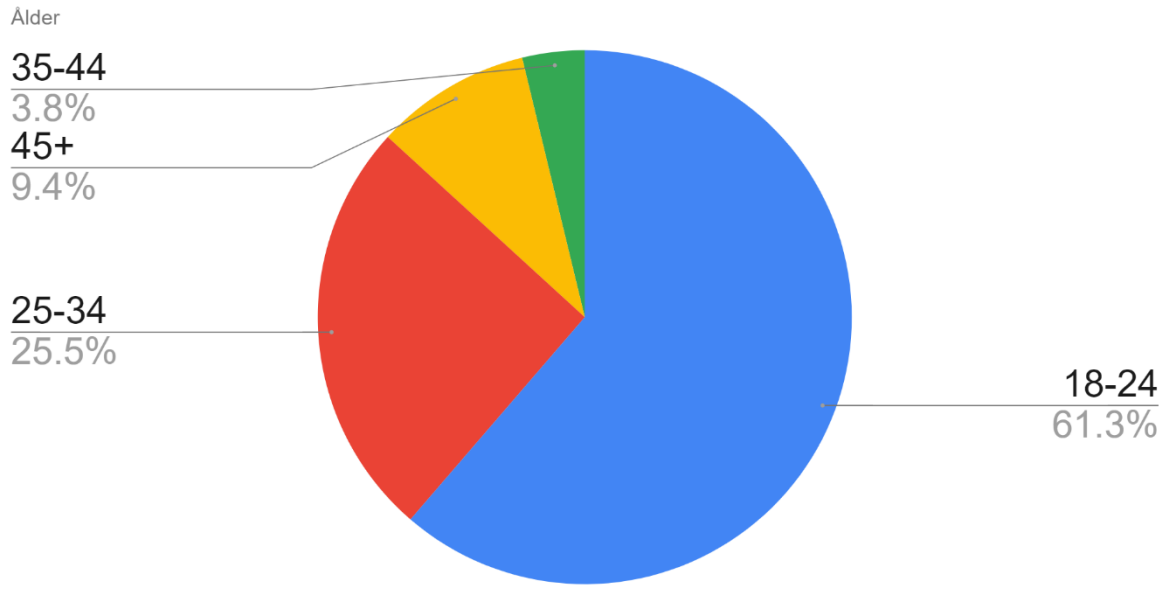
tabell pekade på tydliga tendenser kunde detta användas i diskussion för att delvis förklara de skillnader som hittades. Tabellens syfte var också att användas som stöd för att diskutera teorin om slumpmässigt utfall som standard och hur detta hade tagits emot enligt människors etik (Nagler et al., 2018).

4 Resultat

4.1 Resultat från enkätundersökning

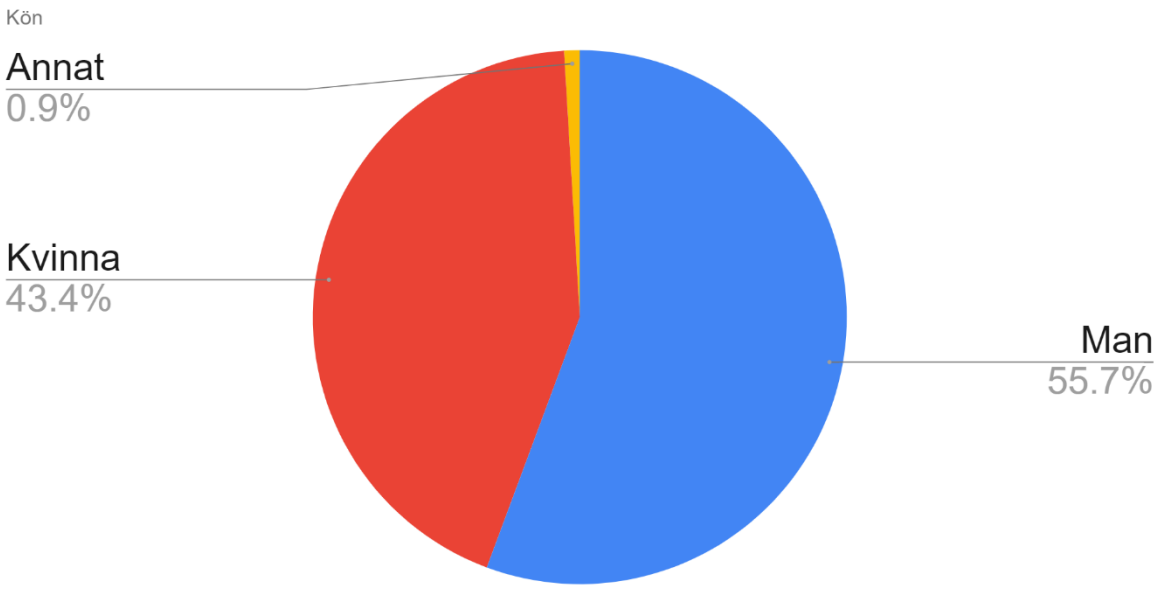
Enkätundersökningen innehöll totalt 15 frågor som var uppdelade i fyra delar. Totalt medverkade 106 respondenter i undersökningen och samtliga uppgav svar på alla 15 frågor. I detta kapitel presenteras de direkta resultaten med kortare beskrivningar för granskning av respondenternas eventuella skillnader i svar utefter demograftillhörighet. För att vara konsekventa i identifiering av betydliga skillnader i frågor noteras bara skillnader av tio procentenheter eller mer mellan demografmotsatser.

4.1.1 Del I: Respondentens demograf



Figur 1: Enkätfråga ålder.

Den dominerande åldersgruppen bland respondenterna var 18–24, med 61% (65). Nästa åldersgrupp var 25–34, med 26% (27). Åldersgruppen 35–44 är minst i storlek med 4% (4). Åldersgruppen 44+ står för 9% (10) av respondenterna.

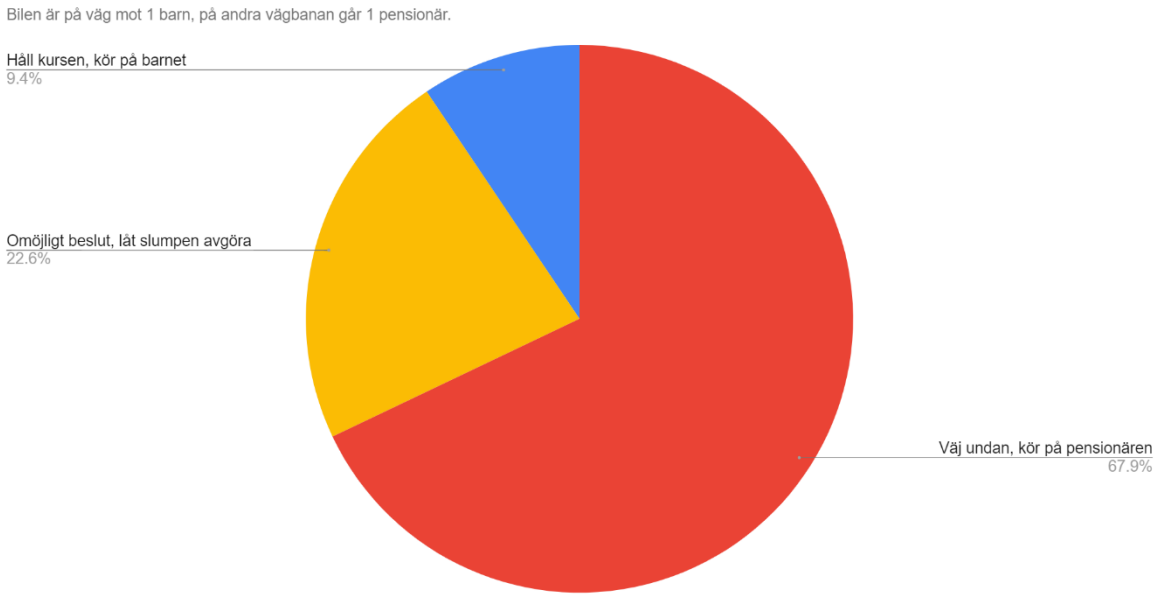


Figur 2: Enkätfråga kön.

Majoriteten av respondenterna var män med 56% (59). Kvinnor utgör 43% (46) och 1% (1) identifierar sig som annat.

