GÖTEBORGS UNIVERSITET

STATSVETENSKAPLIGA INSTITUTIONEN

TRANSPARENS, EN NYCKELFAKTOR?

En experimentell studie om transparens påverkar förtroendet för användning av artificiell intelligens (AI) som beslutstöd inom hälso- och sjukvårdenssektorn

Jenny Svensson

Examensarbete:	15 hp
Program:	Statsvetarprogrammet
Nivå:	Grundnivå
Termin/år:	Vårterminen 2024
Antal ord:	11 913
Handledare:	Jacob Sohlberg

Abstract

The adoption of artificial intelligence (AI) technologies is increasing across various disciplines within society with significant potential in the healthcare sector. As automation increasingly replaces tasks traditionally performed by humans, the public's trust in the new technology becomes crucial. Trust in the healthcare sector as an institution is essential for its functionality and effectiveness. This study investigates public perceptions about the use of AI as decision-making tools within the healthcare sector using an experimental design. The study aims to determine whether enhanced transparency within the decision-making process influences public trust regarding the use of AI, based on theories of institutional trust and transparency. The study involved 201 participants and the data was collected in Gothenburg, Sweden between 16/4-2024 until 28/4-2024. The results are in line with the two hypotheses and statistically significant differences were found between the two groups in the trust levels for AI as decision-making tools. The results imply that transparency has implications for members of society to trust the technology and should therefore be considered for keeping the public trust for the healthcare institution. But more research is needed to investigate other factors that may also impact this trust.

Keywords: Artificial intelligence, decision- making, transparency, institutional trust, healthcare sector

Innehållsförteckning

1. Inledning **1**

1.2 Disposition *3*

2. Teoretisk bakgrund och tidigare forskning **3**

2.1 Redogörelse för begreppet och fenomenet "förtroende" *3*

2.2 Institutionellt förtroende generellt och specifikt inom hälso- och sjukvården *4*

2.3 Betydelsen av transparens och transparens i beslutsfattandeprocesser *5*

2.4 Sambandet mellan transparens och förtroende *6*

2.5 Transparens och användning av AI inom hälso- och sjukvården *8*

2.6 Attityder för användning av AI inom hälso- och sjukvården *10*

2.7. Studiens hypoteser och tillhörande teoretiska modeller *12*

3. Metod **12**

3.1 Val av metod *13*

3.2 Experimentell design *13*

3.3 Distribution av enkät och studiens urval *14*

3.4 Behandling *15*

3.5 Operationalisering av beroende variabler samt andra variabler som avses mätas *17*

3.6 Kontroll för randomisering och manipulation *19*

3.7 Etiska överväganden *19*

4. Datamaterial **20**

4.1 Randomiseringskontroll *21*

4.2 Urvalets representativitet *22*

4.3 Indexkonstruktioner *22*

5. Resultat **23**

5.1 Resultat och analys för studiens första hypotes *24*

5.2 Resultat och analys för studiens andra hypotes *25*

5.3 Potentiella implikationer till studiens resultat *27*

5.4 Test för eventuella samspelsvariabler *31*

6. Diskussion och slutsats **31**

7. Referenslista **34**

8. Appendix **39**

1. Inledning

Artificiell intelligens (AI) introduceras inom flera sektorer i vårt samhälle (Davenport & Kalakota; 2019; Horgan et al., 2019) och en av de sektorerna är hälso- och sjukvården (Davenport & Kalakota, 2019; Horgan et al., 2019; Läkemedelsverket, 2023). Användning av AI inom hälso- och sjukvårdssektorn förväntas ha en betydande roll för framtidens hälsosektor, både för den enskilda individen och för samhället i stort (Horgan et al., 2019). Inom hälso- och sjukvården kan AI-teknik exempelvis användas för att assistera vid diagnostik och utföra administrativa uppgifter (Davenport & Kalakota, 2019). AI-tekniken förväntas bidra med ökad effektivitet, bättre vård utifrån individuella förutsättningar samt bidra med ekonomiska vinster (Horgan et al., 2019; Buch et al., 2018; Davenport & Kalakota, 2019). I och med teknikens genomgripande inflytande ställs det krav på att institutioner såsom exempelvis hälso- och sjukvården implementerar och använder tekniken på ett säkert sätt för att inte riskera negativa konsekvenser (Davenport & Kalakota, 2019). Automatisering av uppgifter som historiskt sett har utförts av människor är problematisk i de fall när samhällets medborgare inte litar på tekniken (Lee & See, 2004). En av de främsta utmaningarna för en framgångsrik implementering av AI inom hälso- och sjukvården är allmänhetens tveksamhet mot tekniken (Buch et al., 2018).

Förtroende, och däribland institutionellt förtroende är viktigt för ett välfungerande samhälle (Warren, 2018) och ett högt institutionellt förtroende för hälso- och sjukvården är avgörande för systemets funktionalitet (Gilson, 2003). När nya interventioner implementeras inom hälso- och sjukvården är medborgarnas förtroende för institutionen avgörande för samhällets acceptans och användning för de nya interventionerna (Gilson, 2003).

Användning av AI inom hälso- och sjukvårdssektorn medför såväl kända som okända risker (Läkemedelsverket, 2023). Forskningen belyser att användning av AI ger upphov till varierande utmaningar och frågor, däribland problematik med transparens (Davenport & Kalakota, 2019; Kiseleva et al., 2022; Rojahn et al., 2023), frågor gällande ansvarsutkrävande (Johansson et al., 2024; Davenport & Kalakota, 2019) och huruvida användningen är etisk och moralisk (Davenport & Kalakota, 2019). Vid implementering av AI inom hälso- och sjukvårdssektorn involveras flertalet aktörer, bland annat vårdgivare, de företag som utvecklar tekniken och de som mottager vården (Davenport & Kalakota, 2019; Rojahn et al., 2023). Tidigare studier belyser att det är viktigt att förstå vårdtagarnas uppfattning för användningen av AI för att kunna

behålla mottagarnas förtroende (Rojahn et al., 2023). Forskare har även påpekat att mer forskning behövs gällande just allmänhetens uppfattning för använding av AI då tidigare studier huvudsakligen har studerat attityder inom vårdproffessionen eller bland AI-experter (Rojahn et al., 2023). Tidigare studier, av både interdisciplinär karaktär (Bankins et al., 2022; Lee et al., 2018) och studier inom hälso- och sjukvården (Rojahn et al., 2023) har huvudsakligen jämfört människors uppfattning för AI-genererade beslut, mot beslut fattade av människor. Studiernas resultat visar huvudsakligen att människor föredrar den mänskliga faktorns beslut, trots medvetenhet om de fördelar som AI-teknik kan medföra (Rojahn et al., 2023). Men i de efterforskningar jag har gjort har jag inte funnit studier som har studerat och jämfört, om människors förtroende för användning av AI varierar beroende hur vi hanterar utmaningarna (transparens, ansvarsutkrävande, etikalitet) som forskningen har identifierat att AI ger upphov till. Transparens har en viktig roll inom hälso- och sjukvården, eftersom beslut avser människors hälsa och liv (Kiseleva et al., 2022). Problematiken med transparens ses som en av de svåraste utmaningarna för en framgångsrik implementering av AI inom hälso- och sjukvården (Davenport & Kalakota, 2019). Därför avser jag att i denna uppsats fokusera på att jämföra om allmänhetens förtroende påverkas av hur vi hanterar utmaningen gällande just transparens i samband med användning av AI som beslutsstöd.

I den här uppsatsen avser jag mot ovanstående bakgrund att fokusera på utmaningen gällande transparens som användning av AI ger upphov till. Studiens syfte är att förstå om graden av transparens påverkar allmänhetens förtroende för användning av AI som beslutsstöd inom hälso- och sjukvården. Forskningsfrågan jag avser besvara är: *Påverkas allmänhetens förtroende för användning av AI som beslutsstöd inom hälso- och sjukvården beroende på hur transparent beslutsfattandeprocessen är?*

För att besvara forskningsfrågan används en experimentell metod med 201 deltagare, vilka har rekryterats i centrala Göteborg mellan 16/4–2024 – 28/4–2024. Den experimentella metoden ger möjligheter till slutsatser om orsak och verkan (Esaiasson et al., 2014) vilket lämpar sig väl för att studera och finna slutsatser om studiens samband mellan transparens och förtroende. Deltagarna i både behandlingsgruppen och jämförelsegruppen tog del av en fiktivt skriven text som jag som författare skapat med inspiration från tidigare studier, nyhetsartiklar men även egen kreativitet. Det som skiljer texterna åt mellan studiens två grupper är *graden av transparens*. Studiens resultat visar att deltagarna som tog del av behandlingen med *hög*

transparens uppfattade transparensen som högre, och erhöll ett högre förtroende för användning av AI som beslutsstöd inom hälso- och sjukvårdssektorn.

1.2 Disposition

Dispositionen för uppsatsen är enligt följande. Studiens introducerades ovan med en beskrivning för det vetenskapliga problem som jag avser studera, en förklaring av studiens syfte och forskningsfrågan jag avser att besvara. Nedan presenteras studiens teoretiska bakgrund och relevant forskning, och kapitlet leder fram till studiens hypoteser. Därefter presenteras studiens metod med tillhörande etiska överväganden samt studiens datamaterial. Sedan presenteras studiens resultat och uppsatsen avslutas med en diskussion och slutsats med förslag till framtida forskning. Studiens behandling presenteras i Appendix.

2. Teoretisk bakgrund och tidigare forskning

I detta kapitel redogörs studiens teoretiska bakgrund och relevant forskning. Kapitlet inleds med en förklaring för två av studiens centrala begrepp, *institutionellt förtroende* och *transparens*. Sedan presenteras sambandet mellan förtroende och transparens, vilket följs av ett avsnitt om transparens och användning av AI specifikt inom hälso- och sjukvårdssektorn. Därefter presenteras tidigare forskning och teori om attityder för användning av AI inom hälso- och sjukvården. Kapitlet avslutas med en redogörelse för studiens hypoteser.

2.1 Redogörelse för begreppet och fenomenet "förtroende"

Förtroende är ett komplext begrepp med relevans inom flera områden (Nannestad, 2008). Begreppet och fenomenets definitioner samt operationaliseringar varierar beroende på kontext (Grimmelikhuijsen et al., 2013; Lee & See., 2004; Nannestad, 2008; Rousseau et al., 1998). Rousseau et al., (1998) har försökt åstadkomma en interdisciplinär definition och definierar förtroende som "Trust is a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" (s. 395). Förtroende är relationellt och uppstår i interaktioner mellan exempelvis människor och människor, eller mellan människor och organisationer (Gilson, 2003). Forskare har länge studerat och intresserat sig för att förstå hur och varför människor litar på andra människor, institutioner eller aktörer i samhället (Grimmelikhuijsen et al., 2013). Denna studie fokuserar på institutionellt förtroende.

2.2 Institutionellt förtroende generellt och specifikt inom hälso- och sjukvården

I demokratier är avsikten att samhällets institutioner ska tillhandahålla gemensamma nyttigheter, exempelvis hälso- och sjukvård, ett fungerande rättssystem samt att leverera social såväl som nationell säkerhet (Warren, 2018). För att kunna erbjuda tjänsterna krävs ett högt institutionellt förtroende från samhällets medborgare. Utan institutionellt förtroende utmanas grundläggande demokratiska förväntningar såsom personlig säkerhet och frihet, välfärd och stöd, samt ekonomisk arbetsfördelning (Warren, 2018). Således är det institutionella förtroendet viktigt för en välfungerande demokrati och ett välfungerande samhälle (Warren, 2018)

Institutionellt förtroende avser förtroende för organisationer såsom den offentliga förvaltningen eller den styrande regeringen (Herreros, 2023). Sœderskov och Dinesen (2015) inspireras av Zmerlis och Newtons (2007) beskrivning av institutionellt förtroende och definierar det som ”an individual’s perception of the credibility, fairness, competence and transparency of state institutions” (s. 181). Institutionellt förtroende beskrivs vara ett samlingsnamn för flera aspekter, och det differentieras mellan förtroende för den representativa sidan (parlament och partier) och den implementerande sidan, det vill säga genomförandet av den offentliga politiken (Rothstein & Stolle, 2003).

Det institutionella förtroendet består av en integrerad relation mellan samhällets medborgare och dess styrning (Rothstein & Stolle, 2003; Herreros, 2023). Huvudsakligen baseras det institutionella förtroendet på institutionens egenskaper snarare än förvaltaren (Warren, 2018). Institutionell opartiskhet, rättvisa och kompetens framhävs som nyckelfaktorer för institutionellt förtroende (Rothstein & Stolle, 2003). Förtroendet påverkas även av demografiska faktorer såsom kön, ålder och utbildning (Robertson, 2023). Exempelvis tenderar högutbildade och äldre personer att erhålla ett högre institutionellt (Hudson, 2006). I samhällen med institutionellt förtroende erhåller samhällets medborgare högre förtroende för offentliga tjänster såsom exempelvis sjukvård (Rothstein & Stolle, 2003).

Hälso- och sjukvårdssystemets funktionalitet baseras på institutionellt förtroende och ett förtroendefullt hälso- och sjukvårdssystem är viktigt för samhället i stort (Gilson, 2003). Hälso- och sjukvården är en sociopolitisk institution som erhåller en större roll i samhället än att leverera sjukvård då institutionen generellt ansvarar för ett mer hälsosamt samhälle (Gilson, 2003). Förtroende för sektorn grundas på ett komplext men integrerat nätverk. Förtroendet baseras delvis på interaktionen mellan vårdgivare och vårdtagarna samt delvis på förhållandet mellan hälso- och sjukvårdssystemet och samhället, det vill säga utifrån samhällets normer och

värderingar. Interaktionen mellan vårdgivare och vårdtagare influeras av den större kontexten, det vill säga det institutionella förtroendet (Gilson, 2003).

