



# Sammenhengen mellom revisors skeptiske holdning og forståelse for beslutningsprosessen til et verktøy med kunstig intelligens

**Maria Hådem og Sandra Marita Nilsen**

**Veileder: Jonas Gaudernack og Carmen Olsen**

Masteroppgave, regnskap og revisjon

NORGES HANDELSHØYSKOLE

Dette selvstendige arbeidet er gjennomført som ledd i masterstudiet i økonomi- og administrasjon ved Norges Handelshøyskole og godkjent som sådan. Godkjenningen innebærer ikke at Høyskolen eller sensorer innestår for de metoder som er anvendt, resultater som er fremkommet eller konklusjoner som er trukket i arbeidet.

## Sammendrag

Det forventes at implementering av kunstig intelligens i revisjonsmetodikken vil gjøre revisjonen mer effektiv (Asklund, 2016). Samtidig har man sett flere eksempler på at tidligere bruk av digitale verktøy og ekspertsystemer i revisjonen endte med at det ikke ble utnyttet fullt ut. Teknologianvendelsen var lite effektivt som følge av at revisor var skeptisk til verktøyene og ikke stolte på dem (Fischer, 1996). Dette kan også være tilfellet ved bruk av verktøy som baserer seg på kunstig intelligens, ettersom kompleksiteten i verktøyet sin bakenforliggende teknologi vil gjøre det vanskelig for revisor å forstå virkemåten, og han vil derfor ikke ha tillit til systemet. Tidligere forskning tilsier at det er større sannsynlighet for at revisor vil bruke systemet når han har tillit, mens skepsis til systemet vil føre til at han ikke bruker det (Elson, Derrick, & Ligon, 2018). Revisors skeptiske holdning til kunstig intelligens kan måles gjennom hvor pålitelig de anser revisjonsbeviset og om de vil utføre flere revisjonshandlinger (Nelson, 2009; Hurtt, Brown-Liburd, Earley, & Krishnamoorthy, 2013; Nolder & Kadous, 2014). Vårt forskningsspørsmål undersøker hvorvidt lavere forståelse for beslutningsprosessen til et verktøy med kunstig intelligens vil øke revisors skeptiske holdning ved vurdering av revisjonsbevis. For å besvare forskningsspørsmålet gjennomførte vi et eksperiment med et 2x1 design, hvor 22 deltakere ble plassert i to grupper. Vi manipulerte den uavhengige variabelen ved å gi deltakerne i den første gruppen en forklaring på beslutningsprosessen til KIT som deltakerne i gruppe to ikke fikk. Dette skulle føre til at deltakerne i gruppe en hadde en høyere forståelse enn deltakerne i gruppe to. Deretter målte vi hvilken effekt dette hadde på den avhengige variabelen, den skeptiske holdningen til deltakerne. Funnene i oppgaven indikerer at forståelse gjør deltakerne mindre skeptisk, men på grunn av manglende signifikante verdier kan vi ikke forkaste nullhypotesene våre. Det er dermed ikke grunnlag for å uttale at det er en sammenheng mellom forståelse og skeptisk holdning.

## Forord

Denne masterutredningen er et selvstendig arbeid, skrevet som en del av vår mastergrad i regnskap og revisjon ved Norges Handelshøyskole i Bergen.

Prosessen med å skrive masteroppgaven har vært utfordrende og givende. Temaet for oppgaven er revisors skeptiske holdning i møte med verktøy som benytter kunstig intelligens. Teknologi og kunstig intelligens er et dagsaktuelt tema, og dette har vi kjent på gjennom hele semesteret. Det har forløpende kommet nye artikler og forskning på feltet, og vi har måttet legge ned mye arbeid for å holde oss oppdatert. Flere selskaper har i 2018 annonsert spennende nye verktøy som benytter kunstig intelligens, og det innblikket vi har fått i hvordan fremtidens revisjon vil se ut har vært svært lærerikt. I tillegg har prosessen testet vår evne innen selvstendig beslutningstaking og samarbeid. Summen av dette har bidratt til en positiv erfaring hvor vi sitter igjen med økt kunnskap og en dypere forståelse av fagområdet og tema.

Vi vil takke våre veiledere, Jonas Gaudernack og Carmen Olsen for god hjelp og nyttige råd, og PwC for masterstipendet vi fikk for å gjennomføre oppgaven. Vi vil også takke våre medstudenter som skriver masteroppgave med samme veiledere som oss. De har gitt oss gode tilbakemeldinger på pilottesten og mange nyttige diskusjoner gjennom semesteret. Til slutt vil vi takke alle som deltok i eksperimentet, og familie og venner som alltid har vært tilgjengelige for støttende og oppmuntrende ord.

Bergen, juni 2018

Maria Hådem

Sandra Marita Nilsen

---

# Innholdsfortegnelse

<b>SAMMENDRAG</b> .....	<b>I</b>
<b>FORORD</b> .....	<b>II</b>
<b>1 INNLEDNING</b> .....	<b>1</b>
1.1 MOTIVASJON FOR OPPGAVEN .....	1
1.1.1 <i>Forskningsspørsmål</i> .....	2
1.2 AVGRENSNINGER.....	2
1.3 DISPOSISJON OVER RESTEN AV OPPGAVEN .....	3
<b>2 LITTERATUR OG FORSKNING</b> .....	<b>4</b>
2.1 KUNSTIG INTELLIGENS.....	4
2.1.1 <i>Beskrivelse av teknologien bak kunstig intelligens</i> .....	4
2.1.2 <i>Kunstig intelligens i revisjon</i> .....	4
2.1.3 <i>Ekspertsystem</i> .....	5
2.1.4 <i>Samarbeid mellom maskin og menneske</i> .....	6
2.2 REVISJONSBEVIS OG SKEPSIS.....	7
2.2.1 <i>Revisjonsbevis</i> .....	7
2.2.2 <i>Skepsis i en revisjonssammenheng</i> .....	8
2.2.3 <i>Konsekvensen av at revisor er for skeptisk</i> .....	9
<b>3 KONSEPTUELL MODELL OG HYPOTESER</b> .....	<b>10</b>
3.1 KONSEPTUELL MODELL FOR EFFEKTE AV FORSTÅELSE PÅ REVISORS SKEPTISKE HOLDNING .....	10
3.2 HYPOTESER .....	11
3.2.1 <i>Revisor vurderer revisjonsbevis som mindre pålitelig når han mangler forklaring på verktøyets beslutningsprosess</i> .....	11
3.2.2 <i>Revisor velger å utføre flere revisjonshandlinger når han mangler forklaring på verktøyets beslutningsprosess</i> .....	12
<b>4 METODE</b> .....	<b>13</b>
4.1 VALG AV FORSKNINGSDESIGN .....	13
4.2 DATAINNSAMLING .....	14
4.2.1 <i>Utvalg</i> .....	14
4.2.2 <i>Utforming av case og spørsmål</i> .....	14
4.2.3 <i>Pilottest</i> .....	15
4.2.4 <i>Praktisk gjennomføring</i> .....	16
4.3 ANALYSE.....	16
4.3.1 <i>Forberedelse til analyse</i> .....	16
4.3.2 <i>Dataanalyse</i> .....	17

4.4	RELIABILITET OG VALIDITET .....	17
<b>5</b>	<b>RESULTATER.....</b>	<b>20</b>
5.1	MANIPULASJONSSJEKK.....	20
5.2	DESKRIPTIV STATISTIKK.....	20
5.3	TEST AV HVORVIDT REVISOR HAR EN LAVERE VURDERING AV PÅLITELIGHET VED MANGLENDE FORKLARING .....	21
5.4	TEST AV HVORVIDT REVISOR VELGER Å UTFØRE FLERE REVISJONSHANDLINGER VED MANGLENDE FORKLARING .....	22
5.5	ANDRE OBSERVASJONER.....	24
5.5.1	<i>Skepsis til revisjonsselskapets arbeidsmetoder .....</i>	24
5.5.2	<i>Forholdet mellom pålitelighet og tilstrekkelighet .....</i>	24
<b>6</b>	<b>DISKUSJON OG KONKLUSJON.....</b>	<b>25</b>
6.1	DISKUSJON AV HYPOTESENE .....	25
6.1.1	<i>Diskusjon av hvorvidt manglende forklaring fører til lavere vurdering av pålitelighet .....</i>	25
6.1.2	<i>Diskusjon av hvorvidt manglende forklaring fører til at revisor velger å utføre flere revisjonshandlinger .....</i>	26
6.1.3	<i>Diskusjon av kontrollvariablene .....</i>	26
6.2	DISKUSJON AV ANDRE OBSERVASJONER .....	27
6.2.1	<i>Diskusjon av forskjellen mellom vurderingen av pålitelighet og videre revisjonshandlinger .....</i>	27
6.2.2	<i>Diskusjon om skepsis til revisors arbeidsmetoder er en del av profesjonell skepsis .....</i>	28
6.3	KONKLUSJON .....	29
6.4	BEGRENSNINGER .....	29
6.5	ANBEFALINGER TIL PRAKSIS OG VIDERE FORSKNINGSTEMA .....	30
	<b>LITTERATURLISTE .....</b>	
	<b>VEDLEGG 1 CASETEKST OG SPØRSMÅL.....</b>	

## Figurliste

Figur 1: Libby boxes .....	10
----------------------------	----

## Tabelliste

Tabell 1: Deskriptiv statistikk .....	20
Tabell 2: Gjennomsnitt og standardavvik til variabelen pålitelighet.....	21
Tabell 3: ANOVA .....	21
Tabell 4: ANOVA med kontrollvariabler.....	22
Tabell 5: Gjennomsnitt og standardavvik til variabelen tilstrekkelighet.....	22
Tabell 6: ANOVA .....	23
Tabell 7: ANOVA med kontrollvariabler.....	23
Tabell 8: Profesjonell skepsis.....	24

# 1 Innledning

I dette kapitlet vil vi først presentere oppgavens tema og forskningsspørsmålet vårt, samt dens plassering innen forskningsuniverset. Deretter vil vi legge frem avgrensningene vi har valgt å gjøre. Vi avslutter kapitlet med en oversikt over resten av oppgavens disposisjon.