4.1.2 Del II: Scenario – självkörande bil

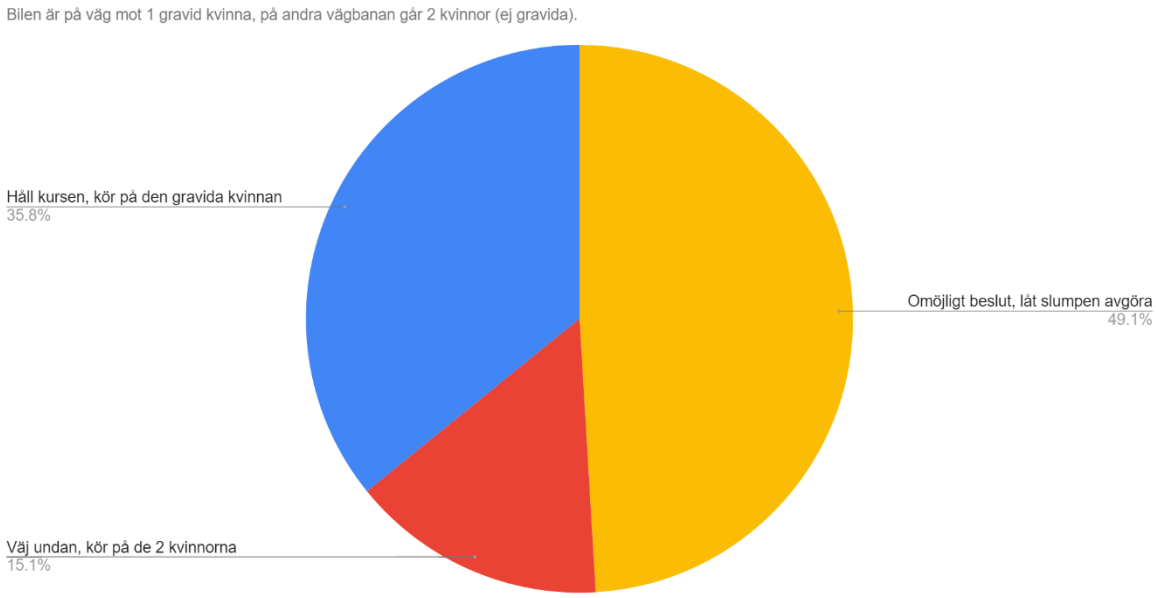
4.1.2.1 Scenario ett



Figur 3: Enkätfråga scenario ett.

Resultaten visade en majoritet med 68% (72) av svaren för svarsalternativet att väja undan och köra på pensionären. 23% (24) valde att det var ett omöjligt beslut, och 9% (10) svarade att hålla kursen och köra på barnet.

4.1.2.2 Scenario två

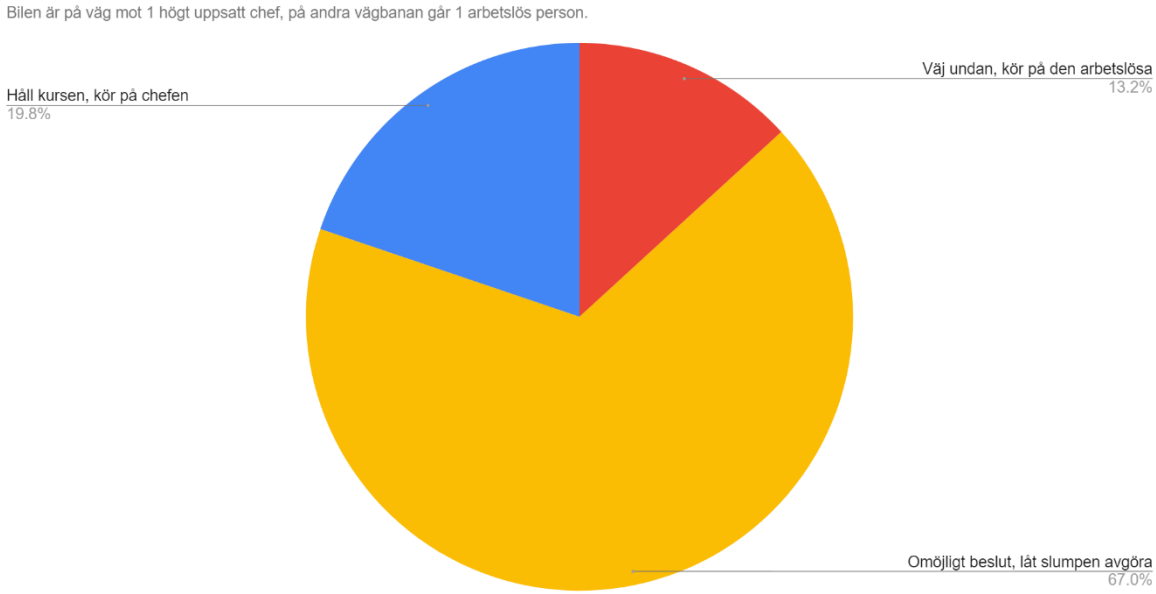


Figur 4: Enkätfråga scenario två.

Resultaten visade att i denna situation valde 49% (52) Omöjligt beslut, låt slumpen avgöra. 15% (16) valde alternativet att väja undan och 36% (38) valde att hålla kursen.

Demograffördelningen i denna fråga visade skillnad med 18 procentenheter mellan könen. 59% av kvinnorna valde omöjligt beslut medan 41% av männen valde samma.

4.1.2.3 Scenario tre

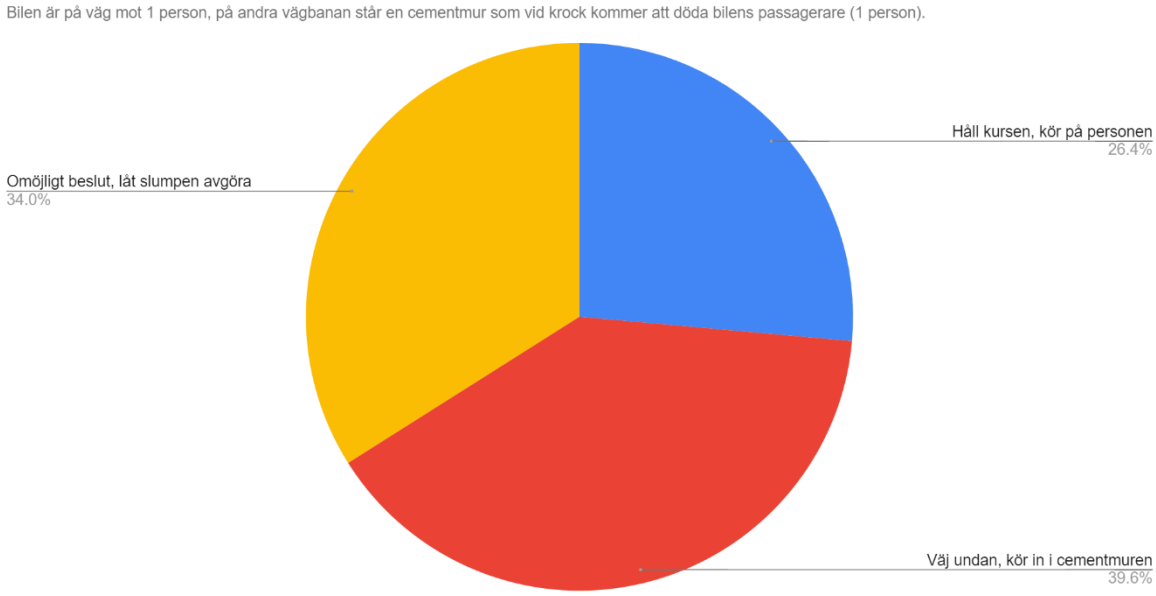


Figur 5: Enkätfråga scenario tre.

Resultaten visade att 67% (71) av respondenterna valde omöjligt beslut. 13% (14) valde att väja undan till andra vägbanan medan 20% (21) valde att hålla kursen på aktuell vägbana.

En skillnad ses mellan kön, då 74% av kvinnorna valde omöjligt beslut jämfört med 63% av männen. Båda grupperna följde dock mönstret från den totala svarsfördelningen med majoritet för omöjligt beslut.

4.1.2.4 Scenario fyra

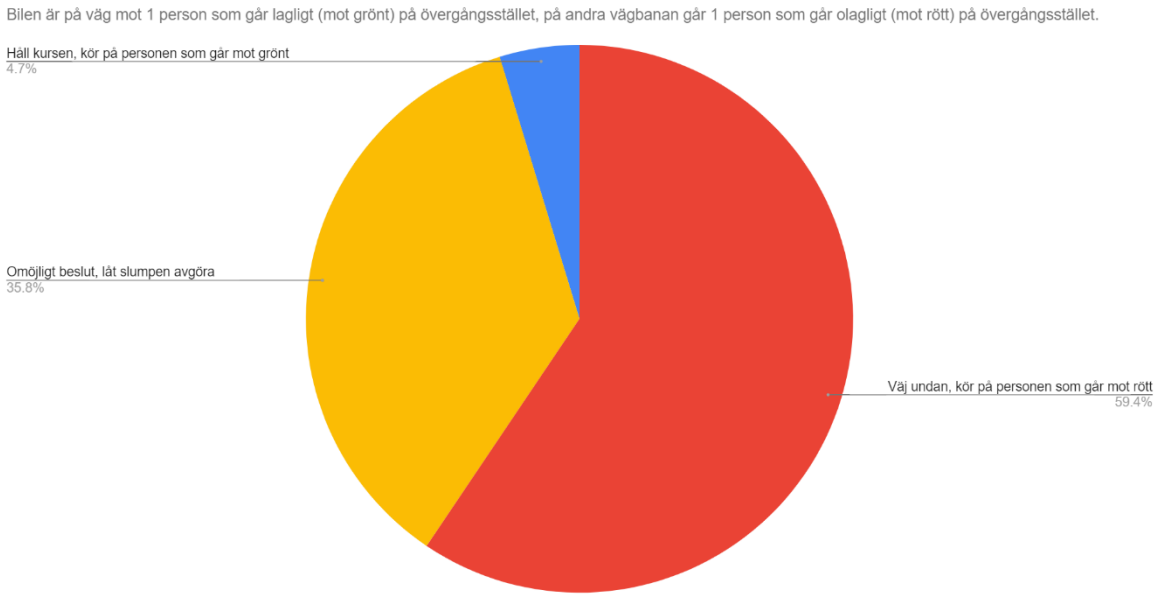


Figur 6: Enkätfråga scenario fyra.