Det institutionella förtroendet för hälso- och sjukvården är avgörande för att systemet ska kunna tillhandahålla de tjänster som samhället behöver (Gilson, 2003). Systemet fungerar endast väl och effektivt om dess handlingar anses legitima från samhällets medborgare. När nya interventioner implementeras inom hälso- och sjukvården är det institutionella förtroendet avgörande för samhällets acceptans och användning (Gilson, 2003).

2.3 Betydelsen av transparens och transparens i beslutsfattandeprocesser

Transparens förklaras vara ett komplext begrepp med varierande definitioner (Hood & Heald, 2006; Hollyer et al., 2011). Heald (2006) förklarar att transparens varierar i dimensioner, riktningar och beroende vilket ”habitat” transparensen tillämpas inom. Demokratier kännetecknas bland annat av “nedåtriktad transparens”, vilket syftar till att medborgare har insyn i styrningen (Heald, 2006). Transparens ger positiva effekter som att exempelvis förhindra korruption och ge ökad rättvisa (Hood, 2006; Heald, 2006), och enligt Birkinshaw (2006) bör transparens dessutom ses som en mänsklig rättighet. Transparens inkluderar alla faktorer som påverkar informationsflödet i ett samhälle (Hollyer et al., 2011). De Fine Licht et al., (2014) beskriver transparens som “Transparency refers to the degree that information is made available about how and why decisions are produced within a certain institution” (s. 112). Således beskrivs transparens inkludera öppenhet för regler och beteenden, vilket varierar beroende på vilka det handlar om och vilka det tillämpas för, exempelvis mellan regeringar, medborgare eller organisationer (Hood & Heald, 2006).

I denna studie fokuserar jag på transparens i beslutsfattandeprocessen, vilket avser processerna mellan vad Heald (2006) benämner som ”händelser”. Händelserna syftar på delarna i beslutsfattandeprocessen som externt sett är synliga och oftast mätbara, det vill säga *insatserna* till, *resultatet* av och *konsekvenserna* från beslutet. Processerna mellan händelserna är svårare att mäta men kan beskrivas om informationen finns tillgänglig (Heald, 2006). Transparens inom beslutsfattandeprocessen innebär öppenhet för beslutets procedur samt dess bakomliggande motiv vilket ger medborgare en möjlighet att förstå och uppfatta huruvida beslutet grundas på normer som tidigare har kommits överens om (Grimmelikhuijsen et al., 2013).

De Fine Licht et al., (2014) skiljer mellan transparens för det *rationella* och transparens för *processen*. Skillnaden är att det förstnämnda fokuserar på logiken till varför ett beslut tas,

medan transparens för processen avser förklaringar för exempelvis diskussionerna runtomkring beslutet. Variationerna av transparens tjänar olika syften och ger därför till viss del varierande effekter. Transparens för det rationella syftar till att öka förståelsen och acceptansen för beslutet genom att redogöra för logik och rationalitet. Transparens i processen syftar till att visa öppenhet för att beslutet grundas på rättvisa principer, vilket avser att öka förtroendet för både processen och beslutsfattarna.

2.4 Sambandet mellan transparens och förtroende

Transparens förklaras vara en nyckelfaktor för ”a better governance”, det vill säga en bättre styrning (Hood & Heald, 2006). Hood (2006) skiljer på ”governance” och ”government” och menar att ”governance” inkluderar styrningen för alla typer av organisationer och institutioner medan ”government” specifikt innebär för samhällets regering. Transparens bidrar till ökat förtroende för de förtroendevalda och för samhällets styrning (Hood, 2006; Heald, 2006). Det förklaras bero på att redogörelser för hur och varför ett beslut tas, ökar medborgarnas förståelse och ger därmed högre förtroende för de styrande. I likhet med Hood och Heald (2006) argumenterar Szompka (1998) för att demokratiska institutioner ger upphov till faktorer som främjar förtroende, och exemplifierar däribland transparens.

Till skillnad från Hood och Heald (2006) är O’Neill (2002) mer skeptisk till ökad transparens. Enligt O’Neill (2002) kan transparens ge insyn i komplexa beslutsprocesser vilket riskerar att skapa förvirring snarare än förtroende. En del forskare (De Fine Licht, 2014; Grimmelikhujisen et al., 2013) menar att effekterna av transparens är mer kontextberoende än vad som tidigare argumenteras för.

De Fine Licht et al., (2014) studerar sambandet mellan transparens och legitimitet både teoretiskt och empiriskt. Författarna definierar legitimitet enligt Tylers (2006) beskrivning, “a psychological property to an authority, institution, or social arrangement that leads those connected to it believe that it is appropriate, proper and just” (s. 375). De Fine Licht et al., (2014) presenterar agentteorin, deliberativ demokratiteori och teorin om procedurrättvisa för sambandet mellan transparens och legitimitet. Enligt agentteorin medför transparens ökad legitimitet eftersom mottagarnas osäkerhet att beslutsfattarna tar beslut som inte följer moraliska principer reduceras (Holmström, 1979). Utifrån deliberativ demokratiteori ger transparens bättre förståelse för bakgrunden till beslutet, vilket ökar förtroendet för både beslutet och beslutsfattarna. Enligt den tredje teorin, procedurrättivseteorin medför transparens

öppenhet i beslutets procedur vilket enligt både Birkinshaw (2006) och De Fine Licht et al., (2014) är ett värde i sig. Förtroendet ökar även genom att visa att proceduren har varit rättvis. Samhällets medborgare kan endast bedöma ett besluts rättvisa om de är medvetna om proceduren (De Fine Licht et al., 2014). Teorierna ger goda skäl att förvänta sig att ett positivt samband mellan transparens och ökad legitimitet (De Fine Licht et al., 2014). Författarna identifierar däremot även mekanismer som kan påverka sambandet negativt. Transparens riskerar medföra en känsla av frustration och maktlöshet, om mottagaren upplever att hen inte har möjlighet att påverka proceduren. Mottagaren kan också uppleva missnöjdhet med beslutet eller missförstå beslutet, vilket författarna beskriver som "Bismarck disappointment effect" (De Fine Licht et al., 2014, 118). Transparensens effekter beror således på kontexten och effekterna påverkas beroende hur involverad mottagaren är i frågor gällande transparens (De Fine Licht et al., 2014)

Empiriskt studerar De Fine Licht et al., (2014) sambandet mellan transparens i beslutsfattandeprocesser och legitimitet genom en experimentell metod inom en skoltext med 400 deltagare i åldrarna 17–19 år. Studiens resultat ger stöd för att transparens kan öka legitimitet, men att transparensens effekter är kontextberoende och komplexa. Det är svårt att avgöra vilka mekanismer som orsakar sambandet men huvudsakligen förklaras effekten bero på faktorer i linje med teorin om procedurrättvisa och/eller deliberativ demokratiteori. Deltagarna som fick en förklaring till varför besluten togs, och insyn för anledningarna till beslutet, accepterade beslutet i större utsträckning. Full transparens förklaras däremot inte nödvändigtvis öka legitimiteten mer än måttlig transparens. Det vill säga, det viktigaste är inte att ta beslut med öppna dörrar utan att orsakerna till beslutet motiveras väl (De Fine Licht et al., 2014).

Även De Fine Licht (2014) studerar sambandet mellan transparens och förtroende, och fokuserar på om politikområdet som transparensen tillämpas inom påverkar sambandet. Urvalet bestod av 1032 deltagare som inte var representativt för Sveriges befolkning då antalet män och antalet högutbildade var överrepresenterade i jämförelse mot Sveriges befolkning. Politikområdena som studien fokuserar på är finansiering av kultur och fritid samt trafiksäkerhet. Politikområdena används som exempel på områden som behandlar olika "trade-offs" med varierande nivåer av moralisk känslighet. Beslut om finansiering av kultur och fritid avser mer rutinmässiga beslut med lägre moralisk känslighet medan beslut om trafiksäkerhet innefattar högre moralisk känslighet. Enligt författaren kan explicita eller implicita redogörelser

för moraliska dilemman påverka sambandet både positivt och negativt (De Fine Licht, 2014). Resultatet ger stöd för att transparens kan öka förtroende, men att transparensens effekter varierar beroende på område det tillämpas inom. Effekterna av transparens var mer positiva när besluten behandlade rutinmässiga procedurer snarare än moraliskt känsliga områden. Således belyser De Fine Licht (2014) att transparensens effekter riskeras förbli negativa när det tillämpas inom mer moraliskt känsliga områden. Beslutsfattare måste därför beakta kontexten som transparens tillämpas inom när effekterna av transparens ska analyseras (De Fine Licht, 2014).

2.5 Transparens och användning av AI inom hälso- och sjukvården

Transparens inom hälso- och sjukvården har en viktig roll eftersom beslut inom hälso- och sjukvården avser människors hälsa och liv (Kiseleva et al., 2022). Inom sektorn fyller transparensen flera funktioner, exempelvis säkerställa kvalitet och säkerhet, möjliggöra ansvarsutkrävande och ge patienter möjligheten att ta informerade beslut (Kiseleva et al., 2022). I likhet med Hood & Heald (2006) förklaring för variationer av transparens, skiljer Kiseleva et al., (2022) inom hälso- och sjukvårdssektorn mellan intern och extern transparens. Intern transparens avser transparens mellan vårdpersonal som arbetar inom systemet. Extern transparens avser transparensen från vårdprofessionen mot patienter, det vill säga vårdtagarna (Kiseleva et al., 2022). Således fokuserar denna uppsats på extern transparens.

Hälso- och sjukvården grundas på etiska principer och när AI introduceras inom sektorn ställs det krav på att tekniken introduceras i linje med principerna för att behålla medborgarnas förtroende (Horgan et al., 2019). Transparens är en av de grundläggande etiska principerna, men även ansvarsutkrävande, säkerhet och rättvisa beskrivs som etiskt grundläggande principer (Horgan et al., 2019). I och med att människor historiskt sett har varit beslutsfattare inom hälso- och sjukvården medför AI en ovana i interaktionen med tekniken, vilket väcker frågor gällande både transparens och ansvarsutkrävande (Davenport & Kalakota, 2019).

AI har funnits sedan 1960-talet, men än idag finns ingen gemensam definition (Davenport & Kalakota, 2019; Läkemedelsverket, 2023). AI beskrivs som en strävan att efterlikna den mänskliga hjärnan (Jiang et al., 2019), och Rojahn et al., (2023) beskriver AI som "Generally, the term AI is often used to refer to algorithms and statistical models that can communicate and reason independently in a variety of scenarios in a way that is similar or even superior to a human" (s.1). AI bör ses som en samling av flera tekniker (Davenport & Kalakota, 2019; Jiang

et al., 2019) och tillämpas med varierande metoder inom hälso- och sjukvården beroende på användningsområde (Davenport & Kalakota, 2019).

AI används inom hälso- och sjukvården idag och förväntas få en större roll i framtiden (Davenport & Kalakota, 2019; Horgan et al., 2019). Inom sektorn används AI bland annat som beslutsstöd och som medicinsk rådgivning (Läkemedelsverket, 2023). Generellt sätt beskrivs AI ha en viktig roll för framtidens hälso- och sjukvård (Davenport & Kalakota, 2019; Horgan et al., 2019), och förväntas förbättra hälsosektorn för både den enskilda individen och för samhället i stort (Horgan et al., 2019). Användning av AI beskrivs bland annat bidra till möjligheter att förbättra sektorns effektivitet och därmed spara tid (Buch et al., 2018), både gällande diagnostik och administrativa uppgifter (Davenport & Kalakota, 2019). Effektivitet är fördelaktigt när den mänskliga expertisen är en begränsad resurs (Buch et al., 2018). AI kan assistera vårdprofesionen i olika beslut och öka sannolikheten att besluten blir korrekta vid första tillfället. AI förväntas även möjliggöra bättre vård utifrån personliga omständigheter och medföra ekonomiska vinster (Horgan et al., 2019).

Vid användning av AI problematiseras den etiskt grundläggande principen transparens (Horgan et al., 2019) då AI-teknik kan betraktas vara som ”en svart låda” (Rojahn et al., 2023). Med ”svart låda” avses att det är svårt att förstå hur och varför systemet har resulterat i dess slutsats (Rojahn et al., 2023). AI-verktyg varierar i vilken grad de anses vara ”en svart låda”, det vill säga i vilken grad verktygen är transparenta (Kiseleva et al., 2022). Tillhörande till debatten om AI som ”svart låda”, är att den mänskliga hjärnan till viss del också förklaras vara som ”en svart låda” eftersom alla överväganden och val som görs inför ett beslut inte alltid explicit förklaras, eller kan förklaras (Rojahn et al., 2023). Trots det menar Rojahn et al., (2023) att människan i större utsträckning har enklare att lita på den mänskliga faktorns beslut.

Problematiken med bristen på transparens beskrivs vara en av de svåraste utmaningarna för implementering av AI inom hälso- och sjukvårdssektorn (Davenport & Kalakota, 2019), och problematiken oroar politiska beslutsfattare världen över (Kiseleva et al., 2022). När en patient informeras om ett beslut eller en diagnos kommer patienten följaktligen vilja veta hur och varför beslutet har tagits (Davenport & Kalakota, 2019). Rojahn et al., (2023) menar att AI:s bristande förklaringar kan vara en huvudsaklig anledning till att människor som saknar kunskap om AI, i mindre utsträckning litar på verktygets beslut inom hälso- och sjukvården. Onekligen kommer även AI- verktyg innebära att misstag görs vid exempelvis diagnostisering (Kiseleva et al., 2022), och således problematiseras även ansvarsutkrävandefrågan. Transparens har således en

särskilt viktig funktion om något går fel eftersom vi behöver förstå hur och varför det gick fel inför framtida beslut (Kiseleva et al., 2022).