## 1.1 Motivasjon for oppgaven

Det har vært en enorm fremgang innen teknologi de siste årene. Spesielt kunstig intelligens har vært i fokus, og det har vært forsket mye på hvordan dette vil påvirke fremtidens jobbmarked. Det er uttalt at så mye som en tredjedel av den norske sysselsettingen vil bli automatisert innen 2035, og revisjonsbransjen er et av områdene som er mest utsatt (Pajarinen, Rouvinen, & Ekeland, 2015). Vi ser allerede at flere av de store revisjonsselskapene har begynt å ta i bruk kunstig intelligens i revisjon (Davenport, 2016; Samuel, 2018). Kunstig intelligens vil kunne utføre jobben raskere og bedre enn revisor, og dette vil i årene fremover føre til en endring av revisjonsbransjen (Greenman, 2017). Det blir dermed viktig å se hvilke muligheter og utfordringer dette vil medføre, slik at man kan ta i bruk teknologien på best mulig måte.

Å ta i bruk de teknologiske verktøyene på best mulig måte vil blant annet si at revisor utøver en optimal grad av profesjonell skepsis ved anvendelsen av dem, for dette har betydning for revisjonskvaliteten og effektiviteten. Revisjonskvaliteten avhenger av hvilke revisjonsbevis som blir innhentet og revisors evaluering av dem (Francis, 2011). I et teknologisk perspektiv kan dette blant annet oversettes til at revisor anvender digitale verktøy på en riktig måte og gjør en riktig beslutning vedrørende bevisene den utarbeider. Det er viktig at revisor gjør en vurdering av når det er hensiktsmessig å bruke digitale verktøy og er skeptisk til hvilke konsekvenser det kan medføre. At revisor er skeptisk har dermed en preventiv funksjon ved at det reduserer risikoen for at revisor tar gale vurderinger eller beslutninger (Olsen, 2014).

Det er samtidig viktig at kravet til effektivitet blir oppnådd. Bruk av kunstig intelligens vil effektivisere revisjonen, men kunstig intelligens inneholder en så komplisert teknologi at det vil være vanskelig, nær sagt umulig, for revisorene å forstå den (Kokina & Davenport, 2017). Dette kan dermed føre til at revisor er så skeptisk at teknologien ikke blir utnyttet på den mest effektive måten.

Den nye teknologien kan spesielt være et problem i revisjonsbransjen fordi man har sett en tidligere tendens til at ny teknologi ble møtt med motvilje og skepsis. Revisor valgte å ikke ta i bruk verktøyene eller utførte tradisjonelle revisjonshandlinger i tillegg, slik at revisjonen ble ineffektiv (Fischer, 1996). Når det vil være vanskelig for revisor å forstå den kompliserte teknologien bak kunstig intelligens, er tillit foreslått som et nødvendig element for å få til en suksessfull adopsjon av teknologien (Nunes & Jannach, 2017). Det er videre foreslått at en forklaring på beslutningsprosessen til et verktøy kan gjøre at revisor har en høyere grad av tillit til verktøyet, og dermed tar det i bruk i en større skala (Ye & Johnson, 1995).

### **1.1.1 Forskningsspørsmål**

Oppgaven har som formål å undersøke sammenhengen mellom skeptisk holdning og forståelse. Dette førte til at vi utledet følgende forskningsspørsmål for oppgaven vår:

«Vil lavere forståelse for beslutningsprosessen til et verktøy med kunstig intelligens øke revisors skeptiske holdning ved vurdering av revisjonsbevis?»

Teknologien bak kunstig intelligens er så ny at det er forsket lite på hvordan den påvirker revisors beslutningsatferd. Svaret på forskningsspørsmålet vårt vil ha en praktisk betydning for videre forskning og adopsjon av kunstig intelligens i revisjon. Det kan bidra til å sikre en optimal utnyttelse av nye digitale verktøy, og dette kan øke kvaliteten og effektiviteten til revisjonen.

## **1.2 Avgrensninger**

Når vi ser på regulatoriske forhold har vi valgt å kun se på de revisjonsstandardene som er aktuelle i norsk revisjon. Disse er de internasjonale revisjonsstandardene (ISA) som er utgitt av International Auditing and Assurance Standards Board (IAASB).

I følge ISA 200, pkt.13 er profesjonell skepsis definert som «en holdning om innebærer at revisor stiller spørsmål og er oppmerksom på forhold som kan indikere vesentlig feilinformasjon som følge av feil eller misligheter, og foretar en kritisk vurdering av revisjonsbevis» (IAASB, 2009, a). Det fremkommer likevel ikke av standarden om dette kun omhandler skepsis ovenfor klienten, eller om det også omhandler skepsis til revisjonsverktøyene og metodene. Profesjonell skepsis er likevel den eneste formen for skepsis som er kjent i revisjonens verden, og selv om det ikke kommer frem av standarden om

dette også omhandler revisors arbeidsmetoder er det dette begrepet vi har ansett det som nødvendig å diskutere i teoridelen for å definere og måle skeptisk holdning.

I oppgaven vår forutsetter vi at tillit er det motsatte av skepsis, hvor lav grad av tillit øker den skeptiske holdningen og høy grad av tillit reduserer den skeptiske holdningen.

### 1.3 Disposisjon over resten av oppgaven

I kapittel to vil vi presentere teori og forskning som er relevant for forskningsspørsmålet vårt. Vi vil først gå gjennom hva kunstig intelligens er og hvordan det kan brukes i revisjonen, før vi går over til å presentere begrepene revisjonsbevis og profesjonell skepsis. I kapittel tre vil vi presentere vår konseptuelle modell og våre hypoteser. I kapittel fire går vi over til å beskrive vår metode og begrunne de valgene vi har tatt innenfor dette. Vi vil presentere resultatene fra analysen i kapittel fem. I kapittel seks vil vi diskutere disse før vi avslutningsvis vil konkludere på forskningsspørsmålet vårt, diskutere hvilke begrensninger oppgaven har og komme med forslag til videre forskning og praksis.



## 2 Litteratur og forskning

I dette kapitlet vil vi presentere litteratur og forskning som er relevant for forskningsspørsmålet vårt. Vi vil først definere hva kunstig intelligens er og hvordan det kan påvirke revisjonsbransjen fremover. Deretter vil vi redegjøre for begrepet revisjonsbevis, og til slutt tar vi for oss hva profesjonell skepsis er og hvordan det kan måles.

### 2.1 Kunstig intelligens

#### 2.1.1 Beskrivelse av teknologien bak kunstig intelligens

Kunstig intelligens er et komplisert fagfelt som oppsto for rundt 70 år siden. Begrepet kunstig intelligens er en samlebetegnelse på forskjellige teknologier som, alene eller i kombinasjon med hverandre, operer på en intelligent måte (Davenport, 2016). Det vil si at teknologien skaper en intelligens i maskiner som ligner den menneskelige intelligens, med evne til blant annet å forstå språk, resonnere og løse problemer.

Maskinene kan både lære over tid og motta og behandle ustrukturert informasjon på samme måte som mennesker (Aurstad, 2017). De har evne til å analysere og evaluere data, de forstår språkmønstre, tekst, bilder og lyd, og de kan bruke dette til å ta beslutninger. Det som gjør denne teknologien ny er at maskinen kan gjøre dette uten å være spesifikt programmert til det. I stedet lærer den av samspill og erfaring, og dette gjør systemet adaptivt og dynamisk (KPMG, 2016).

#### 2.1.2 Kunstig intelligens i revisjon

Flere revisjonsselskaper har allerede begynt å se på hvordan den nye teknologien kan forbedre revisjonen, og hvis de fortsetter å omfavne teknologien vil man kunne oppleve en stor omveltning i revisjonsbransjen hvor regnskapsrevisjonen slik vi kjenner den i dag er helt borte innen 2028 (Fjørtoft, 2018). Det vil være viktig å se på hvilke konsekvenser teknologiadopsjonen kan gi til revisjonsbransjen. Deloitte har tatt i bruk verktøyet Argus, som bruker kunstig intelligens ved gjennomgåelse av kontrakter (Davenport, 2016). I mars 2018 annonserte KPMG at de har utviklet et lignende verktøy. De kaller det «KPMG Contract Abstraction Tool», og den er resultatet av et samarbeid KPMG har med IBM for å bruke deres teknologi i revisjon (Samuel, 2018).