Resultaten visade med relativt jämn fördelning att 40% (42) valde att väja undan. 26% (28) av respondenter valde att hålla kursen medan 34% (36) valde att beslutet är omöjligt.

Demograffördelningen hade en skillnad på 17 procentenheter mellan könen, var 34% av männen valde att hålla kursen och 17% av kvinnorna valde samma. Männen visade sig vara jämnt fördelade mellan svarsalternativen med 34% som svarade att väja undan jämfört med 46% av de kvinnliga respondenterna.

4.1.2.5 Scenario fem



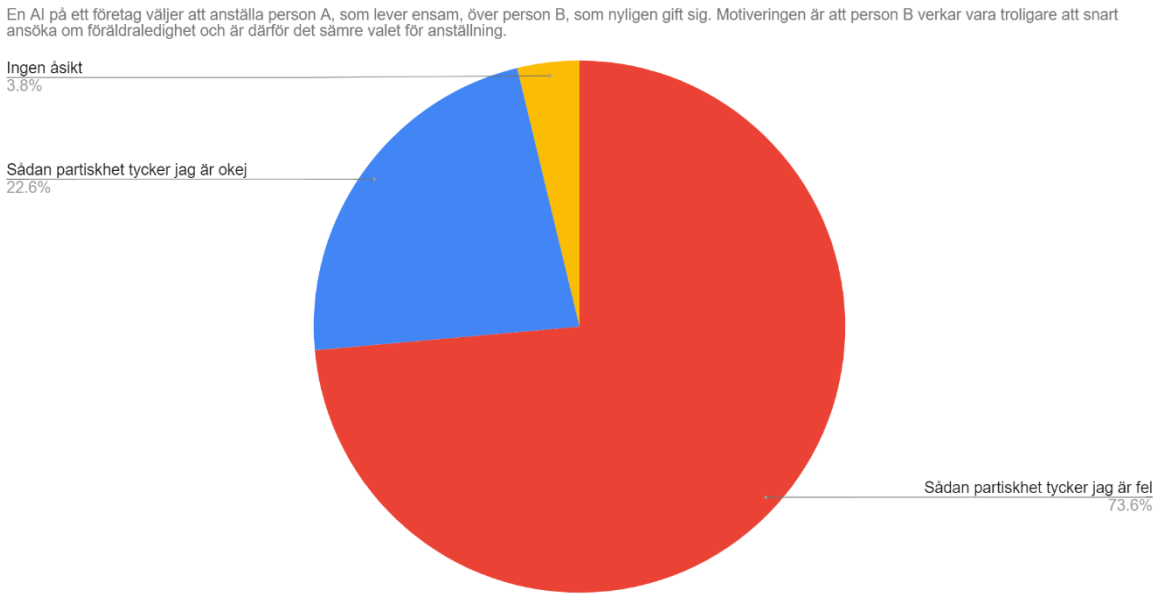
Figur 7: Enkätfråga scenario fem.

Resultaten visade att 59% (63) valde att väja undan. 36% (38) ansåg att beslutet var omöjligt att ta medan 5% (5) valde att hålla kursen på aktuella vägbanan.

Demograffördelningen pekade på skillnader mellan könsgrupperna, då kvinnorna hade en jämnare fördelning med 48% för omöjligt beslut och 50% för väj undan, jämfört med männen som valt 27% för omöjligt beslut och 66% för att väja undan.

4.1.3 Del III: Andra scenarion

4.1.3.1 Scenario sex

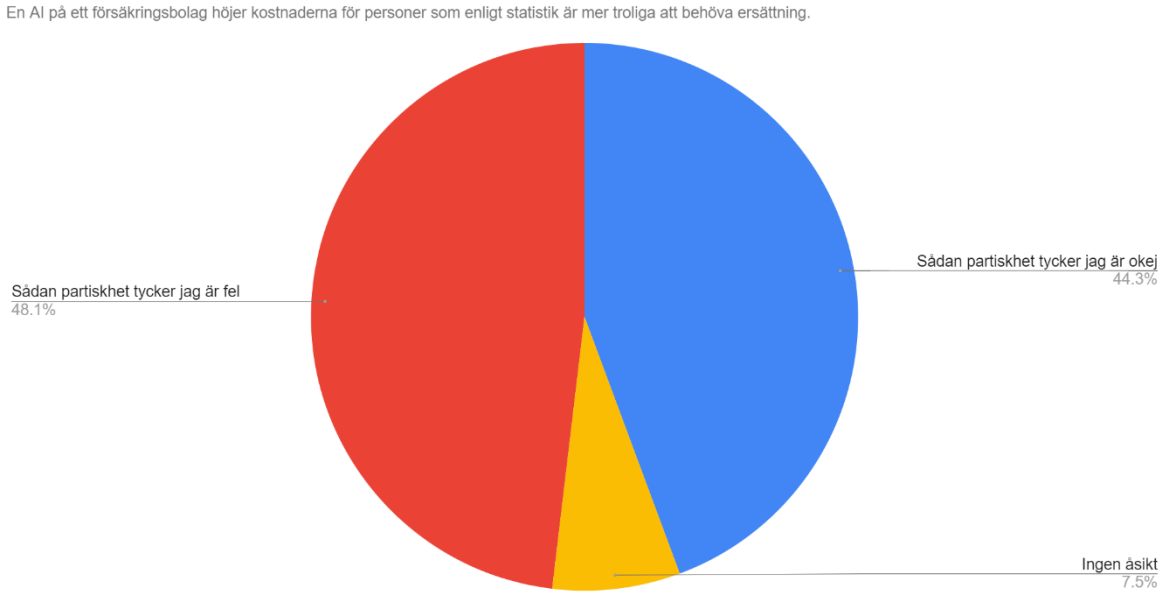


Figur 8: Enkätfråga scenario sex.

Respondenterna uppgav till 74% (78) att denna partiskhet ansågs fel, 23% (24) ansåg att det var okej, och 4% (4) hade ingen åsikt.

Demograffördelningen för denna fråga visade tydliga skillnader mellan könen, då kvinnor till 96% svarade att partiskheten ansågs fel och 2% att partiskheten ansågs okej. Männen svarade också enligt den totala majoriteten för frågan men till mindre utsträckning med 58% för att partiskheten ansågs fel och 37% för att partiskheten ansågs vara okej.

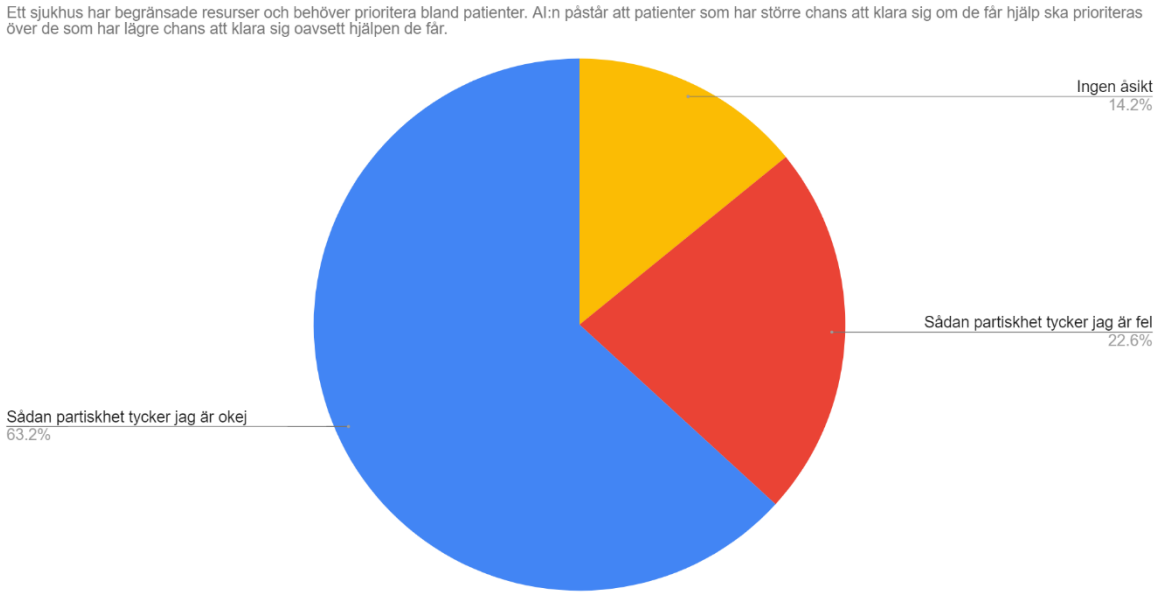
4.1.3.2 Scenario sju



Figur 9: Enkätfråga scenario sju.