2.6 Attityder för användning av AI inom hälso- och sjukvården

Rojahn et al., (2023), Gao et al., (2020) och Johansson et al., (2024) har studerat de allmännas attityd till användning av AI inom hälso-och sjukvårdssektorn. Rojahn et al., (2023) undersöker genom en internetenkät med experimentell metod hur den generella folkopinionen uppfattar och litar på AI inom hälso- och sjukvården. I studien kontrasteras AI mot den mänskliga faktorn (vårdprofessionen) som medicinska beslutsfattare. Deltagarna föredrog den mänskliga faktorns beslut i jämförelse med AI-verktygets, trots att deltagarna var medvetna om att AI kan medföra fördelar sett till väntetider, ge bättre vård utifrån individuella förutsättningar samt att den mänskliga faktorn i större utsträckning tar kulturellt partiska beslut. Författarna menar att resultatet kan tänkas bero på att det finns en etablerad relation till den mänskliga faktorn som vårdgivare, vilket baseras på bekantskap, kunskap och lojalitet (Rojahn et al., 2023). Som också nämnt i avsnitt 2.5 *(transparens och användning av AI inom hälso- och sjukvården)* menar författarna att om människor inte förstår hur eller varför ett beslut har tagits kan vi heller inte lita på tekniken (Rojahn et al., 2023).

Likt Rojahn et al., (2023) undersöker Gao et al., (2020) allmänhetens uppfattning för användning av AI inom hälso- och sjukvården, men studiernas resultat skiljs sig något åt. Datamaterialet hämtas från en social medie i Kina i syfte att förstå orsaken till den publika opinionens oro för AI inom hälso- och sjukvården. Resultatet visar att majoriteten erhöll positiva attityder för användning av AI, men det fanns även deltagare med neutrala och negativa attityder. De positiva attityderna förklaras huvudsakligen av de tekniska fördelar som AI förväntas medföra, exempelvis precision i diagnostik och ökad effektivitet. De positiva attityderna motiverades även av att AI kan reducera diskriminering mot patienter, förbättra rättvisa och reducera konflikter mellan vårdgivare och vårdtagare. De neutrala attityderna bestod bland annat av människor som inte tog ställning åt varken det ena eller andra hållet. De negativa attityderna motiverades av att den tekniska utmaningen står inför svårigheter och en misstro mot företagen som skapar teknologin. De negativa attityderna grundades även på en oro för avsaknad av empati när AI-tekniken jämfördes mot den mänskliga faktorn (Gao et al., 2020). Det liknar Rojahn et al., (2023) studie där människor föredrog den mänskliga faktorn när användning av AI kontrasterades mot den mänskliga faktorn.

I den svenska kontexten har bland annat Johansson et al., (2024) studerat kvinnors uppfattning till användning av AI i samband med mammografi. Studiens resultat visar generellt sett en positiv inställning till användning av AI, men att deltagarna däremot föredrog den mänskliga faktorn. Den mänskliga faktorn ansågs ha bättre förståelse för den större kontexten och kan därför bidra med mer empati. Deltagarna beskrev att det var viktigt att den mänskliga faktorn var involverad i den första bedömningen, och visade en oro för att inte förstå AI-verktyget. Därför belyser Johansson et al., (2024) att transparent användning av AI är avgörande för människors förtroende för tekniken. Gällande ansvarsutkrävande erhöll studiens deltagare varierande åsikter. En del deltagare visade högre tolerans för misstag av den mänskliga faktorn jämfört med ett AI-verktyg. Andra deltagare förväntade sig att AI-verktyg inte begår misstag, medan andra uttryckte att det alltid bör vara vårdprofessionen som har det sista ansvaret när felaktigheter inträffar. Författarna argumenterar därför att det krävs tydliga riktlinjer för ansvarsutkrävandet vid implementering av AI inom mammografi. Deltagarna visade däremot även en förståelse för fördelarna att kombinera AI med den mänskliga faktorns kunskap (Johansson et al., 2024).

I samband till den empiriska forskningen om varför människor föredrar beslut från människor framför AI-verktyg, trots fördelarna, blir socialpsykologen Jonathan Haidt (2013) teorier om etik och moral relevanta. Haidt (2013) menar att moraliska uppfattningar baseras på magkänslan och individens intuition, före det rationella tänkandet. Därför menar Haidt (2013) att det är komplicerat att definiera och finna universella regler för vad som är, och vad som uppsattas som etiskt och moraliskt.

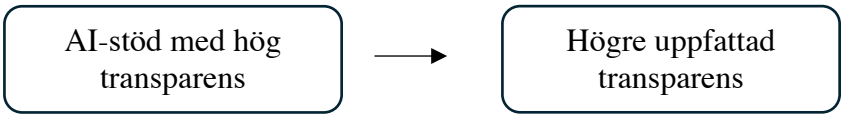
Således visar forskningen att användning av AI ger upphov till fler utmaningar än endast transparens, exempelvis frågorna gällande *ansvarsutkrävande* (Johansson et al., 2024; Davenport & Kalakota, 2019; Kiseleva et al., 2022) och *etik och moral* (Davenport & Kalakota, 2012; Haidt, 2013). Emmanuel et al., (1996) beskriver ansvarsutkrävande som ”At its most general, accountability is about individuals who are responsible for a set of activities and for explaining or answering for their actions” (s.229). Asan och Choudhury (2022) beskriver ansvarsutkrävande enligt följande: “Accountability can be explained as a process in which healthcare practitioners have the potential responsibility to justify their “clinical actions” to patients and are held liable for the consequences” (s.2).

2.7. Studiens hypoteser och tillhörande teoretiska modeller

Studiens första hypotes kan anses något given, men för att säkerställa att deltagarna har uppfattat transparensen på olika sätt mellan de två grupperna väljer jag att mäta fenomenet. Därav lyder studiens första hypotes:

Hypotes 1: Personer som läser beslutet med ***högre transparens*** i beslutsfattandeprocessen kommer att uppfatta transparensen i beslutet som ***högre***.

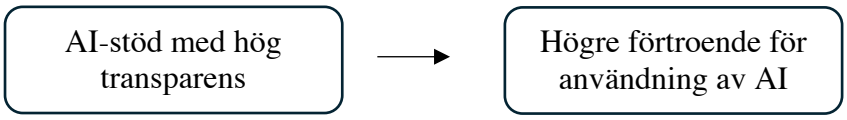
Figur 1: Teoretisk modell, hypotes 1



Studiens andra hypotes baseras på följande bakgrund. Sambandet mellan transparens och förtroende argumenteras av flertalet forskare medföra positiva effekter för de styrandes legitimitet (Hood & Heald, 2006; Szompka, 1998; De Fine Licht et al., 2014; De Fine Licht, 2014). Både Davenport och Kalakota (2019) och Kiseleva et al., (2022) belyser att transparens inom hälso- och sjukvården har en betydande roll eftersom beslut avser människors liv och hälsa. Rojhan et al., (2023) och Johansson et al., (2024) argumenterar för att när AI introduceras inom hälso- och sjukvården är det avgörande att människor ska förstå AI-tekniken för att erhålla ett förtroende. Således att beslutet är transparent. Därav lyder studiens andra hypotes:

Hypotes 2: Personer som läser beslutet med ***högre transparens*** i beslutsfattandeprocessen kommer ***ha ett högre förtroende*** för användning av AI som beslutsstöd.

Figur 2: Teoretisk modell, hypotes 2



3. Metod

För att besvara forskningsfrågan genomförs ett survey-experiment med endast-efter design. I kapitlet nedan redogörs studiens val av metod, design och urval. Sedan förklaras studiens randomisering och kontroll av manipulation. Avsnittet avslutas med en redogörelse för experimentets behandling samt operationaliseringar av övriga variabler.

3.1 Val av metod

Den experimentella metoden är en lämplig metod för välgrundade slutsatser om kausalitet, det vill säga slutsatser om orsak och verkan (Esaiasson et al., 2017). Studiens oberoende variabel tillåts isoleras och manipuleras, och andra förklaringsfaktorer kan uteslutas (Esaiasson et al., 2017). Således kan den oberoende variabeln i studien, "grad av transparens" isoleras och målsättningen är att andra förklaringsfaktorer utesluts. Vid väl utfört experiment erhåller metoden en enkelhet att utläsa om det finns ett kausalt samband mellan studiens oberoende och beroende variabler. I översiktsartikeln beskriver Nannestad att (2008) experiment har genomförts i tidigare studier som studerat förtroende (Nannestad, 2008), och metoden används även i studier som studerar sambandet mellan transparens och förtroende (De Fine Licht, 2014; De Fine Licht et al., 2014).

3.2 Experimentell design

Experimentet genomförs i form av ett survey-experiment (enkätstudie) med endast-efter design. Endast-efter design innebär att enkäten genomförs efter behandlingen, och vanligtvis introduceras momenten vid samma tillfälle (Esaiasson et al., 2017). Studiens deltagare tar därför först del av behandlingen och besvarar sedan den efterföljande enkäten. Momenten introduceras vid samma tillfälle.

Studien är uppdelad i fyra moment, vilka presenteras nedan. Det första momentet avser att informera deltagarna att studien är frivilligt och anonym, samt kort informera om studiens syfte. Sedan följer en binär fråga om deltagarna har tidigare kunskap om AI inom hälso- och sjukvården. Det andra momentet avser studiens behandling och är utformad som en kort, fiktivt skriven nyhetsartikel *(Se avsnitt 3.4 "Behandling")*. Efter behandlingen introduceras enkäten, vilket utgör operationaliseringarna för variablerna som avses mätas. I enkäten tar deltagarna först ställning till sju påståenden och bedömer i vilken utsträckning de instämmer med respektive påstående, svarsalternativen tillhandahålls på en fyrgradig likertskala från "Instämmer helt" till "Instämmer inte alls". Därefter besvaras tre flervalsfrågor och enkäten avslutas med ytterligare tre frågor av deskriptiv karaktär gällande om ålder, utbildningsnivå och kön. Enkäten avslutas med mer utförlig information om studiens syfte.

Studiens design, endast-efter design, har både för- och nackdelar. Designens tydliga fördel är tidseffektiviteten eftersom datainsamlingen endast sker vid ett tillfälle (Esaiasson et al., 2017). Således lämpar sig designen väl i och med den relativt korta tidsomfattning som finns till förfogande (10 veckor). Nackdelen med designen är att det inte går att säkerställa om det på

innan behandlingen finns någon skillnad mellan studiens grupper (Esaiasson et al., 2017). En alternativ design för att på förhand kontrollera eventuella skillnader mellan grupperna är, före-efter design, vilket innebär att två identiska mätningar tillämpas, en före behandling och en efter. Nackdelen med före-efter design är att det inte är går att kontrollera i vilken utsträckning den första mätningen påverkade behandlingen, exempelvis om deltagarna uppmärksammade begrepp som de annars inte hade uppmärksammat eller aktiverade särskilda tankeprocesser (Esaiasson et al., 2017). Mitt val för endast-efter design motiveras dels av tidseffektiviteten som designen erhåller. Sedan kan det även resoneras att designen är den mest relevanta eftersom syftet är att studera effekterna av transparens i det specifika fallet. Enkätfrågorna tillhör och baseras på behandlingen vilket gör att enkäten inte är relevant utan att läsa behandlingen. Däremot kan grupperna tänkas ha olika grundinställningar exempelvis om transparensens betydelse och huruvida AI är transparent. I och med att deltagarna randomiseras *(se avsnitt 3.3 "Distribution av enkät och studiens urval"*) bör grupperna systematiskt inte skilja sig åt. En väl fungerande randomisering gör att exempelvis lika många AI-experter (och andra påverkande faktorer) tilldelas till respektive grupp, vilket således hanterar problematiken att grupperna på förhand riskerar skilja sig åt.

3.3 Distribution av enkät och studiens urval

För att kunna belägga kausalitet är slumpmässig och korrekt allokering, även kallat randomisering till respektive grupp avgörande (Esaiasson et al., 2017). Randomisering är ett avgörande moment i den experimentella metoden och avser att fastställa studiens interna validitet. Om deltagarna inte allokeras till en specifik grupp på ett systematiskt sätt bör egenskaperna inte skiljas åt på ett förutbestämt sätt, förutsatt att urvalet är tillräckligt stort (Esaiasson et al., 2017). För min studie betyder det att randomiseringen är avgörande för adekvata slutsatser om det är variationen av transparens som påverkar förtroendet för användning av AI. Vid lyckad randomisering erhåller studiens två grupper lika många exempelvis kvinnor som män, lika antal högutbildade och lika antal med tidigare kunskap om AI. I studien görs randomiseringen av Qualtrics randomiseringsfunktion och således påverkas randomiseringen inte av mig som forskare, vilket är fördelaktigt eftersom jag av misstag hade kunnat orsaka systematiska misstag i tilldelningen.

Som nämnt är syftet att studera den publika opinionens förtroende för användning av AI inom hälso-sjukvården. Det kan föras ett logiskt resonemang att alla medborgare på ett eller annat sätt kan komma i kontakt med vården eftersom hälsoproblem kan möta vem som när som.