En av de største utfordringene med dagens revisjon er mangelen på tid (Broberg, Tagesson, Argento, Gyllengahm, & Mårtensson, 2017). Kunstig intelligens kan være en løsning på dette ved at data kan bli samlet inn raskere, og analysen av disse kan gjøres på en bedre måte (Asklund, 2016). Et eksempel er gjennomgang av dokumenter, som en maskin kan gjøre på en brøkdel av tiden et menneske ville brukt. Dette er tilfellet med Argus, som gjennom kunstig intelligens lærer fra hver interaksjon den har. Den bruker avansert maskinlæringsteknikker og naturlig språkprosessering til å automatisk identifisere og ta ut nøkkelinformasjon fra alle typer elektroniske dokumenter (Davenport, 2016). Slike verktøy kan bidra til å skape en revisjon som er mer effektiv og av bedre kvalitet ved at revisor kan bruke tid på andre områder.

Det er også andre måter man kan utnytte kunstig intelligens i revisjonen. En regnskapspost det er knyttet mye usikkerhet til er estimat på ukurans på varelager, og nye digitale verktøy kan hjelpe revisor med å skaffe tilstrekkelig og hensiktsmessig revisjonsbevis på dette området (Aurstad, 2017). Et verktøy som benytter kunstig intelligens kan også skanne regnskapet og foreslå risikoområder, og dermed bidra til å fastsette risikoen for vesentlig feil (Davenport, 2016).

En utfordring revisjonsbransjen møter i forbindelse med den teknologiske utviklingen, er at revisjonsstandardene ikke er skrevet på en måte som tar hensyn til bruk av digitale verktøy i revisjonen (Issa, Sun, & Vasarhelyi, 2016). Dette hindrer revisjonsselskapene i å utnytte nye verktøy på best mulig måte. I 2017 vant PwC prisen «audit innovation of the year» for en teknologi de kaller GL.ai. Denne teknologien vil kunne være en enorm hjelp til revisorene, men regulatoriske krav har vært en hindring for å kunne ta det i bruk i revisjonen (PwC, 2018).

### **2.1.3 Ekspertsystem**

Ekspertsystemer er en type kunstig intelligens som hadde stor popularitet innen regnskap og revisjon på 80- og 90 tallet, men på grunn av dårlig teknologi døde bruken ut (Gray, Chiu, Liu, & Li, 2014). Systemene etterligner skjønnsbruken til eksperter ved hjelp av regler som er programmert inn i systemet. Flere selskaper utviklet egne systemer med håp om å oppnå en mer effektiv revisjon (Messier & Hansen, 1987). Over tid fant man at teknologien bak systemene var for simpel og ikke levde opp til forventningene (Gray, Chiu, Liu, & Li, 2014). Systemene var aldri bedre enn programmeringen, kunnskapen den hadde var subjektiv og den var ikke dynamisk, slik at den raskt ble utdatert. Ekspertsystemene ble over tid byttet ut med andre teknologier som benyttet kunstig intelligens (Sutton, Holt, & Arnold, 2016).

I storhetstiden til ekspertsystemene ble det forsket mye på hvorvidt brukerne omfavnet denne typen teknologien, og mange fant at det var en tendens til at revisorer brukte systemene feil, og at de valgte å ikke bruke de. Feilbruk innebar at revisorene lente seg for mye på beslutningene fra maskinene, og dette var ofte tilfellet med revisorer med lite arbeidserfaring (Abou-El-Sood, Kotb, & Allam, 2015; Jensen, Lowry, Burgoon, & Nunamaker, 2010). På den andre siden er det observert mange eksempler på at revisor var for skeptisk til teknologien og vegret seg for å bruke systemene, og det ble funnet at revisorer med arbeidserfaring hadde en overdreven tro på egne evner (Hampton, 2005; Whitecotton, 1996). Det er også funnet eksempler på at alder påvirket teknologibruken, ved at yngre revisorer har større adopsjonsvilje (Morris & Venkatesh, 2000). Fischer (1996) fant at nye verktøy ble møtt med motvilje, og for at de skulle bli tatt i bruk måtte brukerne selv se nytten ved dem. Da revisorene ble tvunget til å bruke digitale verktøy valgte de fortsatt å gjøre de tradisjonelle revisjonshandlingene i tillegg, noe som skapte en ineffektiv revisjon.

#### **2.1.4 Samarbeid mellom maskin og menneske**

Det forventes at en implementering av fremtidens kunstige intelligens i revisjonsbransjen vil føre til en endring i samspillet mellom revisor og maskinen. Fremtidens revisjonsverktøy kan ha evnen til å gjøre skjønnsmessige vurderinger, og disse vil da kunne ta over en enda større del av revisjonsarbeidet enn de gjør i dag. Likevel er de fleste eksperter enige i at kunstig intelligens ikke kan erstatte revisor fullstendig i nær fremtid (Davenport, 2016; Greenman, 2017). De sier at man heller vil se et samarbeid mellom revisor og verktøy som bruker kunstig intelligens. Verktøyet vil bli et teammedlem, og dette vil kunne føre til bedre beslutninger og en mer produktiv revisjon. (Davenport, 2016; Issa, Sun, & Vasarhelyi, 2016; Buchholz, Kulms, & Kopp, 2017).

En utfordring med verktøy som benytter kunstig intelligens er mangelen på gjennomsiktighet som oppstår når verktøyene er en «black box» (Kokina & Davenport, 2017). «Black box» er et begrep som benyttes når man bare kan se input og output, uten mulighet til å få kunnskap om maskinens indre prosesser. Det vil være vanskelig for brukere av slike verktøy å få en forståelse for verktøyets arbeidsprosesser, og dette vil kunne påvirke hvordan revisor bruker verktøyet (Ribeiro, Singh, & Guestrin, 2016).

For å få til en adopsjon av kunstig intelligens må man finne en måte å overkomme de utfordringene «black boxes» medfører. Forskning har funnet at en forståelig og lett tilgjengelig

begrunnelse fra maskinen har hjulpet (van Dongen & van Maanen, 2013; Ye & Johnson, 1995). Gjennom å forklare sin arbeidsprosess kan maskinen gi revisor en forståelse for hva maskinen har gjort, og dette har vist seg å være skape tillit til resultat. Studier viser videre at tillit til system vil føre til bruk av system, og mangel på tillit vil føre til at systemet ikke blir utnyttet fullt ut (Elson, Derrick, & Ligon, 2018; Lewis, Sycara, & Walker, 2018).

## 2.2 Revisjonsbevis og skepsis

### 2.2.1 Revisjonsbevis

Målet til revisor er å oppnå betryggende sikkerhet for at regnskapet ikke inneholder vesentlig feilinformasjon. Dette gjør revisor ved å redusere risikoen for feilinformasjon til lav ved å planlegge revisjonen slik at det innhentes tilstrekkelige og hensiktsmessige revisjonsbevis som vil underbygge konklusjonen i revisjonsberetningen (Appelbaum, 2016). Revisor vil gjøre en vurdering av den iboende risikoen og kontrollrisikoen til klienten, og hvilken revisjonsrisiko som anses som akseptabel. Deretter vil revisor fastsette den oppdagelsesrisikoen som er nødvendig for å oppnå ønsket revisjonsrisiko. Denne setter rammen for hvilke og hvor mange revisjonshandlinger revisor må gjennomføre, og dermed hvilke revisjonsbevis som må innhentes (Gulden, 2016).

Tilstrekkelighet er et mål på omfanget av bevisene revisor må innhente, og bestemmes av oppdagelsesrisikoen fastsatt av revisor og kvaliteten på bevisene (Appelbaum, 2016). Hensiktsmessigheten er at bevisene må ha bevisverdi i relasjon til den regnskapsopplysningen og den påstanden revisor ønsker å bekrefte (Gulden, 2016). At revisjonsbevisene er hensiktsmessige vil si at de må være relevante og pålitelige. Relevans er at revisjonsbeviset har en logisk tilknytning til formålet med revisjonshandlingene som revisor utfører. Pålitelighet betyr hvor mye revisor kan stole på beviset. Påliteligheten påvirkes av informasjonens kilde og type, og avhenger av de enkelte omstendighetene rundt innhenting (IAASB, 2009, b).