Utan majoritet i frågan blev resultatet med 48% (51) att denna partiskhet var fel. 44% (47) ansåg att det var okej, och 8% (8) hade ingen åsikt.

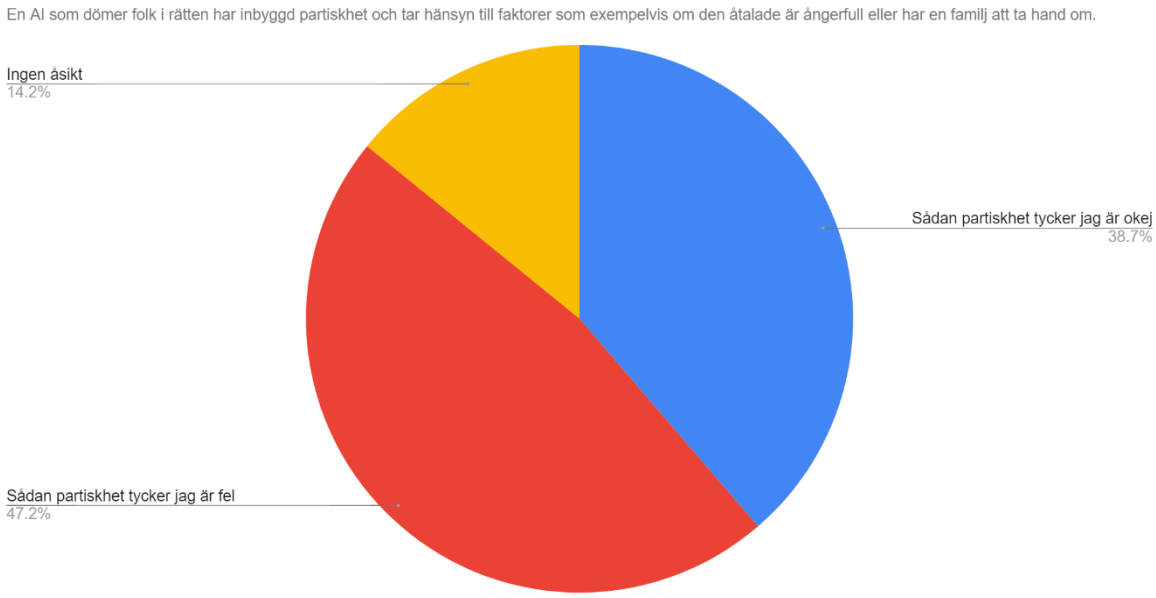
4.1.3.3 Scenario åtta



Figur 10: Enkätfråga scenario åtta.

En majoritet med 63% (67) ansåg att denna partiskhet var okej, med 23% (24) i motsatta åsikten att det var fel. 14% (15) hade ingen åsikt.

4.1.3.4 Scenario nio



Figur 11: Enkätfråga scenario nio.

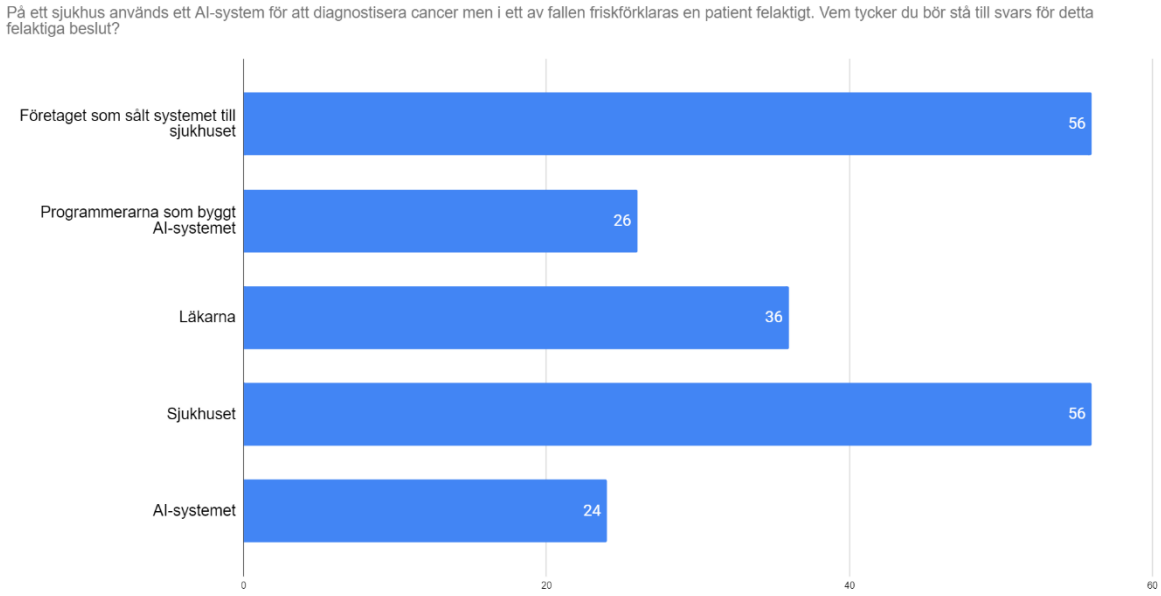
Utan majoritet tyckte respondenterna till störst del med 47% (50) att partiskheten i frågan var fel, och 39% (41) ansåg att den var okej. 14% (15) hade ingen åsikt.

Detta scenario visade en skillnad i svar mellan könsgrupperna, på 11 procentenheter, där män till 44% anser att det är okej jämfört med kvinnor som anser samma till 33%.

4.1.4 Del IV: Ansvarsfördelning

I denna del granskades svaren för vart scenario utifrån demografer på samma vis som i ovanstående delar. Inga skillnader på tio procentenheter eller mer hittades och demograf-granskning kommer därmed inte nämnas för följande fyra scenarion.

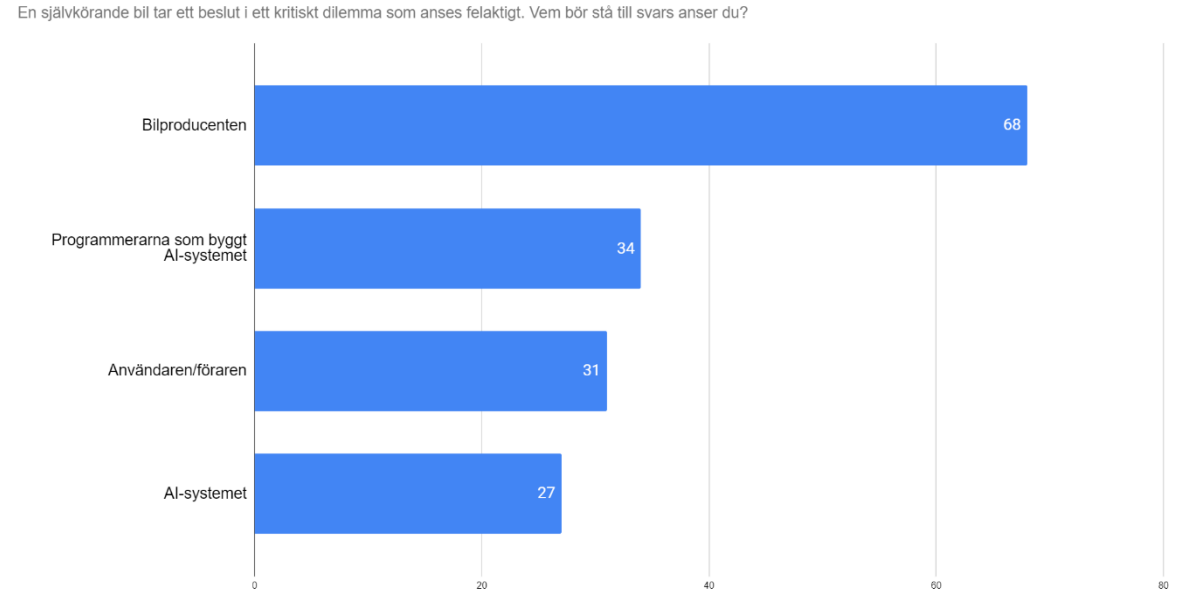
4.1.4.1 Scenario tio



Figur 12: Enkätfråga scenario tio.