Studiens urval är ett bekvämlighetsurval vilket är vanligt förekommande vid experimentstudier (Esaiasson et al., 2017). Bekvämlighetsurval fungerar i experimentstudier eftersom rekryteringsprocessen inte påverkar studiens kvalitet i lika utsträckning som randomiseringsprocessen (Esaiasson et al., 2017). Studiens urval avser personer i samhället, vilka tillfrågas och rekryteras i ”lättillgängliga områden”. Deltagare rekryteras vid lekplatser i Göteborg, på Göteborgs Stadsbibliotek, vid Göteborgs centralstation och närliggande busshållplatser, vid Korsvägen, på Humanistiska Institutionen och Institutionen för Statsvetenskap vid Göteborgs Universitet. Enkäten sprids till ett fåtal via mina sociala medier, däremot inte fullständigt eftersom det riskerar att orsaka en snedvridning i urvalet. Urvalet påverkar studiens externa validitet, därmed möjligheterna att generalisera resultaten till en större population. Urvalet påverkar däremot inte studiens interna validitet.

Urvalsstorleken påverkar möjligheten att identifiera statistiskt signifikanta skillnader mellan grupperna (Esaiasson et al., 2017). Antalet deltagare påverkar experimentets statiska styrka vilket syftar till sannolikheten att förkasta en falsk nollhypotes (Sundell, 2012, 11 november) och ett större urval ger högre statistisk styrka (Esaiasson et al., 2017). Tumregeln för experimentella studier är att varje grupp bör inkludera minst 30–40 deltagare. I denna studie avser jag att rekrytera 100 deltagare till respektive grupp vilket ökar studiens statistiska styrka.

3.4 Behandling

Experimentets behandling kan ses som operationaliseringen för studiens oberoende variabel (Esaiasson et al., 2017). Behandlingen avser att variera den oberoende variabeln mellan studiens grupper, det vill säga *graden av transparens*. Rekommendationen är att kraftfullt variera den oberoende variabeln eftersom det är enklare att analysera signifikanta resultat än nollresultat (Esaiasson et al., 2017). Min avsikt är därför att skapa en stark behandling.

Behandlingen utformas så att samtliga deltagare tar del av en fiktivt skriven text, vilket gör att båda grupperna på sätt och vis utsätts för en behandling. En renodlad kontrollgrupp avser en grupp som inte utsätts för någon behandling överhuvudtaget (Esaiasson et al., 2017). Därför benämns inte grupperna som behandlings alternativt kontrollgrupp. Gruppen som läser texten med hög transparens, avses vara “behandlingsgruppen” men benämns som “AI-stöd med hög transparens”. Gruppen som läser texten med låg transparens, avses vara “kontrollgrupp” men benämns som “AI-stöd med låg transparens”.

Behandlingen centrerar kring “Linda” som söker vård för oro för diabetes. Texten är skriven som en nyhetsartikel och handlar om Lindas vårdbesök där ett AI-verktyg används som beslutsstöd. Det som skiljer texterna åt är *graden av transparens* vid användningen av AI-verktyget som beslutsstöd. I gruppen ”AI-stöd med hög transparens” avses och beskrivs transparensen vara hög, medan transparensen avses och beskrivs vara låg i gruppen ”AI-stöd med låg transparens”. För att tydliggöra nivån av transparens inleds texterna med tydliga rubriker huruvida beslutet erhöll välgrundade eller bristfälliga förklaringar. Gruppen ”AI-stöd med hög transparens” läser rubriken ”AI-verktyg gav rätt diagnos i vården - med tydliga förklaringar och öppna beslut”, medan gruppen ”AI-stöd med låg transparens” läser rubriken ”Trots rätt diagnos av AI-verktyg i vården - stora frågetecken över bristande förklaringar”. Min avsikt är att rubrikerna tydligt ska uppmärksamma deltagarna på transparensens betydelse och nivå. Sedan redogörs nivån av transparens ytterligare en gång genom ett enskilt stycke i varje text. Gruppen ”AI-stöd med hög transparens” ges tydliga förklaringar till varför beslutet togs och transparensen redogörs på följande sätt:

”Beslutet motiverades och förklarades väl med att algoritmen inte identifierade några avvikelser eller fann några risker i Lindas tillstånd/bakgrund för att utveckla diabetes. Linda fick på så sätt en tydlig förklaring om hur och varför AI-verktyget kom fram till dess slutsats. Det gav henne insikt i beslutsprocessen och en tydlig förklaring till varför beslutet om diagnosen togs.”

I gruppen ”AI-stöd med låg transparens” kontrasteras behandlingsgruppens information och det informeras om att Linda inte fick någon förklaring till varför beslutet togs. Båda texterna avslutas med ett identiskt stycke om att ansvarsutkrävandet för beslutet fungerar väl. Således är det som skiljer texterna åt graden av transparensen. Min avsikt är att konstanthålla faktorerna ansvarsutkrävande och etikalitet (samt den mänskliga faktorns involvering) mellan texterna eftersom jag vill isolera den oberoende variabeln för slutsatser om kausalitet. För detaljerad information om studiens behandling se Appendix.

Till följd av att båda grupperna till viss del utsätts för en behandling försvåras möjligheten att dra slutsatser om det är den höga transparens eller låga transparens som påverkar resultatet. Det vill säga, om resultatet visar sig vara statistiskt signifikant försvåras möjligheten att bedöma om det är den låga eller är höga transparensen som har orsakat utfallet.

3.5 Operationalisering av beroende variabler samt andra variabler som avses mätas

För att besvara forskningsfrågan operationaliseras studiens beroende variabler och andra variabler som avses mätas. Samtliga variabler (bortsett från samspelsvariablerna samt variablerna som avser beskriva urvalet) konverteras till ett index på två empiriska operationaliseringar. Indexkonstruktioner är vanligt förekommande inom samhällsvetenskaplig forskning och avser att skapa mer tillförlitliga operationaliseringar för variabeln som studeras (Esaiasson et al., 2017). Avsikten med indexkonstruktioner är att öka mätningens reliabilitet. Det uppnås genom att risken för slumpmässiga och osystematiska fel hanteras och minskas i jämförelse mot om variabeln endast operationaliseras till en indikator. Det är viktigt att belysa att avsikten är att öka reliabiliteten, men det blir inte alltid som tänkt. Förutsättningen för indexkonstruktionernas reliabilitet är att konstruktionerna baseras på valida operationaliseringar (Esaiasson et al., 2017). För att nå så hög begreppsvaliditet som möjligt operationaliseras därför respektive variabel i den mån det går på operationaliseringar från tidigare forskning, däremot kan vissa ändringar komma att göras för att anpassa operationaliseringarna till studien.

Den beroende variabeln, *förtroende för AI-verktygets beslut,* operationaliseras till ett index på två frågor. Den första operationaliseringen avser att deltagarna får ta ställning till i vilken utsträckning de instämmer med följande påstående: "AI-verktygets beslut är av god kvalitet". Svarsalternativen presenteras på en fyrgradig likertskala, från "Instämmer helt" till "Instämmer inte alls". Operationaliseringen baseras på Lee et al., (2018) operationalisering av förtroende "How much do you trust that the [algorithm/ manager] make good-quality [decision specified in the scenario]?" (s.7). Den beroende variabelns andra operationalisering är "Jag litar på att AI-verktyget har ställt rätt diagnos". Operationaliseringen baseras på Körber (2018) operationalisering för förtroende, "I trust the system" (s.21). Svarsalternativen är densamma som för den första operationaliseringen.

Uppfattad transparens operationaliseras till ett index på två frågor. Svarsalternativen för operationaliseringarna är kodade från "Instämmer helt" - "Instämmer inte alls". Den första operationaliseringen avser påståendet "Förklaringen som ges till beslutet är bristfällig" baserat på Blader & Tyler (2003) operationalisering. Höga värden innebär att förklaringen till beslutet uppfattas som bristfällig medan låga värden innebär att beslutet ses som transparent. Den andra operationaliseringen för *uppfattad transparens* avser påståendet "Beslutsprocessen är

transparent" (De Fine Licht et al., 2014). Höga värden innebär att beslutet uppfattas som transparent medan låga värden innebär att transparensen uppfattas som bristfällig.

När transparens manipuleras finns det en risk för att även att de två teoretiskt identifierade faktorerna *ansvarsutkrävande* och *etikalitet* manipuleras. Därför mäts både ansvarsutkrävande och etikalitet. *Ansvarsutkrävande,* operationaliseras till ett index på två frågor. I likhet till uppfattad transparens är svarsalternativen för båda operationaliseringarna kodade från "Instämmer helt" - "Instämmer inte alls". Den första operationaliseringen avser "Vården tar ansvar för sitt beslut" och baseras på Emmanuel et al., (1996) beskrivning av ansvarsutkrävande. Den andra operationaliseringen bestäms till "Vården följer upp ärendet ansvarsfullt och tar fullt ansvar för eventuella konsekvenser av beslutet" (Asan & Choudhury, 2022).

Etikalitet, operationaliseras till ett index på två frågor. Den första operationaliseringen avser påståendet: "Beslutet fattas på ett etiskt och moraliskt sätt". Operationaliseringen baseras på Haidt (2013) socialpsykologiska teorier om att etik och moral baseras på intuition före rationalitet. Den andra operationaliseringen för etikalitet avser frågan "Har AI-verktyget behandlat Linda med värdighet och respekt?" och baseras på Colquitt (2001) operationalisering " Has (he/she) treated you with dignity?" (s. 389). Svarsalternativen för operationaliseringarna är kodade från "Instämmer helt" – "Instämmer inte alls".

För att kunna beskriva urvalet och identifiera eventuella samspelsvariabler (interaktionsvariabler) avslutas enkäten med frågor av deskriptiv karaktär (ålder, kön och utbildning). *Ålder* tillfrågas genom frågan "Vilket år är du född?" följt av att svaret ska besvaras med födelseår. *Kön* tillfrågas genom en flervalsalternativ-fråga med svarsalternativen kvinna, man eller annat. För att mäta *utbildningsnivå* ställs frågan: "Vilken utbildningsnivå har du avslutat?" med svarsalternativen, ej avslutad grundskoleutbildning, grundskola, gymnasieutbildning, yrkeshögskola, högskola/universitetsstudier (kandidatexamen), universitetsstudier (längre än tre år) samt examen från forskarutbildning. I enkäten ställs även frågan " Enligt din mening, i vilken utsträckning går det att lita på människor i allmänhet? ". Frågan är inspirerad från hur frågan efterfrågas i World Value Survey (2017–2021, s.7) och beskrivs av Nannestad (2008) användas i flera enkäter såsom exempelvis General Social Survey och European Social Surveys. Frågan inkluderas i enkäten för att kunna beskriva urvalet.

3.6 Kontroll för randomisering och manipulation

I experimentet kontrolleras både studiens randomisering och manipulation. För att kontrollera att randomiseringen har fungerat väl genomförs en randomiseringskontroll. I randomiseringskontrollen inkluderas kön, ålder och utbildningsnivå, vilket beskrivs vara vanliga faktorer att kontrollera för vid randomiseringskontroller (Esaiasson et al., 2017). Jag avser även att kontrollera för tidigare kunskap för användning av AI inom hälso- och sjukvården och hur deltagarna i grupperna generellt litar på andra människor. Om randomiseringen fungerar väl kommer de två grupperna vara lika på ovan nämnda faktorer.

Manipulationskontroll, även kallat stimuluskontroll, används för att kontrollera om studiens manipulation har uppfattats på det sätt som avses. Stimuluskontroller kan utformas genom att efterfråga om deltagarna minns särskilda detaljer i texten (Esaiasson et al., 2017). Därför ställs frågan "Av vilken anledning sökte Linda vård från början?" med svarsalternativen "Oro och symtom på diabetes", "Oro och symtom på halsfluss" samt "Oro och symtom på pollenallergi". Om en deltagare svarar fel på stimuluskontrollen exkluderas deltagarens data från studien.

3.7 Etiska överväganden

Vid experimentstudier är det viktigt att ta hänsyn till forskningsetiska överväganden och att nyttan för experimentet överväger effekterna av att delta (Esaiasson et al., 2017). För att ta hänsyn till de etiska aspekterna tydliggörs flertalet faktorer innan studien påbörjas. Innan påbörjad studie informeras deltagarna om att studien är frivillig och anonym, samt att det är tillåtet att avbryta deltagandet när som under studiens gång. Deltagarna informeras kort om studiens syfte utan att explicit avslöja moment som kan påverka behandlingen. För att säkerställa att deltagarna tagit del av informationen och att de frivilligt väljer att delta, behöver deltagarna ge sitt medgivande genom att markera "Jag godkänner" på enkätens första sida. Efter avslutad studie förklaras studiens syfte mer utförligt, samt att studien var av experimentell karaktär där samtliga deltagare blev slumpmässigt indelade till antingen kontroll eller behandlingsgrupp. Deltagarna informeras både innan och efter avslutad studie att behandlingen (texten) är fiktivt skriven. För mer information se Appendix. Vidare är avsikten med studien att bidra med kunskap om den allmänna opinionens förtroende/uppfattning för användning av AI till både forskningsfältet och samhällsdebatten. På så sätt är det noga övervägt att ge större nytta än vad det innebär att delta.

En tredje etisk aspekt jag tar hänsyn till är att behandlingen inte ska skapa oro eller rädsla. Min avsikt är att på empirisk väg undersöka om transparens i samband med användning av AI

påverkar samhällets förtroende. Jag avser inte att sprida negativa perspektiv mot användandet av AI inom hälso- och sjukvården, och därför har jag valt att behandlingen mynnar ut i något positivt.

4. Datamaterial

Studiens datainsamling resulterade i 201 deltagare. Totalt tillfrågades 303 personer, varav 45 personer avstod, 56 tackade ja men slutförde ej studien och en deltagare svarade ”fel” på stimuluskontrollen, och därav användes ej svaret. Det resulterade i 100 deltagare i gruppen ”AI-stöd med hög transparens” och 101 deltagare i gruppen ”AI-stöd med låg transparens”. Åldersintervallet i det totala urvalet går från 1942 till 2006 och det genomsnittliga födelseåret i gruppen ”AI-stöd med hög transparens” var 1985 och det genomsnittliga födelseåret i gruppen ”AI-stöd med låg transparens” var 1986. Datamaterialet samlades in från 16/4–2024 – 28/4–2024.