Revisjonsstandardene krever at revisor skal innhente tilstrekkelige og hensiktsmessige revisjonsbevis ved bruk av profesjonelt skjønn, og foreta en kritisk vurdering av revisjonsbevis (IAASB, 2009, a; IAASB, 2009, b). Den digitale verden endrer hvilke former for revisjonsbevis som innhentes og hvordan revisor gjør en vurdering av dem (AICPA, 2015). Bruk av kunstig intelligens i revisjonen kan føre til at det er maskinens skjønn som tar

avgjørelser, og ikke revisor selv. Bruk av kunstig intelligens frigjør ikke revisor for ansvar, for revisor er, uavhengig av hvilke verktøy han bruker eller handlinger han utfører, ansvarlig for beslutningen som blir tatt (Omotoso, 2012). Dermed kan ikke revisor stole blindt på maskinen, men må i stedet bruke skjønn og skepsis til å vurdere maskinens avgjørelser.

Dette gir nye utfordringer til revisors vurdering av tilstrekkelighet og hensiktsmessighet. Spesielt påliteligheten til revisjonsbevisene kan være vanskelig å avgjøre, ettersom elektroniske bevis ofte er innhentet fra eksterne kilder som kan være vanskelig å verifisere. Det vil kreves teknologiske ferdigheter hos revisor for å forstå og gjøre en vurdering av elektroniske data og bevis (Appelbaum, 2016).

### **2.2.2 Skepsis i en revisjonssammenheng**

Blant forskere finnes det ikke en felles forståelse eller definisjon av hva profesjonell skepsis er, noe som gjør det vanskelig å definere og måle profesjonell skepsis. Profesjonell skepsis har blant annet blitt definert som det motsatte av tillit, eller at revisor er uavhengig eller mistenkelig (Hurtt R. K., 2010).

Hurtt (2010) forklarer profesjonell skepsis som en flerdimensjonal individuell karakteristikk. Hun deler det opp i en tilstand og et personlighetstrekk, hvor et personlighetstrekk er mer varig, mens en tilstand er noe som oppstår midlertidig avhengig av situasjonsvariabler. For å måle personlighetstrekket utarbeidet hun en skala kalt «The Hurtt Professional Skepticism Scale» (HPSS). Hvor høy grad av skepsis en revisor har vil bli uttrykt i en profesjonell setting som profesjonell skepsis.

Nolder & Kadous (2014) forklarer profesjonell skepsis som en tankegang og en holdning. En tankegang er en samling av kriterier for å ta en beslutning og kognitive prosesser for å utføre en oppgave. Blant dette ligger det hvordan revisor søker etter informasjon, altså revisors åpenhet, oppmerksomhet og mottakelighet for informasjon når han skal vurdere bevisene. En revisor som har dette vil bruke lenger tid på å ta en avgjørelse. En holdning er noe som forklarer handlingene, og er vanskeligere å se og måle. Holdninger i revisjonssammenheng inkluderer revisors tro på ledelsens påstander og på revisjonsbevisene, nærmere bestemt revisors tanker om risiko for feilinformasjon og revisors vurderinger om revisjonsbevisene er relevante og pålitelige.

For at en revisor skal utøve profesjonell skepsis i en revisjonssammenheng må både en skeptisk vurdering og en skeptisk handling være til stede. En skeptisk vurdering vil si at revisor identifiserer at det er en risiko og at videre undersøkelse behøves, mens en skeptisk handling ikke vil oppstå før revisor velger å utføre handlinger basert på denne vurderingen (Nelson, 2009). Hurtt, Brown-Liburd, Earley & Krishnamoorthy (2013) bygger videre på Nelson sitt rammeverk og definerer fire karakteristika som vil påvirke den skeptiske vurderingen og handlingen til revisor, dette er karakteristika ved revisor, bevisene, klienten og det eksterne miljøet.

### **2.2.3 Konsekvensen av at revisor er for skeptisk**

Når både en skeptisk vurdering og en skeptisk handling er tilstede er en konsekvens at revisor vil utføre flere revisjonshandlinger før han konkluderer med at regnskapet ikke inneholder vesentlig feil (Hurtt, Brown-Liburd, Earley, & Krishnamoorthy, 2013).

Nelson (2009), som har et presumptivt syn på profesjonell skepsis, forklarer bakgrunnen til dette ved at en revisor med høy grad av profesjonell skepsis vil trenge overbevisning om at påstanden er riktig, fordi han antar at det finnes feil i regnskapet, og må dermed utføre flere revisjonshandlinger. Man kan i dette tilfellet havne i en situasjon hvor en revisor har for høy grad av profesjonell skepsis, og revisjonen således blir ineffektiv.

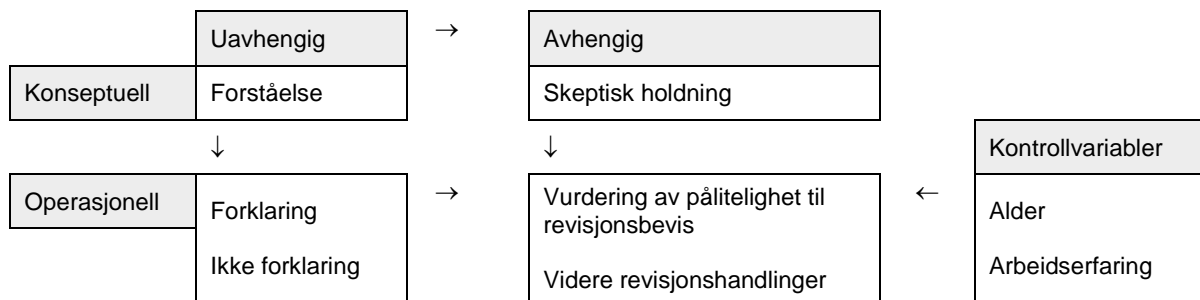
De internasjonale revisjonsstandardene har derimot et nøytralt syn på profesjonell skepsis, da revisor verken skal ha en positiv eller negativ antagelse vedrørende klientens informasjon. Dette ligger også til grunn for HPSS skalaen til Hurtt. Hurtt, Eining og Plumlee fant at revisorer som scorer høyere på HPSS skalaen vil innhente flere revisjonsbevis (Quadackers, Groot, & Wright, 2014). Dette indikerer at både under et nøytralt og presumptivt syn på profesjonell skepsis vil en mer skeptisk revisor innhente flere revisjonsbevis.

### 3 Konseptuell modell og hypoteser

I dette kapitlet vil vi presentere vår konseptuelle modell og hypoteser.

#### 3.1 Konseptuell modell for effekten av forståelse på revisors skeptiske holdning

Figur 1: Libby boxes



Vi har valgt å bruke rammeverket til Libby (1981) for å illustrere vår konseptuelle modell. Formålet med et eksperiment er å undersøke hvordan en faktor, den uavhengige variabelen, påvirker en annen, den avhengige variabelen. For å se på forholdet mellom disse manipuleres den uavhengige variabelen samtidig som effekten på den avhengige variabelen måles. Disse variablene kan ikke måles direkte, så for å kunne teste dem må de operasjonaliseres til målbare variabler.

Forskningsspørsmålet vårt ser på hvordan forståelse påvirker skeptisk holdning. Forståelse er da vår uavhengige variabel og skeptisk holdning er den avhengige variabelen. Vi operasjonaliserer disse til følgende målbare variabler; «forklaring» og «ikke forklaring» for begrepet forståelse og «vurdering av pålitelighet» og «videre revisjonshandlinger» for begrepet skeptisk holdning. Disse variablene vil vi teste gjennom to hypoteser, slik at vi kan si noe om forholdet mellom forståelse og skeptisk holdning.

Vi operasjonaliserer forståelse inn i «forklaring» og «ikke forklaring» fordi vi antar at forklaringen vil føre til en bedre forståelse, mens fravær av forklaring vil føre til lavere forståelse.

Vi valgte å operasjonalisere skeptisk holdning inn i to målbare variabler basert på det faktum at revisors utøvelse av profesjonell skepsis må inneholde både en skeptisk vurdering og en skeptisk handling (Nelson, 2009). Vi valgte revisors vurdering av påliteligheten til

revisjonsbeviset som skeptisk vurdering fordi dette er antatt å være det mest kritiske elementet ved elektroniske revisjonsbevis (Appelbaum, 2016). Videre valgte vi å måle den skeptiske handlingen gjennom å undersøke om deltakerne ville utføre flere revisjonshandlinger, og vi ser på valget om revisjonshandlinger som en vurdering av tilstrekkelighet (Nelson, 2009; Hurtt, Brown-Libur, Earley, & Krishnamoorthy, 2013). Vi anser vurdering av pålitelighet og vurdering av tilstrekkelighet som to uavhengige måter å måle skeptisk holdning på.

Kontrollvariabler er andre variabler enn den uavhengige som kan påvirke den avhengige variabelen. Basert på tidligere forskning og teori anser vi alder og arbeidserfaring som de variablene som mest sannsynlig kan ha en effekt på skeptisk holdning (Abou-El-Sood, Kotb, & Allam, 2015; Fischer, 1996; Hampton, 2005; Jensen, Lowry, Burgoon, & Nunamaker, 2010; Morris & Venkatesh, 2000; Whitecotton, 1996).