Alternativen som ansågs ha högst ansvar var dels företaget som sålt systemet till sjukhuset, dels sjukhuset med 56 röster vardera. Alternativet programmerarna som byggt AI-systemet valdes 26 gånger och läkarna 36 gånger. Minst antal röster lades på själva AI-systemet, 24 röster fick detta alternativ.

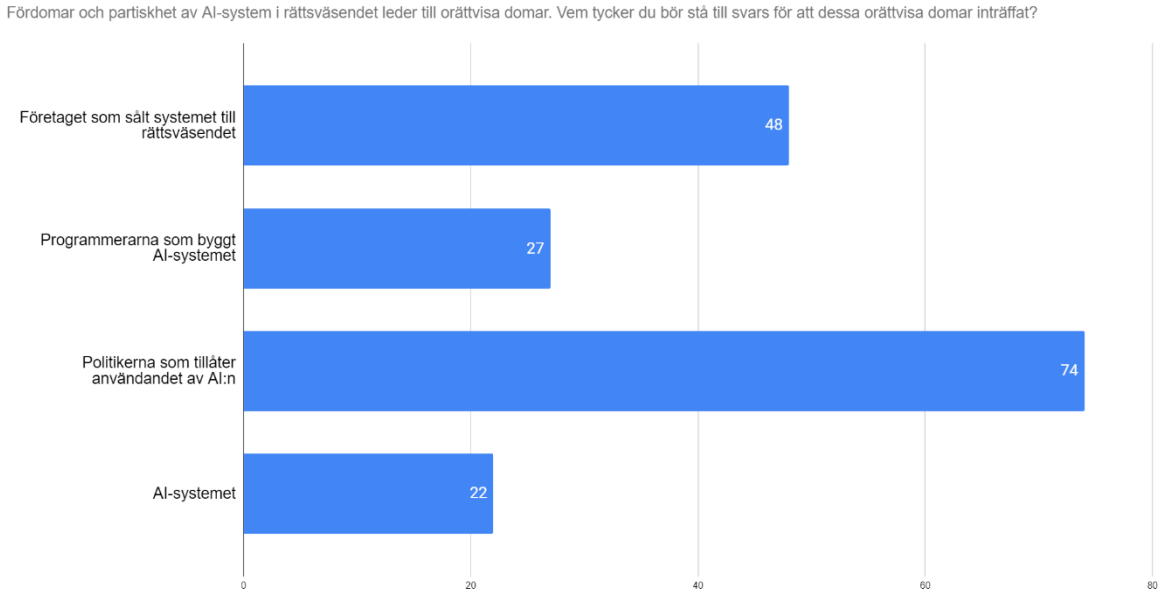
4.1.4.2 Scenario elva



Figur 13: Enkätfråga scenario elva.

Resultaten visade att majoriteten av rösterna lades på alternativet bilproducenten som valdes 68 gånger. Programmerarna som byggt AI-systemet valdes 34 gånger, 31 röster lades på att föraren hade sitt ansvar i denna situation medan AI-systemet själv fick 27 röster.

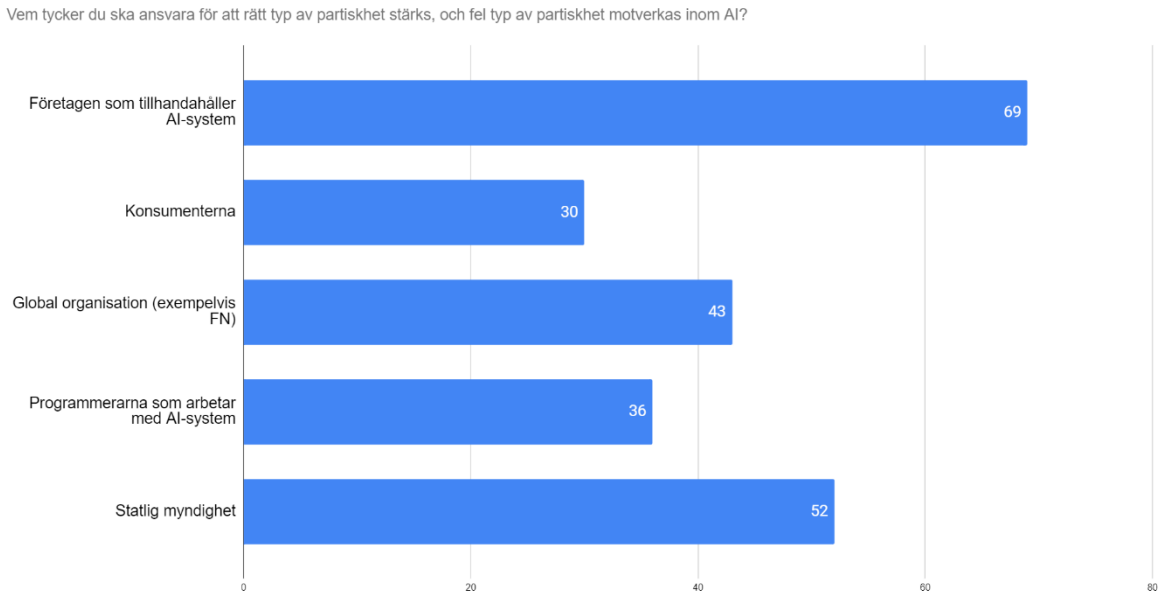
4.1.4.3 Scenario tolv



Figur 14: Enkätfråga scenario tolv.

Resultaten visade att 27 respondenter ansåg att programmerarna som byggt AI-systemet bör ta sitt ansvar. 48 respondenter valde alternativet företaget som sålt systemet till rättsväsendet, en majoritet på 74 respondenter valde alternativet politikerna som tillåter användandet av AI:n. 22 respondenter har valt att AI-systemet själv bör ha sitt ansvar.

4.1.4.4 Scenario tretton



Figur 15: Enkätfråga scenario tretton.

69 av respondenterna ansåg att det var företagen som har huvudansvar. Till stor del ansågs även att statlig myndighet och global organisation bar ansvar, 52 respektive 43 av respondenternas svar. Programmerarna ansågs av 36 respondenter ansvara och minst besvarade alternativet konsumenterna ansåg 30 av respondenterna ha ansvar.

4.2 Analys

Genom att stapla frågorna från respektive enkätdel (se Appendix A) ovanpå varandra och jämföra utefter demografen kön och utan hänsyn till frågans karaktär kan vissa tendenser att svara på ett visst sätt lyftas fram. Tabellen utläses horisontellt, och benämningen ”Total” syftar på den totala svarsfördelningen på frågorna utan hänsyn till demografer. Mindre skillnader identifierades som kan vara intressanta att föra vidare till diskussion med de gick ej över gränsen på 10% för att klassas som noterbar skillnad. Noterbara skillnader förklaras.

Tabell 1: Analys Del 2

	Håll kursen	Väj undan	Omöjligt beslut, låt slumpen avgöra
Kvinna	15%	36%	49%
Man	23%	41%	37%
Total	19%	39%	42%

I ovanstående tabell visas att ingen demograf sticker utanför den totala svarsfördelningen med mer än 7 procentenheter, och oftast inte heller med mer än 4 procentenheter. Dock fanns en skillnad mellan kön på 12 procentenheter, där kvinnor till 49% svarade att scenarierna är omöjliga beslut och låter slumpen avgöra utfallet, jämfört med männen, där 37% tog detta alternativ över hela del II.

Noterbara skillnader:

1. Omöjligt beslut, låt slumpen avgöra: Kvinna 49%, Man 37%, differens 12%

Tabell 2: Analys Del 3

	Sådan partiskhet tycker jag är okej	Sådan partiskhet tycker jag är fel	Ingen åsikt
Kvinna	34%	54%	11%
Man	48%	43%	9%
Total	42%	48%	10%

Denna tabell visar på samma vis som tabellen för Del 2 att skillnader från den totala svarsfördelningen är försumbara. Även i denna tabell skiljer det sig, då män till en högre grad än kvinnor svarade att scenarierna var okej och kvinnor till en högre grad svarade att de var fel.

Kvinnor valde därmed mer frekvent än män att partiskheten i givna scenarion var fel där 54% svarade detta jämfört med männen där det var 43%.