Tabell 1. Könsfördelningen för studiens totala urval och respektive grupp.

	Kvinna	***Man***	***Annat***
AI-stöd hög transparens	60	39	1
AI-stöd låg transparens	55	45	1
Totalt	115	84	2

Tabell 1 visar könsfördelningen för studiens totala urval och inom studiens grupper. Gruppen ”AI-stöd med hög transparens” bestod av 60 kvinnor, 39 män och en som identifierade sig som annat. Fördelningen i gruppen ”AI-stöd med låg transparens” var 55 kvinnor, 45 män och en annat. Totalt deltog 115 kvinnor, 84 män och 2 som identifierade sig som annat.

4.1 Randomiseringskontroll

Tabell 2. Randomiseringskontroll

	Kön	***Ålder***	***Utbildning***	***Förtroende för andra människor***	***Tidigare kunskap***
AI-stöd med hög transparens	0.41	1985	3.75	3.14	0.61
AI-stöd med låg transparens	0.47	1986	3.65	3.20	0.60
P-värde	0.45	0.62	0.59	0.41	0.84
Antal observationer	201	201	201	201	201

Kommentar tabell 2: Tabellen visar gruppernas medelvärde för respektive faktor (kön, ålder, utbildningsnivå, förtroende, tidigare kunskap). Kön representerar 0 = kvinna, 1 = man, 2 = annat. Ålder visas i födelseår. Utbildningsnivå mäts på en 6-gradig skala. Förtroende för andra människor representerar frågan "Generellt sätt, i vilken utsträckning litar du på andra människor" och mäts på en 4-gradig skala. Tidigare kunskap utgörs av en dummyvariabel 0 = har inte tidigare kunskap, 1 = har tidigare kunskap. P-värdet visar signifikansnivån för medelvärdeskillnaden mellan grupperna för respektive faktor.

Randomiseringskontrollen visar att gruppernas medelvärden gällande faktorerna *kön*, *ålder, utbildning, generellt förtroende, och tidigare kunskap* inte statistiskt signifikant skiljer sig åt på en 5% alfanivå. Gruppernas medelvärden för samtliga faktorer är väldigt lik. Exempelvis är medelvärdet för "kön" väldigt lik mellan grupperna (0.43 för "AI-stöd med hög transparens" och 0.45 för "AI-stöd med låg transparens"), vilket betyder att fördelningen mellan kvinnor, män och annat är ungefär densamma inom grupperna. Tabellen visar även att det genomsnittliga födelseåret är ungefär densamma mellan grupperna. Generellt sätt erhåller grupperna relativt högt förtroende för andra människor (3.14 i "AI-stöd med hög transparens och 3.21 "AI-stöd med låg transparens") och därför tolkas resultatet som att deltagarna *ofta* litar på andra människor. På liknande sätt erhåller grupperna liknande medelvärdesnivåer gällande tidigare kunskap för AI inom hälso- och sjukvårdssektorn. Sammantaget visar randomiseringskontrollen att grupperna inte på förhand skiljer sig åt och därmed tolkas det som

att randomiseringen lyckades. Det finns därför ingen anledning att tro att det är dessa faktorer påverkar studiens resultat.

4.2 Urvalets representativitet

För att avgöra huruvida studiens resultat kan generaliseras till Sveriges befolkning jämförs könsfördelningen och den generella utbildningsnivån mellan studiens urval och Sveriges befolkning. Könsfördelningen i Sveriges befolkning är ca 49,7% kvinnor och 50,3% män (Statistiskmydnigheten, 2024). Tabell 1 visar att studiens urval inkluderar 57,2% och män 41,8%. Därmed är urvalet inte representativt för könsfördelningen för Sveriges befolkning. Utbildningsnivån (enligt tabell 2) uppnår något över hälften på den sex-gradiga skalan, då gruppen ”AI-stöd med hög transparens” erhåller medelvärdet 3.75 medan gruppen ”AI-stöd med låg transparens” erhåller medelvärdet på 3.65. Enligt Statistikmyndigheten (2020) har 45% av Sveriges befolkning i åldrarna 25–64 år studerat efter-gymnasialnivå. Det innebär att andelen som studerat efter-gymnasial-nivå i min studie är högre jämfört mot Sveriges befolkning, och därmed är urvalet inte representativt. En förklaring till att det skiljer sig mellan urvalet och nationellt kan bero på att andelen högutbildade tenderar att vara högre i storstadsområden (Statistikmyndigheten, 2020). Således bör det tas i beaktning vid generalisering av studiens resultat att det inte är helt representativt för den Sveriges befolkning. Däremot är det viktigaste för studiens resultat är att studiens grupper inte skiljer sig åt, vilket randomiseringskontrollen visar att de inte gör.

4.3 Indexkonstruktioner

Som nämnt i avsnitt 4.5 skapas ett index för samtliga variabler som avses mätas. Indexen är relativt små eftersom endast två variabler ingår i varje index. Frågorna som i enkäten avser att mäta samma fenomen konverteras till ett index och varje fråga utgör lika stor andel av indexet. Således konverteras frågorna gällande transparens till ett *transparensindex*, förtroende till ett *förtroendeindex*, ansvarsutkrävande till ett *ansvarsutkrävandeindex* och etikalitet till ett *etikalitetsindex*.

Vid indexkonstruktioner används det sammanfattade måttet Cronbachs alfa för att studera konstruktionernas reliabilitet (Esaiasson et al., 2017). Cronbachs alfa varierar mellan 0 – 1, där höga alfavärden indikerar på hög reliabilitet mellan variablerna som ingår i indexet, därmed att variablerna mäter samma sak. Enligt Esaiasson et al., (2017) beskrivs alfavärdet 0.7 vara en

tumregel för vad som anses god reliabilitet. Nedan redovisas Cronbachs alfa för samtliga indexkonstruktioner.

Tabell 3. Cronbachs alfa för studiens indexkonstruktioner.

Variabel	***Namn på index***	***Cronbachs alfa***
Transparens	*Transparensindex*	0.69
Förtroende	*Förtroendeindex*	0.74
Ansvarsutkrävande	*Ansvarsutkrävandeindex*	0.68
Etikalitet	*Etikalitetsindex*	0.61

Tabell 5 visar Cronbachs alfa för studiens indexkonstruktioner. Tabellen visar att förtroendeindexet (alfavärde 0.74) är det enda index som når över gränsen för tumregeln (0.7) för god reliabilitet. Därav tolkas förtroendeindexet erhålla hög reliabilitet och indikerar på att indexets variabler mäter samma sak. Transparensindexet och ansvarsutkrävandeindexet når inte upp till alfavärdet 0.7, men båda indexen erhåller värden nära (0.69 samt 0.68). Reliabiliteten för transparensindexet och ansvarsutkrävandeindexet tolkas därmed vara relativt högt, men det ska ändå belysas att värdena inte når alfagränsen som vanligtvis accepteras. Det sista indexet, etikalitetsindexet uppnår ett alfavärde på 0.61, och erhåller därmed det lägsta alfavärdet av indexkonstruktionerna. Reliabiliteten för etikalitetsindexet kan därmed tolkas som ”måttlig”. I och med att studiens huvudsakliga fokus är förtroende, är det positivt att förtroendeindexet når hög reliabilitet.

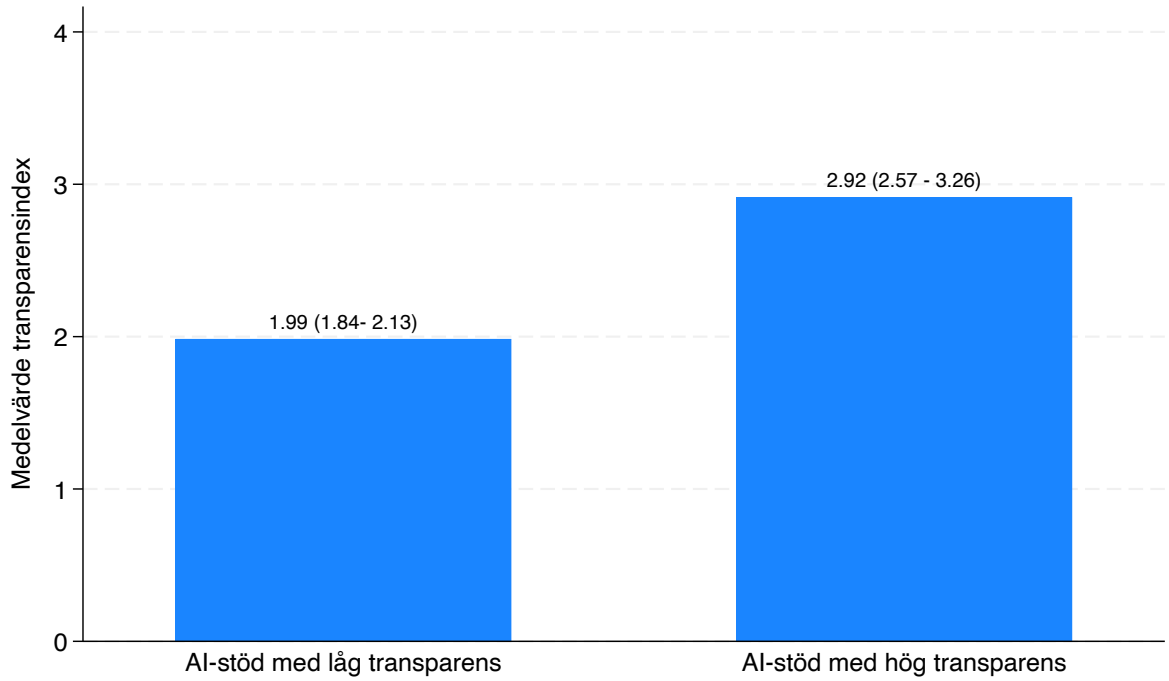
5. Resultat

I kapitel fem redogörs studiens resultat och analys av insamlat datamaterial. Datamaterialet studeras och redovisas genom diagram samt regressionsanalyser. För att resultaten ska ses statistiskt signifikanta ska regressionsanalysernas p-värden erhålla lägre signifikansvärden än alfavärdet 0.05. Min avsikt är att studera om variationen av transparens har orsakat någon skillnad mellan grupperna gällande uppfattad transparens och förtroende för användning av AI som beslutsstöd inom hälso- och sjukvården.

5.1 Resultat och analys för studiens första hypotes

Studiens första resultat studerar studiens första hypotes (H1). Studiens första hypotes är enligt följande: Personer som läser beslutet med ***högre transparens*** i beslutsfattandeprocessen kommer att uppfatta transparensen i beslutet som ***högre***. För att undersöka hypotesen jämförs medelvärdet för respektive grupp gällande transparensindexet i figur 3.

Figur 3. Stapeldiagram för gruppernas medelvärden gällande transparensindex.



Kommentar figur 3: X-axeln visar studiens grupper. Y – axeln visar medelvärdet av transparensindexet (skala 1–4). Konfidensintervall inom parentes.

Figur 3 visar en tydlig medelvärdeskillnad mellan grupperna med avseende på uppfattad transparens (transparensindexet). Gruppen ”AI-stöd med hög transparens” erhåller ett medelvärde för transparensindexet på 2,92 medan gruppen ”AI-stöd med låg transparens” uppnår ett medelvärde på 1,99. Resultatet tolkas att gruppen ”AI-stöd med hög transparens” har uppfattat transparensen som högre än gruppen ”AI-stöd med låg transparens”. För att studera om medelvärdeskillnaden är statistiskt signifikant, skilt från noll genomförs en regressionsanalys vilket presenteras i tabell 4.

Tabell 4: Regressionsanalys, beroende variabel: *transparensindex,* standardfel inom parentes.