## 3.2 Hypoteser

### 3.2.1 Revisor vurderer revisjonsbevis som mindre pålitelig når han mangler forklaring på verktøyets beslutningsprosess

Når verktøy som benytter kunstig intelligens er så kompliserte at de blir en «black box», vil ikke revisor kunne forstå maskinens arbeidsprosess, og dette vil føre til at han ikke har tillit til verktøyet. Når dette verktøyet da produserer et revisjonsbevis vil revisors manglende forståelse øke hans skeptiske holdning.

Økningen i skeptisk holdning vil føre til at revisor vurderer påliteligheten til revisjonsbeviset negativt (Appelbaum, 2016; Nolder & Kadous, 2014).

*H1: «Revisor vurderer revisjonsbevis som mindre pålitelig når han mangler forklaring på verktøyets beslutningsprosess»*

*H0: «Revisor vurderer ikke revisjonsbevis som mindre pålitelig når han mangler forklaring på verktøyets beslutningsprosess»*



### **3.2.2 Revisor velger å utføre flere revisjonshandlinger når han mangler forklaring på verktøyets beslutningsprosess**

Økning i revisors skeptiske holdning som følge av manglende forståelse vil føre til at revisor ikke har tillit til revisjonsbeviset, han vil vurdere det som mindre tilstrekkelig og han vil velge å utføre flere revisjonshandlinger (Nelson, 2009; Hurtt, Brown-Liburd, Earley, & Krishnamoorthy, 2013).

*H2: «Revisor velger å utføre flere revisjonshandlinger når han mangler forklaring på verktøyets beslutningsprosess»*

*H0: «Revisor velger ikke å utføre flere revisjonshandlinger når han mangler forklaring på verktøyets beslutningsprosess»*

## 4 Metode

Her vil vi gjennomgå hvilke valg vi har tatt vedrørende forskningsdesign, samt hvordan vi planla og gjennomførte datainnsamlingen og analysen. Videre vil vi redegjøre for hvilke vurderinger vi har tatt i forhold til validitet og reliabilitet.

### 4.1 Valg av forskningsdesign

Forskningsdesign deles gjerne opp i to typer, kvalitativ og kvantitativ. Forskjellen er hovedsakelig at et kvantitativt forskningsdesign er deduktivt, mens et kvalitativt er induktivt. En deduktiv forskningsstrategi utarbeider hypoteser fra teoretiske perspektiver av det som skal undersøkes (Saunders, Lewis, & Thornhill, 2016). Det faktum at vi ønsket å finne ut hvordan forståelse påvirker revisors skeptiske holdning, og dette kunne besvares ved å utarbeide og teste hypoteser, gjorde at vi valgte et kvantitativt design.

Vi valgte å benytte et eksperimentelt design med to grupper fordi det er betegnet som den beste til å forklare og teste sammenhenger (Trochim, Donnelly, & Arora, 2016). Bakgrunnen for dette er at et eksperiment gir forskeren mulighet til å manipulere en uavhengig variabel, for å se hvilken påvirkning det har på den avhengige variabelen. Dette gjøres ved at deltakerne blir valgt vilkårlig i en av gruppene slik at det ikke er andre faktorer enn manipulasjonen som utgjør forskjellen (Gerber & Green, 2012).

Antall uavhengige variabler i designet er bestemt av forskningsspørsmålet. Desto flere uavhengige variabler man inkluderer, desto flere celler får man i designet, og dermed trenger man flere deltakere i eksperimentet (Trotman, 2001). Det er ofte en utfordring å få nok deltakere, og konsekvensen av å ikke ha nok deltakere per celle er at sannsynligheten for å finne signifikante resultater er svært liten. På grunn av begrensede ressurser og tid, samt det faktum at forskningsspørsmålet vårt kunne besvares ved bruk av én uavhengig variabel, var eksperimentet vårt mest passende i et 1x2 design.

## 4.2 Datainnsamling

### 4.2.1 Utvalg

Ved å gjøre et utvalg velges enheter fra populasjonen som skal undersøke, slik at man gjennom å studere utvalget kan si noe om populasjonen (Saunders, Lewis, & Thornhill, 2016). Det er dermed avgjørende at utvalget er representativt for populasjonen. Revisorer utgjør populasjonen i vårt eksperiment. Utvalget ble trukket fra revisjonsstudenter ved Norges Handelshøyskole (NHH) og Universitetet i Stavanger (UIS), både med og uten arbeidserfaring. Dette var først og fremst på grunn av den rammen som masteroppgaven har. Studentene er på mange måter ikke et representativt utvalg, men forskning har kommet frem til at det ikke er signifikante forskjeller mellom praktiserende revisorer og studenter i revisjonseksperimenter (Mortensen, Fisher, & Wines, 2012). I tillegg gikk undersøkelsen vår ut på at deltakerne skulle vurdere et digitalt revisjonsbevis. Praktiserende revisorer vil nok ha høyere forståelse for hvordan de skal vurdere et revisjonsbevis, men når det gjelder digitale bevis er det ikke usannsynlig at studenter har like mye, om ikke mer, kunnskap om teknologi.

For å dermed sikre at studentene er et representativt utvalg satte vi som forutsetning at deltakerne må ha fullført faget MRR411 revisjon eller tilsvarende revisjonsfag, eller hatt arbeidserfaring innen revisjon. Dette var først og fremst på grunn av forskningsspørsmålet vårt. Caseteksten vår og spørsmålene inneholder teori og begreper fra revisjon, og inngående kunnskaper om dette er dermed nødvendig for å sikre troverdige resultater.

### 4.2.2 Utforming av case og spørsmål

Caset og tilhørende spørsmål ble utarbeidet på en slik måte at det skulle være enkelt og forståelig for deltakerne, og ikke tok mer tid enn nødvendig. Dette anså vi som viktig for å oppnå en høy svarprosent. Caset gikk ut på at deltakerne er revisorer på et oppdrag hvor de måtte benytte revisjonsverktøyet KIT. KIT er et fiktivt verktøy vi fant opp på bakgrunn av hvilke verktøy som eksisterer i dag og hvordan eksperter tror verktøyene vil se ut i fremtiden. KIT benytter kunstig intelligens teknologi for å ta en vurdering og konkludere på hvorvidt varelageret er uten vesentlige feil. Verktøyet produserer deretter et dokument til revisor, som revisor da benytter som revisjonsbevis. For gruppe en inneholdt dette dokumentet konklusjonen fra KIT i tillegg til en nærmere forklaring på hva KIT har gjort, i form av tre problemer den oppdaget, hvordan den handlet og hva resultatet ble. For gruppe to inneholdt

dokumentet kun konklusjonen på vurderingen. Vi valgte å bruke varelageret som revisjonsområde fordi dette ofte er en vesentlig post hos mange revisjonsklienter, det er en del av pensumet i revisjonsfag og det er blitt pekt ut som et område som sannsynligvis vil kunne revideres av kunstig intelligens i fremtiden.

For å måle den skeptiske holdningen til deltakerne gjorde vi en forenkling av rammeverkene til Nelson (2009) og Hurtt et al (2013). Vi brukte videre rammeverket til Nolder og Kadous (2014) for å utforme spørsmålene. Det er utarbeidet mange forskjellige måter å måle profesjonell skepsis på, men vi anså disse rammeverkene som de som var best tilpasset å måle den skeptiske holdningen til deltakerne i vårt eksperiment. I tillegg laget vi tre spørsmål som omhandlet det kompliserte verktøyet for å se om det er forskjell på hvor skeptiske deltakerne er til verktøyet, og hvor skeptiske de er når de vurderer revisjonsbeviset.

Spørsmålene var prestrukturerte på en 5-punkts skala, dette gir oss muligheten til å få frem nyansene hos deltakerne, og gjør det enkelt og tidsbesparende for deltakerne å svare. For at vi ikke skulle lede deltakerne i en bestemt retning formulerte vi noen av spørsmålene negativt.

Vi inkluderte spørsmål vedrørende kontrollvariablene våre for å kunne se om de påvirker resultatene. I tillegg inkluderte vi et spørsmål om deltakerne anså skepsis til revisjonsselskapets arbeidsmetoder som en del av profesjonell skepsis. Vi målte den skeptiske holdningen til deltakerne gjennom rammeverk som måler profesjonell skepsis, selv om det ikke er en klarhet i om dette også gjelder skepsis til egne verktøy og metoder. Det var dermed interessant å kartlegge hva deltakerne mener og dermed få en bedre innsikt vedrørende begrepet profesjonell skepsis.

I et eksperimentelt design bør man inkludere en manipulasjonssjekk for å bekrefte at manipulasjonen er mottatt (Trotman, 2001). For å undersøke om deltakerne i gruppe en hadde lest forklaringen fra KIT ba vi de ta stilling til følgende påstand: «dokumentet fra KIT viser en tabell over de faktorene som var av størst betydning i dens revisjon av varelageret». Det var ikke mulig for deltakerne å gå tilbake i undersøkelsen for å lese informasjonen på nytt.

### **4.2.3 Pilottest**

Pilottesten ble utført på fem personer, som har samme veiledere som oss. Formålet med en pilottest er å forbedre undersøkelsen slik at det ikke er noen problemer for deltakerne å forstå spørsmålene eller for oss å analysere svarene senere (Saunders, Lewis, & Thornhill, 2016).