Noterbara skillnader:

1. Sådan partiskhet tycker jag är okej: Man 48%, Kvinna 34%, differens 14%
2. Sådan partiskhet tycker jag är fel: Kvinna 54%, Man 43%, differens 11%

Tabell 3: Analys individsvar

	Håll kursen	Väj undan	Omöjligt beslut, låt slumpen avgöra	Sådan partiskhet tycker jag är okej	Sådan partiskhet tycker jag är fel	Ingen åsikt
Antal 0	44	14	16	13	11	72
Antal 1	41	25	29	33	20	28
Antal 2	9	31	19	42	49	5
Antal 3	7	25	22	10	19	0
Antal 4	3	10	12	8	7	1
Antal 5	2	1	8			

För att undersöka om det finns mönster i hur respondenterna svarar, har individsvar räknats och granskats utan hänsyn till demografer. Varje individs svar har räknats för sig för att se hur många av varje svarsalternativ den gett, utav scenarierna i del 2 och 3. De värdena har sedan räknats tillsammans med resterande individer för att räkna hur frekvent förekommande svaren är. Antal *n* syftar således på antalet gånger en individ angett samma svarsalternativ. Värdet *x* visar då antalet individer som har angett kolumnens svarsalternativ *n* antal gånger. Som exempel var det 44 individer som svarade ”Håll kursen” 0 gånger, och det var 42 individer som svarade att partiskheten är okej 2 gånger. Vad detta bidrar med är en inblick i om det finns mönster för att hålla sig till ett svarsalternativ av princip.

5 Diskussion

I detta kapitel går vi genom våra empiriska resultat utifrån vad vi hittat enligt litteratur, samt diskuterar påverkan detta har på Moral Machine och vidare AI i stort.

5.1 Resultatdiskussion

Vår undersökning grundande sig i olika scenarion med understöd från litteratur för att stärka kredibiliteten. Att utgå ifrån etiska AI-situationer som endast är hypotetiska skulle inte vara gynnsamt då relevansen för slutsats inte skulle vara hög. En större del av den besvarade enkäten utgick ifrån frågor som härstammar ifrån programmet Moral Machine, och det var även här vi kunde identifiera problemområdet. Då Moral Machine, som i övrigt är den största datainhämtningsprogrammet på området med självkörande bilar, påpekat att urvalet är öppet kunde vi identifiera förbättringsmöjligheter (Awad et al., 2020b). Demografer som ålder och kön är inte något som spåras trots dess potentiella väsentlighet i att utvinna givande data. Här kunde denna undersökning komma in för att se om detta har resulterat i förlorade värdefulla data för skaparna av programmet och dess väsentlighet i samhället kring AI och dess partiskhet idag.

5.1.1 Demograf - kön

Totalt genom frågorna i enkäten identifierades noterbara skillnader i sju frågor av tretton. Samtliga av dessa skillnader finns i de nio frågorna som delas upp i Del II och Del III. Skillnaderna vi enligt litteraturen undersöker är en mer strikt etisk standard hos kvinnor (Nikoomaram et al., 2013), och mer utilitaristiska tendenser hos män (Arutyunova et al., 2016; Banerjee et al., 2010). Vad vi kan använda från resultaten för att testa kvinnors mer strikta etiska standard är den tydliga tendensen att i Del II ange svarsalternativet för omöjligt beslut i jämförelse med männen. Då resterande data från tabellerna över Del II och Del III är från olika scenarion går det inte att argumentera vare sig för eller emot denna teori ytterligare.

Mäns påstådda utilitaristiska tendenser kan testas med de frågor som behandlar praktiskhet utifrån ett samhällsperspektiv. Dessa är scenario två, scenario tre, scenario sex och scenario åtta (se Appendix A). I scenario två, som ställer en gravid kvinna på vägbanan med två kvinnor på den andra vägbanan, visar män en högre grad av utilitaristisk tendens än kvinnor för att hålla kursen och offra ett liv med ett annat liv i magen, för att rädda två liv. Män visar samtidigt i denna fråga en större tendens än kvinnor för att väja undan och offra två liv för att rädda ett liv med ett annat liv i magen. Att männen svarar i högre grad för båda alternativen kan alltså förklaras med att kvinnor hellre har svarat att det är ett omöjligt beslut. Det finns alltså inte något som pekar på att män i denna fråga är mer utilitaristiska än kvinnor, då skillnaden i svaren verkar vara en effekt av att män mer drar sig för att svara omöjligt beslut i frågorna. I scenario tre finns det också en skillnad i att männen i högre grad kommer att köra på den arbetslösa över chefen, men som i föregående scenario kan detta delvis förklaras med kvinnornas preferens för att välja omöjligt beslut. I scenario sex syns skillnaderna för utilitaristiska

tendenser då männen menar i högre grad än kvinnor att det skulle vara okej att anställa baserat på partiskhet för vem som beräknas vara trolig att ansöka om föräldraledighet, var det då hade varit mer praktiskt för företaget att anställa någon som är mindre trolig att göra detta. Fastän att majoritetsfördelningen ser lika dan ut mellan män och kvinnor så är skillnaderna signifikanta i frågan. För scenario åtta finns inga skillnader.

Då det finns antydan till att män skulle vara mer utilitaristiska, men att två av tillfällen av detta kan förklaras med ett annat fenomen och att ett tillfälle inte räcker för att påstå något eftersom det kan var på grund av andra orsaker än utilitaristiska, kan vi inte ge ett mer tydligt svar på teorin.

Vidare har en del av frågan varit att inte bara undersöka vilka skillnader som dyker upp, men även om skillnader som en helhet visar sig vara tillräckligt stora för att vara noteringsvärda. Litteraturen påpekar att det finns en anledning till varför man använder sig av demografer såsom ålder och kön, då de ofta kan förekomma en skillnad i svaren som anges (Eljertsson, 2019; Trost & Hultåker, 2016). Med den data vi samlat har vi, som sagt, sett skillnader över tio procentenheter förekomma i nio av frågorna på enkäten. Även med frågorna samlade tillsammans för att ge helhetsbild i tabeller har dessa skillnader visats. Vi menar att det med detta går att stärka argumentet för att det i denna undersökning finns signifikanta skillnader mellan könen.

5.1.2 Demograf – ålder

Skillnader som uppkom inom demografen ålder var inom enkätens delar inget noterbart. Bortfallet av respondenter inom olika åldersgrupper var alldeles för stort för att vi skulle kunna dra korrekta jämförbara analyseringar (Jacobsen & Andersson, 2017). Detta var olyckligt då tendensskillnader inom denna demograf skulle kunna vara väldigt givande att undersöka för att sedan kunna tänka sig ifall demografen hade kunnat vara applicerbar i Moral Machine.

5.1.3 Ett slumpmässigt utfall

För att också testa Nagler's et al. (2018) teori om att låta slumpen avgöra vid ett etiskt dilemma i syfte att undvika partiskhet, kan vi använda våra undersökningsresultat. Utan att vare sig bekräfta eller dementera teorin påvisar det enligt dessa resultat att ett slumpmässigt utfall hade varit i linje med 42% av respondenterna genom samtliga frågor från trafikscenarierna. Det är dock en annan sak att påstå att ett slumpmässigt utfall skulle vara standard och tillämpas i alla situationer, vilket enligt våra resultat enbart åtta respondenter (8%) visade visst stöd för genom att ange svarsalternativet på samtliga scenarion. Det ska dock noteras att dessa åtta respondenter inte nödvändigtvis stödjer en standardisering av slumpmässiga utfall, det kan alltså inte uteslutas att de svarade så utifrån scenariernas innehåll snarare än som en princip för slumpmässigt utfall.

5.2 Metoddiskussion

I mån att framtida studier ska kunna genomföra liknande undersökning på ett utvecklat sätt finns det flertalet punkter som hade kunnat förbättras och effektiviseras. Först och främst kan vi inte på ett tillförlitligt sätt säga att respondenterna representerar en hel population. Vad detta beror på är dels att vi inte kan stödja att vårt urval varken i storlek eller fördelning är

representativt nog för att kunna anses vara tillräckligt. Detta hade behövt läggas mer fokus på att få till på ett bra sätt och en absolut optimering behövs ifall forskningen av demografiska egenskapers påverkan skalas upp. Att utgå ifrån ett bekvämlighetsurval vid publikation av enkät är inte optimalt och nyanserade, grundande resultat blir därmed svåra att uppnå. Att lyfta frågan och identifiera tendensen till åsiktsskillnader inom AI-partiskhet är dock något som absolut är värt att poängtera, att möjligheten för framtida studier finns är något som trots allt värderas högt.

5.2.1 Begränsningar

Metoden som använts för själva enkäten har medfört begränsningar i resultaten med avseende på mängden data som samlats in och vilka faktorer som kan ha räknats med. En interaktiv och mer utförlig enkät hade varit fördelaktig, då följdfrågor kan ställas beroende på respondenternas svar för att bättre precisera vilken faktor i scenariot som spelade roll. Om en respondent svarar att exempelvis skona ett barn över en pensionär, kan nästa fråga anpassas till att ställa ett barn gentemot två pensionärer. På samma vis kan en följdfråga anpassas för att ställa en pensionär gentemot två barn ifall respondenten valt att skona pensionären. Vad detta hade bidragit med, om implementerat korrekt och till en större skala på hela enkäten, är att bättre kunna undersöka om respondenten föredrar att skona själva åldern i sig eller om åsikten ändrar sig utilitaristiskt när det då också behandlar aspekten av antal människor.