	Transparensindex
AI-stöd med hög transparens	0,930 *** (0.10)
Konstant	1,985 *** (0.07)
Antal observationer	201
R^2	0,296

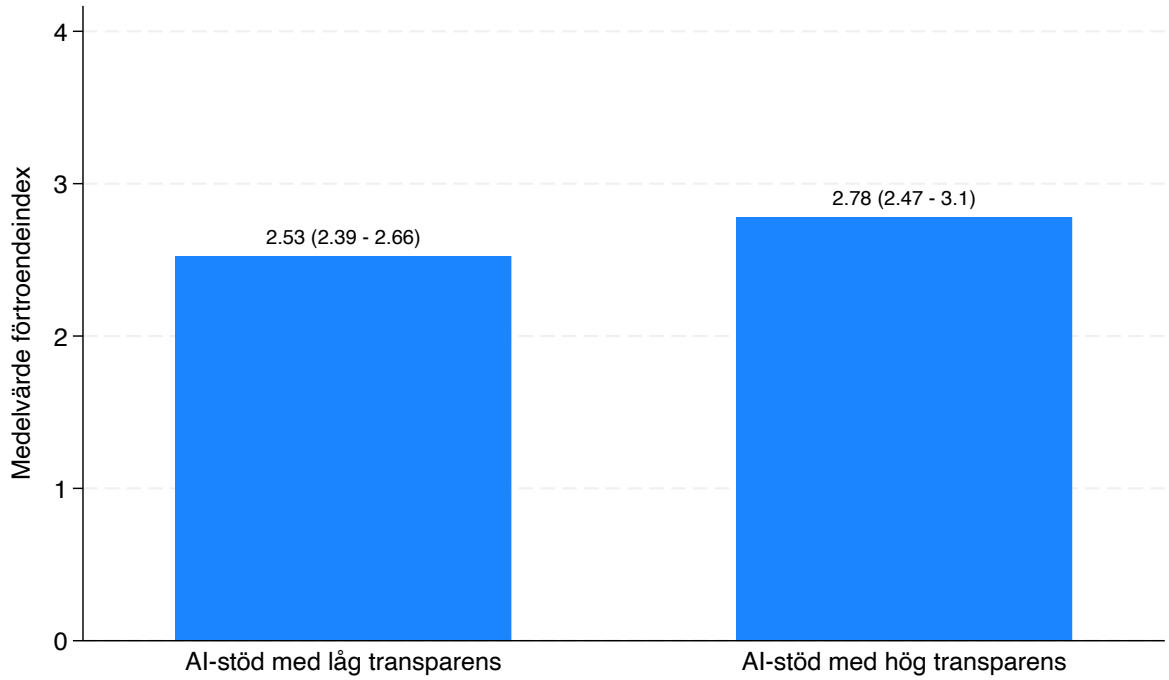
Kommentar tabell 4: P-värden visas med signifikanstjärnor,* $p < 0.05$, ** $p < 0.01$,*** $p < 0.001$

Regressionsanalysen i tabell 4 visar positiva estimat för β -koefficienten vilket indikerar på ett positivt samband mellan den oberoende variabeln (grupperna) och transparensindexet. β-värdet uppnår 0,930 vilket betyder att gruppen ”AI-stöd med hög transparens” i genomsnitt erhöll 0,930 poäng högre transparensindex än gruppen ”AI-stöd med låg transparens”. Konstanten visar gruppen ”AI-stöd med låg transparens” (dvs jämförelsegruppens) medelvärde. P-värdet kan utläsas till lägre än 0,001 och därmed är risken att resultatet beror på slump lägre än 0,1%. Den bestämda alfa-nivån är 0.05, vilket innebär att det erhållna p-värdet (0,001) är lägre än 0,05, och således kan nollhypotesen för hypotes 1 förkastas. Resultatet från regressionsanalysen visar att determinationskoefficienten är 0,296 vilket kan tolkas som att 29,6 % av variansen i den beroende variabeln förklaras av den oberoende variabeln. Resultatet är i linje med hypotes 1, och statistiskt signifikant på 5% nivå. Därför tolkas resultatet som att gruppen ”AI-stöd med hög transparens” uppfattade transparensen i texten högre än gruppen ”AI-stöd med låg transparens”.

5.2 Resultat och analys för studiens andra hypotes

Studiens andra resultat avser att undersöka studiens andra hypotes (H2): ”Personer som läser beslutet med ***högre transparens*** i beslutsfattandeprocessen kommer ha ett ***högre förtroende*** för användning av AI som beslutsstöd inom hälso- och sjukvården”. För att studera hypotesen redovisas nedan medelvärdet för respektive grupp gällande förtroendeindexet.

Figur 4. Stapeldiagram för gruppernas medelvärde gällande förtroendeindexet.



Kommentar figur 4: X-axeln visar studiens grupper. Y – axeln visar medelvärdet av förtroendeindexet (skala 1–4). Konfidensintervall inom parentes.

Figur 4 visar att gruppernas medelvärden skiljer sig åt gällande förtroendeindexet. Gruppen ”AI-stöd med hög transparens” erhåller ett högre medelvärde (2.78) jämfört med gruppen ”AI-stöd med låg transparens” som erhåller ett medelvärde på 2.53. Nedan presenteras regressionsanalysen för att studera om skillnaden är statistiskt signifikant.

Tabell 5: Regressionsanalys, beroende variabel: *förtroendeindex,* standardfel inom parentes.

	Förtroendeindex
AI-stöd med hög transparens	0,255** (0.10)
Konstant	2,525*** (0.07)
Antal observationer	201
R^2	0,035

*Kommentar tabell 5: P-värden visas med signifikiansstjärnor, * p < 0.05, ** p < 0.01, *** p < 0,001*

Regressionsanalysen visar positiva estimat för β -koefficienten vilket indikerar på ett positivt samband mellan den oberoende variabeln (grupperna) och förtroendeindexet. β-värdet uppnår 0,255 vilket innebär att gruppen ”AI-stöd med hög transparens” i genomsnitt uppfattade förtroendeindexet 0,255 poäng högre än gruppen ”AI-stöd med låg transparens”. Konstanten visar medelvärdet för gruppen ”AI-stöd med låg transparens” (det vill säga jämförelsegruppen). Sambandets determinationskoefficient är 0,035 vilket kan tolkas som att den oberoende variabeln förklarar 3,5% av den beroende variabeln. Resultatet erhåller ett p-värde lägre än 0.01 vilket betyder att risken att resultatet beror på slump är lägre än 1%. P-värdet värdet är därför lägre än signifikansnivån 0.05, och därmed kan nollhypotesen förkastas. Således får hypotes två stöd och är statistiskt signifikant på en 5% -nivå.

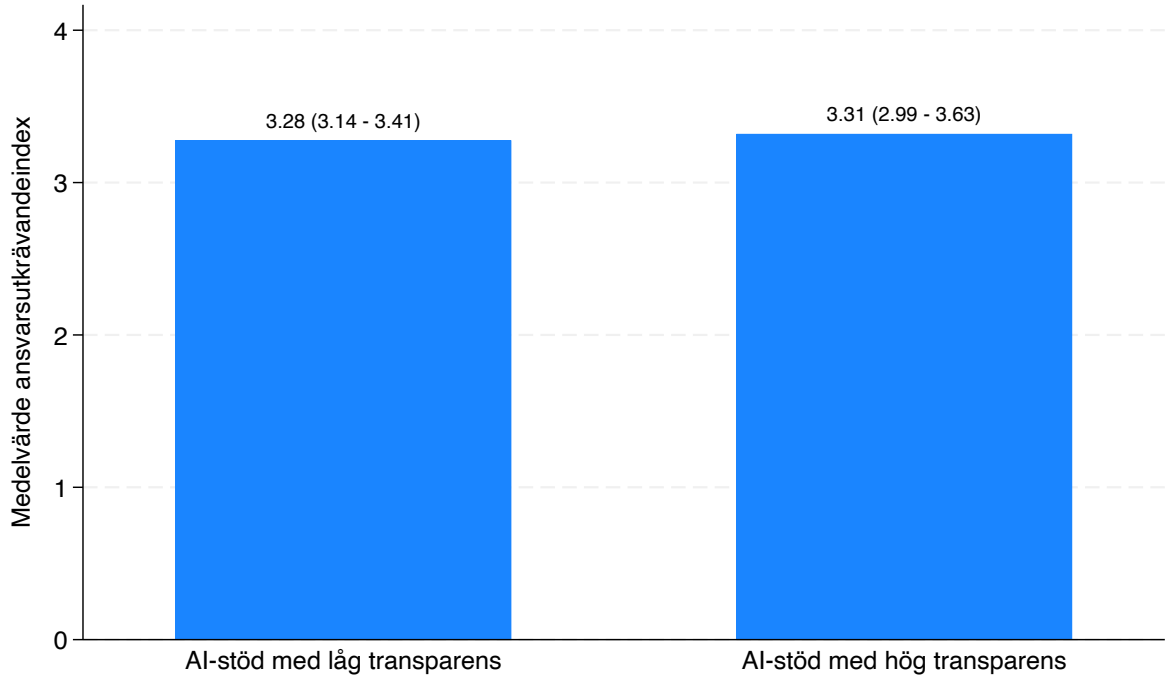
Sammanfattningsvis tolkas resultatet från de två regressionsanalyserna vara i linje med studiens hypoteser och båda hypoteserna får därför stöd. Resultatet tolkas som att deltagarna som läste beslutet med högre transparens i beslutsfattandeprocessen uppfattade transparensen som högre och gruppens deltagare erhöll ett högre förtroende för användning av AI som beslutsstöd.

5.3 Potentiella implikationer till studiens resultat

Trots att hypoteserna får stöd är det av betydelse att analysera de två teoretiskt identifierade faktorerna *ansvarsutkrävande* och *etikalitet*. Min avsikt är att endast manipulera transparens, och att ansvarsutkrävandet samt etikaliteten, hålls och uppfattas konstant mellan grupperna. Det finns en risk för att både ansvarsutkrävande och etikalitet manipuleras i samband med att

transparens manipuleras, det vill säga uppfattas på varierande sätt. Om så är fallet så finns det en risk för att det är de faktorerna som driver effekten på förtroendet.

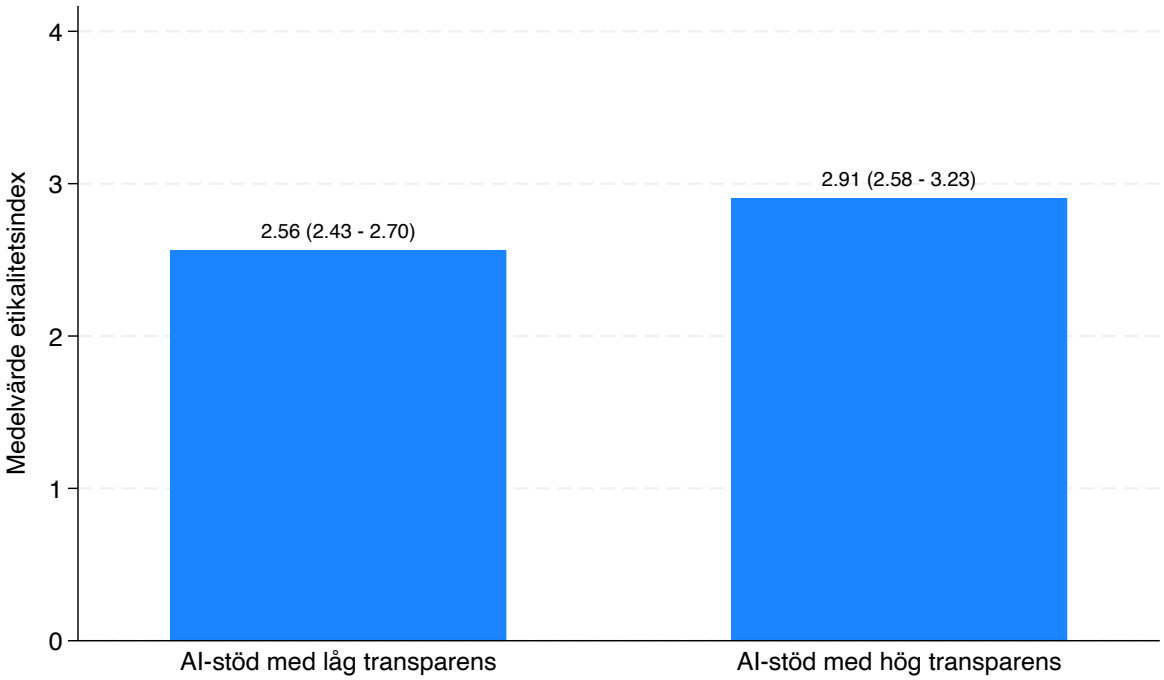
Figur 5. Stapeldiagram för gruppernas medelvärden gällande ansvarsutkrävandeindex.



Kommentar figur 5 : X-axeln visar studiens grupper. Y – axeln visar medelvärdet av ansvarsutkrävandeindexet (skala 1–4). Konfidensintervall inom parentes.

Figur 5 visar en liten medelvärdesskillnad mellan grupperna gällande *ansvarsutkrävandeindexet*. Gruppen ”AI-stöd med hög transparens” erhåller ett medelvärde på 3.31 medan gruppen ”AI-stöd med låg transparens” når ett medelvärde på 3,28. Skillnaden kontrolleras genom regressionsanalys som presenteras i tabell 6.

Figur 6. Stapeldiagram för gruppernas medelvärden gällande etikalitetsindexet.



Kommentar figur 6 : X-axeln visar studiens grupper. Y – axeln visar medelvärdet av etikalitetsindexet (skala 1–4). Konfidensintervall inom parentes.

Figur 6 visar en relativt tydlig medelvärdeskillnad mellan grupperna gällande *etikalitetsindexet*. Gruppen ”AI-stöd med hög transparens” uppnår medelvärdet 2,91 medan gruppen ”AI-stöd med låg transparens” når medelvärdet 2,56. Regressionsanalysen för medelvärdeskillnaden presenteras i tabell 6.

Tabell 6: Regressionsanalys med beroende variabler: *ansvarsutkrävandeindex* och *etikalitetsindex,* standardfel inom parentes.

	Ansvarsutkrävandeindex	**Etikalitetsindex**
AI-stöd med hög transparens	0,033 (0.10)	0,341*** (0.10)
Konstant	3,277*** (0.07)	2,563*** (0.06)
Antal observationer	201	201
R^2	0,001	0,060

Kommentar: P-värden visas med signifikanstjärnor, * $p < 0.05$, ** $p < 0.01$, *** $p < 0{,}001$

Tabell 6 visar regressionsanalyserna för faktorerna ansvarsutkrävande och etikalitet. Nedan redogörs först resultatet för ansvarsutkrävande och därefter etikalitetsindexet.

I tabell 6 visar regressionsanalysen för ansvarsutkrävandeindexet ett β- värde på 0,033 vilket betyder att gruppen ”AI-stöd med hög transparens” i genomsnitt uppfattade ansvarsutkrävandet 0,033 poäng högre än jämförelsegruppen. Avsaknaden av signifikanstjärnor innebär att p-värdet inte är lägre än 0.05. Därmed är skillnaden mellan grupperna inte statistiskt signifikant.

Tabell 6 visar även regressionsanalysen för etikalitetsindexet. Regressionsanalysen visar ett β-värde på 0,341. Resultatet tolkas därför som att gruppen ”AI-stöd med hög transparens” uppfattade etikaliteten i genomsnitt 0,341 poäng högre än gruppen ”AI-stöd med låg transparens”. Determinationskoefficienten är 0,060 vilket tolkas som att den oberoende variabeln förklarar 6 % av den beroende variabeln. P-värdet är lägre än 0,001 och därmed lägre än 0.05, och medelvärdeskillnaden kan tolkas statistiskt signifikant.