Pilottestingen var derfor viktig for å få innspill på hvordan spørsmålene var formulert og om det var noe som var uklart i caseteksten, samt hvor lang tid deltakerne brukte på undersøkelsen. Tilbakemeldingene fra pilottesten resulterte i enkelte justeringer i caseteksten.

#### **4.2.4 Praktisk gjennomføring**

Selve undersøkelsen ble utarbeidet i Qualtrics. Det var flere fordeler ved dette, blant annet ga det oss muligheten til å skjule IP adressen til deltakerne slik at de var helt anonyme, og det gjorde det enkelt for oss å lage, og for deltakerne å gjennomføre, undersøkelsen.

Sannsynligheten for at gjennomsnittet i utvalget vil nærme seg det faktiske gjennomsnittet i populasjonen vil øke ettersom hvor mange deltakere som deltar i eksperimentet (Jacobsen, 2005). Det var derfor viktig for oss å få et høyt antall deltakere, og vi brukte det nettverket vi hadde for å skaffe deltakere. Vi delte undersøkelsen på Facebook gruppene for revisjonsstudenter ved NHH og UIS. I tillegg ble det sendt ut melding igjennom Itslearning og mail til studenter som går MRR på NHH. Etter en uke sendte vi en purring på Facebook sidene. Undersøkelsen var åpen i tre uker.

### **4.3 Analyse**

#### **4.3.1 Forberedelse til analyse**

Før vi kunne starte med selve analysen måtte vi ekskludere en deltaker fra undersøkelsen, da personen ikke oppfylte forutsetningen om at deltakerne måtte ha gjennomført et revisjonsfag eller hatt arbeidserfaring innen revisjon. Resten av deltakerne oppfylte forutsetningen.

Siden vi hadde brukt Qualtrics til undersøkelsen var det enkelt å overføre resultatene vi sto igjen med, til analyseverktøyet. Dette sparte oss for tid og gjorde svarene mer pålitelige ved at det minimerte risikoen for feiltasting av tall.

Som en forberedelse til selve analysen grupperte vi også de svarene som hørte sammen, samt speilet de spørsmålene som var vinklet i motsatt retning, slik at alle svarene ble på den samme skalaen. Videre regnet vi ut et gjennomsnitt og standardavvik på hvert spørsmål og hver gruppe. Dette ga oss en oversikt og en indikasjon på hvilke resultater vi kunne vente oss i analysen. Her innså vi også at spørsmålene som omhandlet KIT ikke var avgjørende for å

besvare forskningsspørsmålet og hypotesene våre. For å ikke komplisere analysen unødvendig eliminerte vi spørsmålene om KIT.

### 4.3.2 Dataanalyse

Analysen gjennomførte vi i IBM SPSS. Formålet med analysen er å teste nullhypotesene gjennom en statistisk analyse.

Vi utførte en ensidig variansanalyse (ANOVA) for å se om vi kunne bekrefte eller forkaste de to nullhypotesene våre. En ANOVA foretar en analyse av sammenhengen mellom en avhengig og uavhengig variabel ved å se om variansen i den avhengige variabelen er blitt signifikant forskjellig på grunn av den uavhengige variabelen (Andersen & Jakobsen, 2004). En ANOVA sier dermed indirekte om gjennomsnittet er signifikant forskjellig i gruppene på grunn av den uavhengige variabelen.

I utgangspunktet skal det kun være manipulasjonen som utgjør forskjellen mellom gruppene i et eksperimentelt design, dette ønsket vi å oppnå ved tilfeldig fordeling gjennom Qualtrics. Tidligere forskning har derimot vist at alder og arbeidserfaring har hatt en påvirkning på hvordan revisor har vurdert bruken av digitale revisjonsverktøy. Vi utførte derfor en ny ANOVA som tar hensyn til kontrollvariablene.

Målet med analysene er å kunne påstå at det er forskjeller mellom gruppene, og at denne forskjellen ikke skyldes tilfeldigheter, men reelle forskjeller. Forutsetningen for å påstå dette er at forskjellene er signifikante (Grenness, 2004). Dette er uttrykt gjennom en p-verdi som forteller hvor sannsynlig det er at resultatet har forekommet på grunn av en tilfeldighet. En lav p-verdi gir en høy sannsynlighet for at forskjellen mellom gruppene er forårsaket av det vi ønsker å måle, og dermed er et troverdig resultat. Det er vanlig å bruke en p-verdi på fem prosent for at man skal forkaste nullhypotesen (Andersen & Jakobsen, 2004). Vi ønsker en lav risiko for type-to feil, som er å forkaste nullhypotesen selv om den er sann, og velger derfor et signifikansnivå på fem prosent.

## 4.4 Reliabilitet og validitet

Reliabiliteten refererer til hvor pålitelige målingene er (Halvorsen, 2006). Det vil si at undersøkelsen gir konsistente resultater på forskjellige tider og under forskjellige forhold (Saunders, Lewis, & Thornhill, 2016). Trusler mot en undersøkelses reliabilitet er graden av

systematiske og tilfeldige feil, samt det faktum at operasjonalisering av begreper vil utfordre reliabiliteten (Grenness, 2004).

Pilottestingen skal bidra til å sikre høy reliabilitet da denne avdekker feil eller uklarheter i teksten eller spørsmålene som kan føre til systematiske målingsfeil. Systematiske målingsfeil kan også oppstå hvis det er mange som ikke gjennomfører undersøkelsen. Konsekvensen av dette er at det kan oppnås skjeve resultater (Saunders, Lewis, & Thornhill, 2016). For å sikre en høy svarprosent laget vi caset så enkelt og kort som mulig. Tilfeldige feil er det vanskeligere å unngå, da dette ikke er noe vi kan styre fordi vi ikke er tilstede under eksperimentet, men tilfeldige feil vil gjerne utligne hverandre og dermed ha mindre betydning for resultatet (Trochim, Donnelly, & Arora, 2016).

Reliabilitet er en nødvendig, men ikke tilstrekkelig, forutsetning for validitet. Validitet er definert ut fra hvor høy grad av samsvar det er mellom teori og empiri. Det finnes mange former for validitet, men vi vil her konsentrere oss om intern og ekstern validitet, da dette ofte blir betegnet som kvalitetsindikatorer for eksperimentelle design (Grenness, 2004).

Intern validitet omhandler om man kan tilordne endringen i den avhengige variabelen til den uavhengige variabelen (Peecher & Solomon, 2001). Høy intern validitet vil si at endringen i den avhengige variabelen ikke er forstyrret av andre variabler. Disse andre variablene er det derfor viktig å kontrollere for i forskningsdesignet (Gerber & Green, 2012).

Den interne validiteten blir forsøkt styrket i eksperimentet vårt ved at vi har en kontrollgruppe som ikke mottar manipulasjonen og ved at den gruppen som får manipulasjonen også utsettes for en manipulasjonssjekk. Dette gjør det mulig å undersøke om vi har lyktes med å manipulere den uavhengige variabelen (Trotman, 2001). I tillegg vil tilfeldig fordeling gjennom Qualtrics minske skjevheter mellom gruppene og dette vil kunne øke den interne validiteten.

Ekstern validitet referer til om de sammenhengene man finner i eksperimentet kan bli generalisert til sammenhenger i populasjonen. Dette oppnås ved et representativt utvalg, som igjen oppnås ved tilfeldig utvalg (Trochim, Donnelly, & Arora, 2016). Studenter anses gjerne ikke som et representativt utvalg, men en stor del av revisjonsbransjen består av nyutdannede revisorer, og disse er ikke så forskjellige fra revisjonsstudenter. I tillegg er vår undersøkelse et bilde på hvordan fremtiden kan se ut i revisjonsbransjen, og det gjør det viktig å se på fremtidens arbeidstakere. Bruk av studenter kan dermed ses på som et representativt utvalg for

vårt eksperiment. Men validiteten er samtidig svekket fordi vi utførte en ikke-sannsynlighetsprosedyre ved at utvalget vårt kun besto av studenter ved NHH og UIS. Studenter fra forskjellige utdanningsinstitusjoner har forskjellige fagkunnskaper og bakgrunner, og dermed kan vi ha fått skjeve resultater ved å kun bruke studenter fra to utdanningssteder.



## 5 Resultater

I dette kapitlet vil vi først presentere resultatet fra manipulasjonssjekken vår, og deretter deskriptiv statistikk over utvalget vårt. Vi vil gå videre til å presentere resultatene fra analysene vi utførte for å svare på hypotesene, før vi til slutt presenterer andre interessante funn vi gjorde.

### 5.1 Manipulasjonssjekk

Vi hadde et spørsmål i eksperimentet som fungerte som manipulasjonssjekk. Denne ble alle deltakerne i gruppe en utsatt for. Av de 10 svarene vi fikk i denne gruppen svarte 50% feil.