Även att slumpmässigt rotera scenariernas innehåll på ett vis som exempelvis skiftar mellan vilken vägbana som personerna står på i scenarierna om självkörande bilar, hade kunnat bättre visa om det faktiskt var scenarioobjektens egenskaper eller att välja mellan att hålla kursen och väja undan som spelade roll. Detta är en funktion som implementerats i Moral Machine, men som vi inte kunnat använda oss av utan ett lika sofistikerat datainsamlingsprogram. Med rotering av frågor mellan olika respondenter hade även fler scenarion kunnat ställas upp som kollar samma aspekter men i olika sammanhang, vilket hade gett en större mängd data även om respondentantalet för vår undersökning vore detsamma.

5.3 Nyttan med utveckling av Moral Machine

Det analyserade resultatet som framställts kan användas som argument till att lyfta frågan om att Moral Machine potentiellt går miste om värdefulla data vid sin datainhämtning. Verktyget Moral Machine framställdes med främsta syftet att få data på vad människor egentligen tycker om AI-partiskhet och att de också får göra sin åsikt hörd (Awad et al., 2020a). Att kunna optimera detta datainsamlingsverktyg parallellt med framfarten av självkörande bilar samt andra samhällsområden där AI appliceras skulle vara värdefullt för att se implementerad etik som folket har varit med och tagit fram. Att bidra med människans synpunkter kring AI-partiskhet och ge utrymme för folket att få göra sin röst hörd är gynnsamt för målet att AI-etik ska optimeras i sin helhet. Moral Machine har i nuläget samlat in mer än 40 miljoner svar vilket innebär 40 miljoner åsikter kring vad som ska anses rätt eller fel i givna scenarion. Vid en sådan massiv datainhämtning skulle hela verktyget och forskningen gynnas av mer nyanserade svar från respondenterna. Detta skulle resultera i att man till en högre grad kan ta hänsyn till responsen och applicera det som träningsdata till AI-system när parametrarna räknats om för att ge en mer rättvis bild av populationen (Danks & London, 2017). I nuläget är det endast geografisk plats som tas i åtanke med denna studie och då finns utrymme för inkludering av fler demografiska egenskaper. Att inkludera respondenters kön som en variabel i analysering av

Moral Machine's datainhämtning skulle bidra med ett helt nytt perspektiv till undersökningen. I denna studie hade man även helst sett ett givande resultat utifrån demografiska egenskapen ålder. Hade det skett så hade potentialen till inkludering av ytterligare variabel funnits. Utifrån resultatet på kön identifierades noterbara skillnader mellan könen. En hypotes baserad på litteraturen (Eljertsson, 2019; Trost & Hultåker, 2016; Arutyunova et al., 2016; Nikoomaram et al., 2013) som kan utformas därefter är, ifall urvalet av ålder hade varit korrekt hade somliga skillnader förmodligen förekommit. Även fast inte resultat kan ses här så är fortfarande denna demograf något som är gynnsamt att inkludera i vidare forskning. Även andra demografer som ekonomisk status, utbildningsnivå, religion är egenskaper som även dessa hade varit intressanta att undersöka för att se ytterligare utvecklingsmöjligheter med Moral Machine.

6 Slutsats

Syftet med uppsatsen har varit att presentera en grund för att kunna lyfta frågan om vidare forskning i ämnet och framför allt om en utveckling av Moral Machine behövs och om det hade varit givande. Med litterär grund, empiriska data och tillämpning av formulerad forskningsfråga har man kunnat identifiera en antydan till demografiska skillnader i åsikter inom AI-partiskhet, med vilket anses uppfylla syftet.

När det kommer till att explicit svara på forskningsfrågan, som är följande: ”*Hur skiljer sig åsikter mellan de demografiska egenskaperna ålder och kön angående frågan om partiskhet inom AI?”* kan man se att, efter analys av den empiriska data som hämtats av respondenter, noterbara skillnader i responsen hittas. Samtliga skillnader hittades inom demografen kön som jämförde respons mellan män och kvinnor vilket därmed påvisar en åsiktsskillnad. Demografen ålder kunde dessvärre inte analyseras då fördelningen mellan åldrar var förvrängd. Däremot kan fortfarande argumentet att frågan angående en demografisk påbyggnad av Moral Machine lyftas trots framställt resultat.

Sammanfattningsvis kan man argumentera för att en påbyggnad av Moral Machine bör ske då skaparna själva har identifierat bristen samt att denna studie har framställt ett resultat som indikerar att det förekommer skillnader i åsikter och respons. En utveckling av världens mest omfattande datainhämtningsmetod inom området, från att endast kunna spåra geografisk plats till att se en inkludering av andra variabler, kommer bidra till en AI-etik som är grundat i vad människor faktiskt tycker. Detta är även vad skaparna själva vill att forskningen ska resultera och bidra till i slutändan.

6.1.1 Vidare forskning

Som syftet för denna uppsats framför, är undersökningen till för att presentera en grund för om det finns belägg för vidare forskning i ämnet. Om demograferna ålder och kön har en påverkan på respondenters svar inom Moral Machine är det något som anses finnas belägg för – detta för att föra forskningen framåt. Andra områden inom AI-etik, som Moral Machine inte berör ser vi också ett behov för att utforska vidare, både för aspekten att samla in data på liknande vis som Moral Machine, samt aspekten att inkludera respondenters demografer för att undersöka om detta även här har en betydande del. Andra demografer utöver ålder och kön bör därför även undersökas för att så nyanserade data som möjligt kan genereras.

För att kunna uppnå en AI utveckling som är rättvis och representerar alla demografer behövs verktyg som Moral Machine optimeras och det är med denna studie man kan föra denna vision vidare.

För att gå vidare i forskningen om AI och dess beslutstagande utifrån mänsklig etik, med storskalig datainsamling som metod för skapandet av databas för AI:ns inlärning, är det relevant att även undersöka vidare hur respondenternas demografer i en sådan datainsamling behandlas.

Appendix A - Enkätguide

Tabell 4: Enkätguide

Enkätfråga	Svarsalternativ
Ålder	• 18-24 • 25-34 • 35-44 • 45+
Kön	• Kvinna • Man • Annat
DEL II: En självkörande, AI-styrd bil med passagerare hamnar i en kritisk situation där den inte hinner bromsa inför ett övergångsställe och är på väg att köra på någon. Det enda alternativet är att väja till andra vägbanan och i stället köra på någon annan. Båda alternativen har dödlig utgång. Ditt jobb är att berätta för bilen vad du tycker är det etiskt rätta att göra i följande scenarion. I alla scenarion har du ett helikopterperspektiv på situationen, så du är inte själv med i olyckan.	
1. Bilen är på väg mot 1 barn, på andra vägbanan går 1 pensionär.	• Håll kursen, kör på barnet • Väj undan, kör på pensionären • Omöjligt beslut, låt slumpen avgöra
2. Bilen är på väg mot 1 gravid kvinna, på andra vägbanan går 2 kvinnor (ej gravida).	• Håll kursen, kör på den gravida kvinnan • Väj undan, kör på de 2 kvinnorna • Omöjligt beslut, låt slumpen avgöra
3. Bilen är på väg mot 1 högt uppsatt chef, på andra vägbanan går 1 arbetslös person.	• Håll kursen, kör på chefen • Väj undan, kör på den arbetslösa • Omöjligt beslut, låt slumpen avgöra
4. Bilen är på väg mot 1 person, på andra vägbanan står en cementmur som vid krock kommer att döda bilens passagerare (1 person).	• Håll kursen, kör på personen • Väj undan, kör in i cementmuren • Omöjligt beslut, låt slumpen avgöra
5. Bilen är på väg mot 1 person som går lagligt (mot grönt) på övergångsstället, på andra vägbanan går 1 person som går olagligt (mot rött) på övergångsstället.	• Håll kursen, kör på personen som går mot grönt • Väj undan, kör på personen som går mot rött