Resultatet från tabell 6 visar således att regressionsanalysen gällande ansvarsutkrävandeindexet inte är statistiskt signifikant, men att regressionsanalysen för etikalitetsindexet är statistiskt signifikant. Därför tolkas resultatet som att det inte finns någon skillnad hur grupperna uppfattade ansvarsutkrävandet, men att grupperna skiljer sig åt gällande uppfattad etikalitet. I och med att uppfattningen av ansvarsutkrävande inte skiljer sig åt kan slutsatsen dras att det

inte finns någon anledning att tro att ansvarsutkrävandet har drivit effekten för grundsambandet (Hypotes 2), och kan därför ”uteslutas”. Gruppen ”AI-stöd med hög transparens” uppfattade däremot etikaliteten som högre än jämförelsegruppen. Det innebär att det finns en risk att etikaliteten driver effekten för grundsambandet (Hypotes 2), eller på något sätt involveras.

5.4 Test för eventuella samspelsvariabler

För att studera eventuella samspelsvariationer i urvalet genomförs interaktionsanalyser för sambandet mellan transparens och förtroende, med interaktion av kön, ålder och utbildningsnivå. Om resultaten visar statistiskt signifikanta skillnader tolkas det som att det finns heterogena effekter i urvalet, det vill säga att sambandet varierar i undergrupper. Om så är fallet så påverkas studiens generaliserbarhet.

Regressionstabellerna för interaktionsanalyserna visas i Appendix. Resultaten visar inga statistiskt signifikanta interaktionseffekter gällande, kön, ålder eller utbildning. Därför tolkas resultatet som att det inte finns heterogena effekter i urvalet.

6. Diskussion och slutsats

I denna studie har jag genom en experimentell metod studerat om allmänhetens förtroende för användning av AI som beslutsstöd inom hälso- och sjukvården påverkas av graden av transparens i beslutsfattandeprocessen. Totalt deltog 201 personer och datamaterialet samlades in 16/4–2024 – 28/4–2024. Tidigare forskning har fokuserat på att jämföra människors uppfattning av AI-genererade beslut, mot beslut som tagits av människor, vilket generellt visat att människors beslut föredras (Rojahn et al., 2023; Lee et al., 2018; Bankins et al., 2022). I denna studie fokuserar jag i stället på att jämföra två AI-genererade beslut, där den enda skillnaden mellan besluten är graden av transparens. Syftet är att förstå om allmänhetens förtroende för AI som beslutsstöd inom hälso- och sjukvården påverkas graden av transparens i beslutsfattandeprocessen, det vill säga hur problematiken med transparensen hanteras.

Studiens resultat ger stöd för studiens hypoteser. Därför tolkas resultatet som att deltagarna i gruppen ”AI-stöd med hög transparens”, det vill säga ”behandlingsgruppen” som tog del av texten med högre transparens i beslutsfattandeprocessen, uppfattande transparensen som högre. Resultatet tolkas även som att gruppen ”AI-stöd med hög transparens” erhöll ett högre förtroende för användning av AI som beslutsstöd.

Resultatet är i linje med tidigare forskning och teori för sambandet mellan ökad transparens och ökat förtroende (Hood & Heald, 2006; De Fine Licht et al., 2014; De Fine Licht, 2014). Enligt Hood & Heald (2006) kan förklaringar till hur och varför ett beslut tas, öka mottagarnas förtroende för beslutet, vilket kan tänkas vara en förklaring för denna studies resultat. Rojahn et al., (2023) belyser att en tänkbar förklaring till varför människor inte litar på AI-teknik är för att vi inte förstår. Tillämpat på denna studie kan det innebära att deltagarna som läste texten med låg transparens inte förstod hur eller varför beslutet togs och därav erhöll lägre förtroende för tekniken jämfört med gruppen ”AI-stöd med hög transparens” som kunde förstå hur och varför beslutet togs, och kunde därför erhålla ett högre förtroende.

Gällande uppfattningen av ansvarsutkrävande identifierades ingen skillnad gällande mellan grupperna, men etikaliteten uppfattades som högre i gruppen ”AI-stöd med hög transparens”. Resultatet och sambandet för transparensens effekter på förtroendet (hypotes 2), försummas inte trots att etikaliteten visar en statistiskt signifikant skillnad mellan grupperna. Däremot försvåras möjligheterna för att dra slutsatser om hur den teoretiska modellen ser ut, eftersom etikalitet bevisligen påverkar på något eller flera sätt. Det är intressant att etikaliteten uppfattades olika eftersom gruppernas texter är identiska bortsett från variationen av transparens. Har deltagarna tagit egna slutsatser gällande etikalitet, eller finns det ett samband mellan uppfattad transparens och etikalitet, som påverkar förtroendet för användning av AI? Kan transparens och etikalitet särskiljas? Enligt De Fine Licht (2014) kan medborgare endast bedöma ett besluts rättvisa om de är medvetna om proceduren. Kan det vara en förklaring till varför deltagarna i min studie uppfattade etikaliteten på varierande sätt? Det vill säga att gruppen ”AI-stöd med hög transparens” fick insyn i proceduren och kunde på så sätt bedöma procedurens rättvisa (etikaliteten), medan gruppen ”AI-stöd med låg transparens” inte fick möjlighet att bedöma proceduren i och med avsaknaden av transparens.

O’Neill (2002) argumenterar för att ökad insyn i komplexa beslutsfattandeprocesser skapar förvirring snarare än förtroende. På liknande sätt har De Fine Licht et al., (2014) identifierat att mekanismerna frustration, maktlöshet samt missnöje kan påverka sambandet mellan transparens och förtroende negativt. Trots att resultatet i min studie visar ett positivt samband mellan transparens och förtroende, motbevisas inte O’Neill (2002) och De Fine Licht et al., (2014) argument. Författarnas argument ger upphov till frågor gällande i vilken utsträckning AI genererade beslut kan och bör vara transparenta för att skapa förtroende snarare än förvirring.

Onekligen erhåller studien och studiens metod brister. Problematiken med generalisering och extern validitet har diskuterats i uppsatsen, och är den experimentella metodens såväl som denna studies brister. Andelen högutbildade är högre i studiens urval jämfört med Sveriges befolkning vilket gör att resultatet inte är generaliserbart för den större befolkningen. Det är svårt att avgöra vilken population och vilka kontexter resultatet är i linje med, och resultatet bör därför generaliseras med försiktighet till andra kontexter. Den experimentella metoden kritiseras för att studera förenklade bilder av verkligheten (Esaiasson et al., 2017). I studien fokuserar jag på transparens. Som visat är exempelvis etikalitet också involverat för hur förtroendet påverkas. Det belyser att ja, transparens är viktigt, men det förklarar inte allt och troligen är sambandet mer komplext än vad studiens resultat implicerar på.

Den tydliga fördelen med den experimentella metoden är den interna validiteten för studiens resultat. I och med att randomiseringen fungerande väl kan kausal interferens beläggas. Det betyder att resultatet kan tolkas som att det finns ett positivt samband mellan högre transparens i beslutsfattandeprocessen vid användning av AI och allmänhetensförtroende för användning av tekniken som beslutsstöd inom hälso- och sjukvårdssektorn. Implikationen är således att transparent användning av AI är viktigt att beakta när AI introduceras som beslutsstöd inom hälso- och sjukvården för att människor ska erhålla ett förtroende för både AI-användningen och hälso- och sjukvården som institution.

Så, påverkas allmänhetens förtroende för användning av AI som beslutsstöd inom hälso- och sjukvården, beroende på hur transparent beslutsfattandeprocessen är? Svaret på forskningsfrågan är, ja transparens påverkar allmänhetens förtroende för användning av AI som beslutsstöd inom hälso- och sjukvården. Deltagarna som tog del av behandlingen med högre transparens erhöll ett högre förtroende för användning av AI som beslutsstöd inom hälso- och sjukvården. Därmed är transparens en faktor som bör beaktas när AI introduceras inom sektorn för att samhällets medborgare ska erhålla ett förtroende för AI-tekniken och för hälso- och sjukvården som institution. Det institutionella förtroendet är avgörande för hälso- och sjukvården funktionalitet och effektivitet samt för att människor ska acceptera nya interventioner (Gilson, 2003).

Resultatet visar att transparens är en faktor som bör beaktas vid implementering av AI inom hälso- och sjukvården för att behålla samhällets förtroende för institutionen, men mer forskning behövs för att belysa andra potentiellt påverkande faktorer. I och med att studiens båda grupper tog del av en behandling med antingen hög eller låg transparens är det svårt att avgöra vad av

mekanismerna som orsakat resultatet. Det vill säga om det var den höga transparensen som ökade förtroendet för AI-verktygets beslut, eller om det var den låga nivån av transparens som minskade förtroendet för AI-verktygets beslut. riktningen för sambandet, det vill säga om ökad transparens ökar eller om minskad transparens minskar förtroendet. Ett andra förslag för framtida forskning är att genom en kumulativ experimentell design studera i vilken utsträckning beslut bör vara transparenta vid användning av AI inom hälso- och sjukvården för att erhålla så högst möjliga förtroende, eftersom studiens resultat endast jämför avsaknad av transparens mot befintlig transparens. Den kumulativa designen möjliggör att studera hur mängden av information påverkar effekten (Esaiasson et al., 2017). Det tredje förslaget för framtida forskning är att studera etikalitetens involvering. Resultatet gällande etikalitet väcker frågor om etikalitetens roll för sambandet mellan transparens och förtroende, men även frågor gällande sambandet mellan transparens och etikalitet. Det krävs således mer forskning för att förstå om, hur och på vilket sätt mekanismerna och sambanden ser ut.

7. Referenslista

Bankins, Formosa, P., Griep, Y., & Richards, D. (2022). AI Decision Making with Dignity? Contrasting Workers' Justice Perceptions of Human and AI Decision Making in a Human Resource Management Context. *Information Systems Frontiers*, 24(3), 857–875. DOI: 10.1007/s10796-021-10223-8

Birkinshaw, Patrick J. 2006. Transparency as a Human Right. In Transparency: The Key to Better Governance? edited by Christopher Hood and David Heald, 47–58. Oxford, UK: Oxford University Press.

Blader, S.L. & Tyler, T.R., 2003. A Four-Component Model of Procedural Justice: Defining the Meaning of a "Fair" Process. Personality & social psychology bulletin, 29(6), 747–758. DOI: 10.1177/0146167203252811

Buch, V. H., Ahmed, I., & Maruthappu, M. (2018). Artificial intelligence in medicine: current trends and future possibilities. *British Journal of General Practice*, *68*(668), 143-144. DOI: 10.3399/bjgp18X695213

Choudhury, & Asan, O. (2022). Impact of accountability, training, and human factors on the use of artificial intelligence in healthcare: Exploring the perceptions of healthcare practitioners in the US. Human Factors in Healthcare (Online), 2, 100021. DOI: 10.1016/j.hfh.2022.100021

Colquitt, J. A. (2001). On the Dimensionality of Organizational Justice: A Construct Validation of a Measure. *Journal of Applied Psychology*, *86*(3), 386–400. DOI: 10.1037/0021-9010.86.3.386

Davenport, & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future Healthcare Journal, 6(2), 94–98. DOI: 10.7861/futurehosp.6-2-94

De Fine Licht, Naurin, D., Esaiasson, P., & Gilljam, M. (2014). When Does Transparency Generate Legitimacy? Experimenting on a Context-Bound Relationship. Governance (Oxford), 27(1), 111–134. DOI: 10.1111/gove.12021

Emanuel, & Emanuel, L. L. (1996). What is accountability in health care? *Annals of Internal Medicine*, *124*(2), 229–239. DOI: 10.7326/0003-4819-124-2-199601150-00007

Esaiasson, Peter, Gilljam, Mikael, Oscarsson, Henrik, Wängerud, Lena, Towns Ann, Metodpraktikan 5:e upplagan 2017, Norstedts förlag. Lund: Studentlitteratur.

Grimmelikhuijsen, Porumbescu, G., Hong, B., & Im, T. (2013). The Effect of Transparency on Trust in Government: A Cross-National Comparative Experiment. Public Administration Review, 73(4), 575–586. DOI: 10.1111/puar.12047

Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E. & Puranen B. (2002): World Values Survey Wave 7 (2017-2022) Cross-National Data- Set. Version: 4.0.0. World Valyes Survey Association. DOI: 10.14281/18241.18

Haidt, J. (2013). Moral psychology for the twenty-first century. *Journal of Moral Education*, *42*(3), 281-297. DOI: 10.1080/03057240.2013.817327

Heald, David. 2006. Transparency as an instrumental value. I Transparency: The Key to Better Governance? edited by Hood Christopher and Heald David, 58–73. Oxford, UK: Oxford University Press.

Herreros, F. (2023). The state and trust. *Annual Review of Political Science*, *26*, 117-134. DOI: 10.1146/annurev-polisci-051921-102842

Hollyer, J. R., Rosendorff, B. P., & Vreeland, J. R. (2011). Democracy and transparency. *The Journal of Politics*, *73*(4), 1191-1205. DOI: 10.1017/S0022381611000880

Holmström, Bengt. 1979. “Moral Hazard and Observability.” Bell Journal of Economics 10: 74–91. DOI: 10.2307/3003320

Hood, Christopher. 2006. Beyond Exchanging First Principles? Some Closing Comments. I Transparency: The Key to Better Governance? edited by Christopher Hood and David Heald, 211–26. Oxford, UK: Oxford University Press.