### 5.2 Deskriptiv statistikk

Tabell 1: Deskriptiv statistikk

Alder	Gruppe 1	Gruppe 2	Totalt
Under 20 år	0	1	1
20-24 år	2	6	8
25-29 år	6	4	10
30-34 år	1	1	2
35-40 år	0	0	0
Over 40 år	1	0	1

Arbeids- erfaring	Gruppe 1	Gruppe 2	Totalt
Nei	7	5	12
Under 1 år	0	4	4
1-2 år	1	3	4
3-4 år	2	0	2

Fullført revisjonsfag	Gruppe 1	Gruppe 2	Totalt
Nei	1	2	3
Ja	9	10	19

Det var totalt 22 personer som fullførte undersøkelsen vår, hvor 10 personer var i gruppe en og 12 personer var i gruppe to. 81,8 % av de som svarte var mellom 20-29 år, litt over halvparten hadde ingen arbeidserfaring og de fleste med erfaring hadde under tre år av dette. 3 av de 22 deltakerne hadde ikke fullført revisjonsfag, men de hadde arbeidserfaring innen revisjon slik at de kunne tas med i analysen. Vi ser at alder er skjevt fordelt i gruppene, hvor de fleste deltakere i gruppe en var i alderskategorien 25-29 år, mens i gruppe to var halvparten i kategorien 20-24 år. Det samme gjelder for arbeidserfaring, hvor det var flere med arbeidserfaring i gruppe to enn i gruppe en.

## 5.3 Test av hvorvidt revisor har en lavere vurdering av pålitelighet ved manglende forklaring

I vår første hypotese ønsker vi å finne ut hvorvidt deltakerne i gruppe to rapporterer lavere pålitelighet til revisjonsbeviset, da de ikke får en forklaring fra verktøyet som gruppe en får.

Tabell 2: Gjennomsnitt og standardavvik til variabelen pålitelighet

Pålitelighet	Gjennomsnitt	Standardavvik
Gruppe 1 (forklaring)	3,67	0,737
Gruppe 2 (ikke forklaring)	3,03	0,745
Totalt	3,32	0,793

Gruppen som ikke fikk forklaring har et lavere gjennomsnitt og vurderte revisjonsbeviset som mindre pålitelig enn den gruppen som fikk forklaring. Gjennomsnittet til gruppe to ligger på middels grad pålitelig, mens gjennomsnittet til gruppe en ligger nærmere stor grad pålitelig. Vi ser altså en forskjell mellom gruppene, hvor gruppen som ikke fikk forklaring var mer skeptisk. Standardavvikene er veldig like, og går fra 0,737 i gruppe en til 0,745 i gruppe to.

Tabell 3: ANOVA

Pålitelighet	Type III Sum of Squares	df	Mean Square	F	Sig.
Forklaring / Ikke forklaring	2,226	1	2,226	4,051	0,058
Error	10,991	20	0,550		
Total	255,444	22			
Corrected Total	13,217	21			

a. R Squared = .168 (Adjusted R Squared = .127)

For å finne ut om denne forskjellen er signifikant utførte vi en ANOVA. Analysen ga oss en p-verdi på 5,8%. Denne er over signifikansnivået vårt på 5%, så vi kan ikke forkaste nullhypotesen vår.

Tabell 4: ANOVA med kontrollvariabler

Pålitelighet	Type III Sum of Squares	df	Mean Square	F	Sig.
Forklaring / Ikke forklaring	1,002	1	1,002	1,839	0,192
Arbeidserfaring	0,235	1	0,235	0,431	0,520
Alder	0,939	1	0,939	1,724	0,206
Error	9,806	18	0,545		
Total	255,444	22			
Corrected Total	13,217	21			

a. R Squared = .258 (Adjusted R Squared = .134)

Fra den deskriptive statistikken fremkommer det at både alder og arbeidserfaring er skjevt fordelt mellom gruppene. Det kan tenkes at dette kan ha hatt en påvirkning på resultatet vårt, så for å ta hensyn til kontrollvariablene utførte vi en ny ANOVA. Her ser vi at ingen av kontrollvariablene våre har hatt en signifikant påvirkning, da de har en p-verdi på 20 og 52%, men p-verdien økte til 19% så vi kan ikke forkaste nullhypotesen vår.

## 5.4 Test av hvorvidt revisor velger å utføre flere revisjonshandlinger ved manglende forklaring

I hypotese to ønsker vi å finne ut hvordan de to gruppene vurderer tilstrekkeligheten til revisjonsbeviset.

Tabell 5: Gjennomsnitt og standardavvik til variabelen tilstrekkelighet

Tilstrekkelighet	Gjennomsnitt	Standardavvik
Gruppe 1 (forklaring)	3,20	0,789
Gruppe 2 (ikke forklaring)	2,67	0,739
Totalt	2,91	0,791

Gruppen som ikke fikk forklaring har et lavere gjennomsnitt og ville gjøre flere revisjonshandlinger enn gruppen som fikk forklaring. Gjennomsnittet til gruppe to ligger nærmere liten grad tilstrekkelig og gjennomsnittet til gruppe en ligger på middels grad tilstrekkelig. Også her ser vi altså at gruppen som ikke fikk forklaring var mer skeptisk. Standardavvikene er ganske like, men gruppe en har høyere avvik på 0,789, mens gruppe to ligger på 0,739.

Tabell 6: ANOVA

Tilstrekkelighet	Type III Sum of Squares	df	Mean Square	F	Sig.
Forklaring / Ikke forklaring	1,552	1	1,552	2,675	0,118
Error	11,600	20	0,580		
Total	199,333	22			
Corrected Total	13,152	21			

a. R Squared = .118 (Adjusted R Squared = .074)

En ANOVA av variabelen tilstrekkelighet ga oss en p-verdi på 12%, og vi kan ikke forkaste nullhypotesen.

Tabell 7: ANOVA med kontrollvariabler

Tilstrekkelighet	Type III Sum of Squares	df	Mean Square	F	Sig.
Forklaring / Ikke forklaring	1,254	1	1,254	2,391	0,139
Arbeidserfaring	2,142	1	2,142	4,085	0,058
Alder	0,015	1	0,015	0,029	0,867
Error	9,439	18	0,524		
Total	199,333	22			
Corrected Total	13,152	21			

a. R Squared = .282 (Adjusted R Squared = .163)

Når vi gjorde en ny ANOVA, hvor vi inkluderte kontrollvariablene, ble p-verdien 14%, og vi kan altså heller ikke her forkaste nullhypotesen vår. Vi ser at arbeidserfaring har en p-verdi på 5,8%, så den er svært nær signifikansnivået på 5%. Dette tyder på at skjevfordelingen av arbeidserfaring, hvor gruppe to hadde flere deltakere med erfaring, kan ha hatt en påvirkning på resultatet. Samtidig ser vi at p-verdien på den uavhengige variabelen økte, noe som styrker antakelsen om at resultatet på denne hypotesen kan ha vært drevet av en, eller begge, av kontrollvariablene.

## 5.5 Andre observasjoner

### 5.5.1 Skepsis til revisjonsselskapets arbeidsmetoder

Tabell 8: Profesjonell skepsis

	Totalt
Ja	73%
Nei	27%

På spørsmålet om hvorvidt deltakerne anser skepsis til revisjonsselskapets arbeidsmetoder som en del av profesjonell skepsis svarte 16 av 22 ja. Dette indikerer et sterkt flertall som mente at profesjonell skepsis også inneholder skepsis til revisors arbeidsmetoder, selv om det ikke er omtalt i revisjonsstandardene.

### 5.5.2 Forholdet mellom pålitelighet og tilstrekkelighet

Tabell 9: Forholdet mellom pålitelighet og tilstrekkelighet

	Gjennomsnitt Pålitelighet	Gjennomsnitt Tilstrekkelighet
Gruppe 1	3,67	3,20
Gruppe 2	3,03	2,67
Totalt	3,32	2,91

Vi ser at begge gruppene vurderte revisjonsbeviset som mer pålitelig enn tilstrekkelig. Samlet gjennomsnitt for vurdering av pålitelighet er over «middels», og for vurdering av tilstrekkelighet er den litt under «middels». Dette betyr at deltakerne hadde en lavere grad av skeptisk vurdering enn skeptisk handling.

## 6 Diskusjon og konklusjon

I dette kapitlet vil vi drøfte resultatene våre opp mot teori og tidligere forskning. Vi vil starte med en diskusjon av resultatene knyttet til hypotesene, før vi diskuterer andre observasjoner. Vi vil deretter komme med en konklusjon på oppgaven, før vi vil redegjøre for begrensningene i oppgaven vår. Til slutt vil vi gå gjennom våre anbefalinger til praksis og videre forskning.