	• Omöjligt beslut, låt slumpen avgöra
DEL III: Anställningsprocesser, försäkringar, sjukvård och rättsväsendet är exempel på andra sektorer där AI används som antingen beslutsstöd eller helt självständigt system. Partiska beslut kan ibland vara något bra och ibland något dåligt. Vi undrar nu vad du tycker är det rätta att göra i följande scenarion:	
6. En AI på ett företag väljer att anställa person A, som lever ensam, över person B, som nyligen gift sig. Motiveringen är att person B verkar vara troligare att snart ansöka om föräldraledighet och är därför det sämre valet för anställning.	• Sådan partiskhet tycker jag är okej • Sådan partiskhet tycker jag är fel • Ingen åsikt
7. En AI på ett försäkringsbolag höjer kostnaderna för personer som enligt statistik är mer troliga att behöva ersättning.	• Sådan partiskhet tycker jag är fel • Sådan partiskhet tycker jag är okej • Ingen åsikt
8. Ett sjukhus har begränsade resurser och behöver prioritera bland patienter. AI:n påstår att patienter som har större chans att klara sig om de får hjälp ska prioriteras över de som har lägre chans att klara sig oavsett hjälpen de får.	• Sådan partiskhet tycker jag är okej • Sådan partiskhet tycker jag är fel • Ingen åsikt
9. En AI som dömer folk i rätten har inbyggd partiskhet och tar hänsyn till faktorer som exempelvis om den åtalade är ångerfull eller har en familj att ta hand om.	• Sådan partiskhet tycker jag är okej • Sådan partiskhet tycker jag är fel • Ingen åsikt
DEL IV: Du har nu fått svara på vilka etiska värderingar du tycker ska föras vidare till artificiell intelligens. Denna sista del av undersökningen handlar nu om ansvar. Avancerad AI kan ha lärande funktioner och fatta beslut på helt fristående fot utan förklaringar, medan viss AI är tydligare i sina beräkningar och går efter fasta regler. Ibland kan en AI göra fel och då är det svårt att säga vem i ledet som borde bära ansvaret för det. Du kan ange flera svar.	
10. På ett sjukhus används ett AI-system för att diagnostisera cancer men i ett av fallen friskförklaras en patient felaktigt. Vem tycker du bör stå till svars för detta felaktiga beslut?	• Programmerarna som byggt AI-systemet • Företaget som sålt AI-systemet till sjukhuset • Sjukhuset • Läkarna • AI-systemet
11. En självkörande bil tar ett beslut i ett kritiskt dilemma som anses	• Programmerarna som byggt AI-systemet

felaktigt. Vem bör stå till svars anser du?	• Bilproducenten • Användaren/föraren • AI-systemet
12. Fördomar och partiskhet av AI-system i rättsväsendet leder till orättvisa domar. Vem tycker du bör stå till svars för att dessa orättvisa domar inträffat?	• Programmerarna som arbetar med AI-systemet • Företaget som sålt AI-systemet till rättsväsendet • Politikerna som tillåter användandet av AI-systemet • AI-systemet
13. Vem tycker du ska ansvara för att rätt typ av partiskhet stärks, och fel typ av partiskhet motverkas inom AI?	• Konsumenterna • Programmerarna som arbetar med AI-systemet • Företaget som tillhandahåller AI-systemet • Statlig myndighet • Global organisation (exempelvis FN)

Referenser

Angwin, J., Larson, J., Mattu, S. & Kirchner, L. (2016). Machine Bias. *Propublica*, 23 May 2016.

Arutyunova, K. R., Alexandrov, Y. I. & Hauser, M. D. (2016).Sociocultural Influences on Moral Judgments: East–West, Male–Female, and Young–Old, *Frontiers in Psychology*, vol. 7, no. 1334**,** pp

Awad, E., Dsouza, S., Bonnefon, J.-F., Shariff, A. & Rahwan, I. (2020a). Crowdsourcing Moral Machines. *Communications of the ACM.*

Awad, E., Dsouza, S. & Chang, P. (n.d.). Moral Machine.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F. & Rahwan, I. (2018).The Moral Machine Experiment, *Nature*, vol. 563, no. 7729**,** pp 59-64

Awad, E., Dsouza, S., Shariff, A., Rahwan, I. & Bonnefon, J.-F. (2020b).Universals and Variations in Moral Decisions Made in 42 Countries by 70,000 Participants, *Proceedings of the National Academy of Sciences*, vol. 117, no. 5**,** pp 2332

Balasubramanian, R., Libarikian, A. & McElhaney, D. (2021).Insurance 2030 - the Impact of Ai on the Future of Insurance, *McKinsey & Company*, vol. no.

Banerjee, K., Huebner, B. & Hauser, M. (2010).Intuitive Moral Judgments Are Robust across Variation in Gender, Education, Politics and Religion: A Large-Scale Web-Based Study, *Journal of Cognition and Culture*, vol. 10, no. 3-4**,** pp 253-281

BBC. (2018). Amazon Scrapped 'Sexist Ai' Tool. *BBC*, 10 October 2018.

Danks, D. & London, A. (2017). Algorithmic Bias in Autonomous Systems. in*:* Carles Sierra, I.-C. (ed.) *Twenty-Sixth International Joint Conference on Artificial Intelligence.* Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization.

Eljertsson, G. (2019). Enkäten I Praktiken - En Handbok I Enkätmetodik.

EuropeiskaKommissionen. 2018. Etiska Riktlinjer För Tillförlitlig Ai, EuropeiskaKommissionen

Fernández, C. & Fernández, A. (2019).Ethical and Legal Implications of Ai Recruiting Software, *ERCIM NEWS,* vol. no. 116**,** pp 22-23

Ferrer, X., Nuenen, T. v., Such, J. M., Coté, M. & Criado, N. (2020).Bias and Discrimination in Ai: A Cross-Disciplinary Perspective, vol. no.

Guillot, J. D. (2021a).Artificial Intelligence: Threats and Opportunities, *European Parliament News,* vol. no.

Guillot, J. D. (2021b).What Is Artificial Intelligence and How Is It Used?, *European Parliament News,* vol. no.

Hao, K. & Stray, J. (2019). Can You Make Ai Fairer Than a Judge? Play Our Courtroom Algorithm Game. *MIT technology review*, 17 October 2019.

Havens, J. (2019). Heartificial Intelligence - Embracing Our Humanity to Maximize Machines - John C. Havens.

Jacobsen, D. I. & Andersson, S. (2017). Hur Genomför Man Undersökningar? : Introduktion Till Samhällsvetenskapliga Metoder, Lund: Studentlitteratur.

Johansson, S. & Vogelgesang, U. (2016).Automating the Insurance Industry, *McKinsey & Company,* vol. no.

Kaushal, A., Altman, R. & Langlotz, C. (2020). Health Care Ai Systems Are Biased. *Scientific American*.

Kompella, K. (2020).The Trolley Problem and Ethical Dilemmas in Ai, *Information Today,* vol. 37, no. 5**,** pp 35-35

Li, L., Ota, K. & Dong, M. (2018).Humanlike Driving: Empirical Decision-Making System for Autonomous Vehicles, *IEEE Transactions on Vehicular Technology,* vol. 67, no. 8**,** pp 6814-6823

Luthi, B. (2020). Study Points to Rate Bias in U.S. Auto Insurance Industry. *Investopedia*, 11 December 2020.

Matthias, A. (2004).The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata, *Ethics and Information Technology*, vol. 6, no. 175-183

McCarthy, J. (2007).What Is Artificial Intelligence, vol. no.

Merriam-Webster. (n.d.-a). Arificial Intelligence. *Merriam-Webster.com dictionary.*

Merriam-Webster. (n.d.-b). Intelligence. *Merriam-Webster.com dictionary.*

Nagler, J., Van den Hoven, J. & Helbing, D. (2018).An Extension of Asimov's Robotics Laws, *SSRN Electronic Journal*, vol. no.

Nikoomaram, H., Roodposhti, F. R., Ashlagh, A. T., Lofti, F. H. & Taghipourian, Y. (2013).The Effects of Age, Gender, Education Level and Work Experience of Accountant on Ethical Decision Making by Using Fuzzy Logic, *International Research Journal of Applied and Basic Sciences*, vol. 4, no.

Oates, B. J. (2006). Researching Information Systems and Computing.

I, Robot, 2004. Directed by Proyas, A.

Silberg, J. & Manyika, J. (2019).Tackling Bias in Ai (and in Humans), *McKinsey Global Institute*, vol. no.

Sinclair, M. (2013).Attitudes, Norms and Driving Behaviour: A Comparison of Young Drivers in South Africa and Sweden, *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 20, no. 170-181

Socialstyrelsen. 2020. Nationella Principer För Prioritering Inom Intensivvård under Extraordinära Förhållanden, Socialstyrelsen

Thomson, J. J. (1985).The Trolley Problem, *The Yale Law Journal*, vol. 94, no.

Trost, J. & Hultåker, O. (2016). Enkätboken.

Unesco. (n.d.). *Artificial Intelligence: Examples of Ethical Dilemmas* [Online]. Available online: https://en.unesco.org/artificial-intelligence/ethics/cases [Accessed april 27 2021].

Violago, V. & Quevada, N. (2018).Ai: The Issue of Bias, *Managing Intellectual Property,* vol. 277, no. 32-36

Wang, P. (2019).On Defining Artificial Intelligence, *Journal of Artificial General Intelligence,* vol. 10, no. 1-37