Horgan, Romao, M., Morré, S. A., & Kalra, D. (2019). Artificial Intelligence: Power for Civilisation – and for Better Healthcare. Public Health Genomics, 22(5/6), 145–161. DOI: 10.1159/000504785

Hudson, J. (2006). Institutional trust and subjective well-being across the EU. *Kyklos*, *59*(1), 43-62.

Jiang, Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. Stroke and Vascular Neurology, 2(4), 230–243. DOI: 10.1136/svn-2017-000101

Kiseleva, A., Kotzinos, D., & De Hert, P. (2022). Transparency of AI in healthcare as a multilayered system of accountabilities: between legal requirements and technical limitations. Frontiers in Artificial Intelligence, 5, 879603. DOI: 10.3389/frai.2022.879603

Körber, M. (2019). Theoretical Considerations and Development of a Questionnaire to Measure Trust in Automation. I *Advances in Intelligent Systems and Computing* (Vol. 823, pp. 13–30). Springer International Publishing. DOI: 10.1007/978-3-319-96074-6_2

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 205395171875668. DOI: 10.1177/2053951718756684

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human factors, 46(1), 50-80. DOI: 10.1518/hfes.46.1.50_30392

Gilson, L. (2003). Trust and the development of health care as a social institution. *Social science & medicine*, *56*(7), 1453-1468. DOI: 10.1016/S0277-9536(02)00142-9

Läkemedelsverket. (2023). *Vägledning rörande användning av artificiell intelligens i svensk sjukvård*. Läkemedelsverket: Swedish medical products agency. Tillgänglig:

https://www.lakemedelsverket.se/4a5f16/globalassets/dokument/medicinteknik/artificiell-intelligens-ai/vagledning-anvandning-av-artificiell-intelligens-i-svensk-sjukvard.pdf

Nannestad, P. (2008). What have we learned about generalized trust, if anything?. *Annu. Rev. Polit. Sci.*, *11*, 413-436. DOI: 10.1146/annurev.polisci.11.060606.135412

O'Neill, O. 2002. A Question of Trust. The BBC Reith Lectures, 2002. Cambridge: Cambridge University Press.

Robertson, Felicia. (2023). *Vilka litar på våra samhällsinstitutioner och varför*. Tillgänglig: *https://www.gu.se/sites/default/files/2023-03/Vilka%20litar%20p%C3%A5%20v%C3%A5ra%20samh%C3%A4llsinstitutioner%20och%20varf%C3%B6r%20-%20Felicia%20Robertson.pdf*

Rojahn, Palu, A., Skiena, S., & Jones, J. J. (2023). American public opinion on artificial intelligence in healthcare. PloS One, 18(11), e0294028–e0294028. DOI: 10.1371/journal.pone.0294028

Rothstein, B., & Stolle, D. (2003). Social capital, impartiality and the welfare state: An institutional approach. I *Generating social capital: Civil society and institutions in comparative perspective* (s. 191-209). New York: Palgrave Macmillan US.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of management review*, *23*(3), 393-404. DOI: 10.5465/amr.1998.926617

Statistikmyndigheten. (2020). *Högst utbildningsnivå bland kvinnor och i våra universitetsstäder*. Hämtad: 2024 – 01 – 05, från https://www.scb.se/hitta-statistik/statistik-efter-amne/utbildning-och-forskning/befolkningens-utbildning/befolkningens-utbildning/pong/statistiknyhet/befolkningens-utbildning-2019/

Statistikmyndigheten. (2024). *Sveriges befolkning*. Hämtad: 2024 – 01 – 05, från https://www.scb.se/hitta-statistik/sverige-i-siffror/manniskorna-i-sverige/sveriges-befolkning/

Sundell, Anders. (2012, 11 november). *Guide: Statistisk "power" och urvalsstorlek vid experimentell desgin [Blogginlägg]*. Hämtad från: https://spssakuten.com/2012/11/11/guide-statistisk-power-och-urvalsstorlek-i-experimentell-design/

Sztompka. (1998). Trust, Distrust and Two Paradoxes of Democracy. European Journal of Social Theory, 1(1), 19–32. DOI: 10.1177/136843198001001003

Sønderskov, K. M., & Dinesen, P. T. (2016). Trusting the state, trusting each other? The effect of institutional trust on social trust. *Political Behavior*, *38*, 179-202. DOI: 10.1007/s11109-015-9322-8

Tyler, Tom R. 2006. "Psychological Perspectives on Legitimacy and Legitimation." *Annual Review of Psychology* 57: 375- 400 DOI: 10.1146/annurev.psych.57.102904.190038

Warren. (2018). Trust and Democracy. In The Oxford Handbook of Social and Political Trust (pp. 75–94). Oxford University Press. DOI: 10.1093/oxfordhb/9780190274801.013.5 .

Zmerli, S., & Newton, K. (2007). Trust in people, confidence in political institutions, and satisfaction with democracy. I *Citizenship and involvement in European democracies* (pp. 59-89). Routledge.

8. Appendix

8.1 Enkät

Enkätundersökning för kandidatarbete vid Göteborgs Universitet

Hej!

Denna undersökning genomförs i samband med mitt examensarbete på Statsvetarprogrammet på Göteborgs universitet. Syftet med studien är att undersöka upplevelser inom vården.

Deltagande i studien är helt frivilligt och studien genomförs anonymt. Det betyder att varken jag eller någon annan kan spåra dina svar tillbaka till dig. Du kan när som helst välja att avbryta din medverkan i studien om du längre inte vill delta. Om du väljer att avbryta deltagandet kommer jag inte använda dina svar. Svaren från de enkäter som fullföljs kommer att användas som data i min kandidatuppsats.

Med ditt deltagande får du först ta del av en kort fiktiv text för att sedan besvara en enkät med ett antal frågor. Deltagande i studien tar inte mer än cirka 5 minuter.

Genom att fylla i rutan nedan "Jag godkänner" anger du att du har tagit del av informationen ovan och att du frivilligt väljer att delta i studien. Om du har några frågor om studien går det går jättebra att kontakta mig via min mejl: Gussvejeaz@student.gu.se

Genom att klicka i rutan nedan godkänner du att du frivilligt deltar i studien:

- o Jag godkänner

Artificiell intelligens (AI) används till viss del inom vården idag och förväntas att användas mer i framtiden. Har du sedan tidigare kunskap om att AI används inom vården alternativt kommer att användas i framtiden?

- o Ja
- o Nej

I texten på nästa sida får du ta del av en något omskriven och nedkortad nyhetsartikel. Texten behandlar det ämne som presenterades i den tidigare frågan, det vill säga "användning av artificiell intelligens (AI) i vården". Läs först texten och besvara därefter de tillhörande frågorna i tur och ordning.

I vilken utsträckning instämmer du med följande påståenden?

(För att komma vidare till nästa fråga måste du besvara samtliga frågor).

	Instämmer helt	Instämmer delvis	Instämmer knappast	Instämmer inte alls
AI-verktygets beslut är av god kvalitet	o	o	o	o
Jag litar på att AI-verktyget har ställt rätt diagnos	o	o	o	o
Vården följer upp ärendet ansvarsfullt och tar fullt ansvar för eventuella konsekvenser av beslutet	o	o	o	o
Vården tar ansvar för sitt beslut	o	o	o	o
Förklaringen som ges till beslutet är bristfällig	o	o	o	o
Beslutsprocessen är transparent	o	o	o	o
Beslutet fattas på ett etiskt och moraliskt sätt	o	o	o	o

Har AI-verktyget behandlat Linda med värdighet och respekt?

- o Instämmer helt
- o Instämmer delvis
- o Instämmer knappast
- o Instämmer inte alls

Enligt din mening, i vilken utsträckning går det att lita på människor i allmänhet?

- o Det går att lita på människor i allmänhet
- o Jag litar ganska ofta på andra människor
- o Det går sällan att lita på andra på andra människor
- o Det går inte att lita på andra människor

I texten du läste, av vilken anledning sökte Linda vård från första början?

- o Oro och symtom på halsfluss
- o Oro och symtom på diabetes
- o Oro och symtom för pollenallergi

Tre avslutande frågor

Vilket år är du född?

_____

Kön?

- o Kvinna
- o Man
- o Annat

Vilken utbildningsnivå har du avslutat?

- o Ej avslutad grundskola
- o Grundskola
- o Gymnasieutbildning
- o Yrkeshögskola
- o Högskola/Universitetsstudier (Kandidatexamen)
- o Högskola/Universitetsstudier, längre än 3 år
- o Examen från forskarutbildning

Tack för att du deltog i min studie!

Syftet med studien handlar mer specifikt om att jämföra om tilliten till användningen av artificiell intelligens (AI) i vården påverkas av om tillämpningen genomförs med hög eller låg transparens. Studien du deltog i är en experimentell studie där texten som presenterades är en fiktiv text. Du blev slumpmässigt indelad till antingen behandlingsgrupp eller kontrollgrupp.

Har du frågor om min undersökning är du välkommen att hör av dig till mig mejl; Gussvejeaz@student.gu.se

Tack!

8.2 Behandling

8.2.1 Text som läses av ”Behandlingsgrupp” (*Användning av AI som beslutstöd + hög transparens)*

AI-verktyg gav rätt diagnos i vården: Med tydliga förklaringar och öppna beslut

Linda, 37, sökte vård då hon oroade sig för symtom på diabetes. Linda hoppades att ett vårdbesök skulle kunna ge tydliga svar och lugna hennes oro. Vid vårdbesöket kartlades Lindas tillstånd, bakgrund och risk att utveckla diabetes. Det togs också blodprover.

Vid bedömningen av Lindas tillstånd användes ett AI-verktyg som beslutsstöd till läkaren. AI-verktyget visade inga indikationer på avvikelser eller risker att utveckla diabetes. Därför informerades Linda att hon inte behövde oroa sig.

Beslutet motiverades och förklarades väl med att algoritmen inte identifierade några avvikelser eller fann några risker i Lindas tillstånd/bakgrund för att utveckla diabetes. Linda fick på så sätt en tydlig förklaring om hur och varför AI-verktyget kom fram till dess slutsats. Det gav henne insikt i beslutsprocessen och en tydlig förklaring till varför beslutet om diagnosen togs.

Trots bedömningen sökte Linda efter sju månader vård igen eftersom symtomen fortfarande kvarstod. Linda kände sig orolig för att det hade ställts fel diagnos. Vården tog ett bra ansvar genom att boka en ny tid åt Linda. Vårdens bemötande visar att Lindas oro togs på allvar och att vården är redo att ta itu med eventuella misstag i diagnostiseringen. Åtgärden visar på tydliga riktlinjer för bra ansvarsutkrävande där patienternas välmående prioriteras.

8.2.2 Text som läses av ”Kontrollgrupp” *(Användning av AI som beslutsstöd + låg transparens)*

Trots rätt diagnos av AI-verktyg i vården: Stora frågetecken över bristande förklaringar

Linda, 37, sökte vård då hon oroade sig för symtom på diabetes. Linda hoppades att ett vårdbesök skulle kunna ge tydliga svar och lugna hennes oro. Vid vårdbesöket kartlades Lindas tillstånd, bakgrund och risk att utveckla diabetes. Det togs också blodprover.

Vid bedömningen av Lindas tillstånd användes ett AI-verktyg som beslutsstöd till läkaren. AI-verktyget visade inga indikationer på avvikelser eller risker att utveckla diabetes. Därför

informerades Linda att hon inte behövde oroa sig. Däremot gavs ingen förklaring till hur eller varför AI-verktyget har kommit fram till beslutet.

Trots bedömningen söker Linda efter sju månader vård igen eftersom symtomen fortfarande kvarstår. Linda känner sig orolig för att det har ställts fel diagnos. Vården tog ett bra ansvar genom att boka en ny tid åt Linda. Vårdens bemötande visar på att Lindas oro togs på allvar och att de är redo att ta itu med eventuella misstag i diagnostiseringen. Åtgärden visar på ett gott ansvarsutkrävande där patienternas välmående och tillit prioriteras.

8.3 Interaktionsanalyser med kön, ålder, utbildning

8.3.1 Interaktionsanalys med faktorn kön

	Förtroendeindex
AI-stöd med hög transparens	0,32 ** (0.13)
Kön 1. (man)	0.20 (0.14)
2. (annat)	0.06 (0.68)
AI-stöd med hög transparens x kön (man)	(-0.15) (0.19)
AI-stöd med hög transparens x kön (annat)	0.18 (0.96)
Konstant	2,44*** (0.09)
Antal observationer	201
R^2	0,05

*Kommentar: Standardfel inom parentes. P-värden visas med signifikiansstjärnor, * p < 0.05, ** p < 0.01, *** p < 0,001. Kön, 0 = kvinna, 1 = man, 2 = annat.*

8.3.2 Interaktionsanalys med faktorn ålder

	Förtroendeindex
AI-stöd med hög transparens	-14.67 (11.50)
Ålder	- 0.00 (0.00)
AI-stöd med hög transparens x ålder	(0.01) (0.01)
Konstant	6.28 (7.94)
Antal observationer	201
R^2	0,04

*Kommentar: Standardfel inom parentes. P-värden visas med signifikiansstjärnor, * p < 0.05, ** p < 0.01, *** p < 0,001*

8.3.3 Interaktionsanalys med faktorn utbildning

	Förtroendeindex
AI-stöd med hög transparens	0.63 * (0.29)
Utbildning	0.13 * (0.10)
AI-stöd med hög transparens x utbildning	(-0.11) (0.07)
Konstant	2.10 *** (0.20)
Antal observationer	201
R^2	0,04

*Kommentar: Standardfel inom parentes. P-värden visas med signifikiansstjärnor, * p < 0.05, ** p < 0.01, *** p < 0,001*