### 6.1 Diskusjon av hypotesene

#### 6.1.1 Diskusjon av hvorvidt manglende forklaring fører til lavere vurdering av pålitelighet

Hypotese en undersøker hvordan revisors vurdering av pålitelighet ble påvirket av manglende forklaring, og resultatet viser at gruppen som ikke fikk forklaring fra KIT vurderte revisjonsbeviset som 12,8% mindre pålitelig. Påliteligheten til revisjonsbevis er blant annet påvirket av dens kilde. I vårt eksperiment var kunstig intelligens kilden og det var sannsynligvis vanskelig for gruppen som ikke fikk forklaring å gjøre en vurdering av kilden når de ikke har utdypende kunnskap om teknologien. Tidligere forskning viser at revisor ofte er skeptisk når de ikke forstår kompliserte digitale verktøy, og for deltakerne i gruppe to blir dette uttrykt gjennom det at de vurderte revisjonsbeviset fra KIT som middels pålitelig. Gruppe en derimot, fikk en forklaring fra verktøyet og vurderte revisjonsbeviset som mer pålitelig. Dette er i tråd med tidligere forskning som fant at en begrunnelse fra verktøy skapte tillit til output. Gjennom å forklare valg den tok har antagelig verktøyet gitt gruppe en forståelse for hvordan revisjonsbeviset ble produsert, og det gjorde beviset mer pålitelig.

Det at gruppen som fikk forklaring anså beviset som mer pålitelig kan bety at mer informasjon, eller annen type informasjon, ville redusert skepsisen enda mer. Tidligere forskning har vist at jo lettere tilgjengelig og forståelig informasjonen er, jo mer tillit skaper den. Samtidig kan det tenkes at mer informasjon vil gi mer forståelse, og dette kan føre til en høyere pålitelighetsvurdering.

Selv om det er en forskjell mellom gruppene så er ikke dette resultatet signifikant da p-verdien til variabelen pålitelighet er på 5,8%. Nullhypotesen kan altså ikke forkastes. Samtidig er p-verdien svært nær signifikansnivået på 5%. Hadde signifikansnivået vært på 10 % kunne altså nullhypotesen blitt forkastet basert på dette resultatet, men det ville da vært høyere

sannsynlighet for feilaktig forkastelse. På grunn av et lite utvalg i eksperimentet er det vanskelig å oppnå signifikante resultater. Det kan derfor tenkes at med et høyere utvalg kunne resultatet vært signifikant.

### **6.1.2 Diskusjon av hvorvidt manglende forklaring fører til at revisor velger å utføre flere revisjonshandlinger**

Hypotese to undersøker hvordan manglende forklaring påvirket deltakerne når det kom til utføring av flere revisjonshandlinger, og her viser resultatene at gruppe to, som ikke fikk forklaring fra KIT, vurderte revisjonsbeviset som 10,6% mindre tilstrekkelig enn gruppe en. Gruppen som ikke fikk forklaring vurderte det som mer nødvendig å utføre ytterligere handlinger utover hva verktøyet hadde gjort. Forskning viser at manglende forståelse øker skepsis, og gruppen som ikke fikk forklaring uttrykte dette gjennom et ønske om å utføre flere revisjonshandlinger slik at de kunne være sikre på at det ikke forelå feil i varelageret som KIT ikke hadde funnet. Gruppe en derimot, fikk en forklaring fra verktøyet på hvordan den har arbeidet, og dette har sannsynligvis redusert deres skepsis da de vurderte dokumentet som mer tilstrekkelig.

Begge gruppene vurderte tilstrekkeligheten til revisjonsbeviset som lav, med et felles gjennomsnitt på 2,91, noe som er like under «middels» grad tilstrekkelig. Selv om gruppe en fikk forklaringen ble de altså ikke helt overbevist om at revisjonsbeviset fra KIT var tilstrekkelig, og de ønsket å gjøre flere revisjonshandlinger. Dette kan forklares ved at de følte at den informasjonen de fikk ikke var nok til å konkludere på regnskapsposten, selv om de fikk mer informasjon enn gruppe to. Hensikten med forklaringen var å gi deltakerne en innsikt i de faktorene som var av størst betydning for verktøyets konklusjon, men de fikk ikke oversikt over alle handlinger KIT hadde gjort. De fikk heller ikke utdypende informasjon om input og prosess, og verktøyet er fortsatt en «black box». Slik at, selv om det er en forskjell mellom gruppene, var gruppe en fortsatt skeptisk til revisjonsbeviset.

I likhet med hypotese en, er ikke forskjellen mellom gruppene her signifikante. P-verdien er 12%, så nullhypotesen kan ikke forkastes.

### **6.1.3 Diskusjon av kontrollvariablene**

Analysene som tok hensyn til kontrollvariablene, alder og arbeidserfaring, resulterte i en økning i signifikansnivået til begge hypotesene. Dette tilsier at forskjellen mellom gruppene

kan være drevet av en, eller begge, av kontrollvariablene. Dette er i tråd med tidligere forskning som har funnet at alder og erfaring kan påvirke teknologiadopsjon ved at eldre revisorer med mer arbeidserfaring er mer skeptiske til ny teknologi.

P-verdien til variabelen alder er 20,6% i hypotese en og 86,7% i hypotese to, og dermed ikke signifikant i noen av hypotesene. Det kan likevel ikke utelukkes at alder har hatt en påvirkning på resultatet, men samtidig var 81,8 % av deltakerne mellom 20-29 år og den største forskjellen i gruppene er ulikheten mellom intervallet 20-24 år og 25-29 år. En person på 29 år vil ikke bli omtalt som å være i det eldre siktet i revisjonsbransjen, og det kan heller ikke sies å være stor forskjell på om en deltaker er 20 eller 29 år, så teorien rundt alder og teknologiadopsjon vil sannsynligvis ikke ha virket inn på eksperimentet. Resultatet vil heller være påvirket av at vi har så lite utvalg at den lille spredningen vi har i alder blir dominerende.

Variabelen arbeidserfaring har en p-verdi på 52% i hypotese en og 5,8% i hypotese to. I hypotese to er denne altså tett opp mot å være signifikant, så dette tyder på at arbeidserfaring kan ha bidratt til at det er en høyere skeptisk holdning i gruppe to. Tidligere forskning har funnet at revisorer med mer arbeidserfaring ikke stolte på digitale verktøy og utførte tradisjonelle revisjonshandlinger i tillegg. Likevel tar også denne forskningen utgangspunkt i et større spenn i arbeidserfaring enn vi har i vårt utvalg. På den andre siden kan det tenkes at kun det faktum om deltakerne har eller ikke har arbeidserfaring kan utgjøre en stor forskjell på hvordan de vurderer videre revisjonshandlinger. De deltakerne uten arbeidserfaring har ikke noe å sammenligne revisjonshandlingene med og kan dermed ha lettere for å vurdere de som hensiktsmessige, men de med arbeidserfaring kan se det opp mot de revisjonshandlingene de er vanligvis utfører og er dermed mer skeptiske hvis handlingene avviker fra dette.

## 6.2 Diskusjon av andre observasjoner

### 6.2.1 Diskusjon av forskjellen mellom vurderingen av pålitelighet og videre revisjonshandlinger

Et annet interessant funn vi gjorde var at tilstrekkelighet ble vurdert lavere enn pålitelighet innen begge gruppene, noe som tilsier at den skeptiske vurderingen hos deltakerne var mindre enn den skeptiske handlingen. Dette kan bety at det er lettere for deltakerne å si at «dette revisjonsbeviset er pålitelig, og kan brukes som bevis» enn det er for de å si at «dette revisjonsbeviset er tilstrekkelig og vi trenger ikke gjøre flere revisjonshandlinger». I



eksperimentet opplever ikke deltakerne tidspress eller krav til at revisjonen skal være effektiv. Dermed vil det ikke ha noen konsekvenser å utføre flere handlinger. Dette kan medføre at det er lettere for deltakerne å si at de vil gjøre noe mer, for det er alltid en risiko for at det er feil som ikke oppdages.

At deltakerne mener at revisjonsbeviset er mindre tilstrekkelig enn pålitelig kan være overraskende, fordi pålitelighet er ansett som det elementet ved revisjonsbeviset som kan være det mest kritiske ved digitale bevis. Tilstrekkeligheten ved beviset burde det derimot ikke være mye tvil om, for maskinen kan gå gjennom langt flere dokumenter og gjøre langt flere handlinger enn revisor. Samtidig er det en utfordring i eksperimentet at deltakerne ikke vet alle handlingene som er utført, og dermed kan det være vanskelig å si at de er tilstrekkelige.

### **6.2.2 Diskusjon om skepsis til revisors arbeidsmetoder er en del av profesjonell skepsis**

Det var et klart flertall av deltakerne som mente at skepsis til revisors egne arbeidsmetoder er en del av profesjonell skepsis. En mulig forklaring til resultatet kan være at deltakerne i undersøkelsen vår er hovedsakelig studenter og uavhengig av arbeidserfaring kan man ikke forvente at de har inngående kunnskaper om hva profesjonell skepsis er. Det kan derfor tenkes at på grunn av manglende alternative begreper å plassere det under, anser deltakerne det som profesjonell skepsis. Dette er en sterk indikasjon på at standardsetterne bør presisere begrepet.

En annen forklaring til resultatet kan være at deltakerne har gjort seg opp en mening om hva profesjonell skepsis er, og dermed at det faller inn under dette begrepet. Det er uenighet blant forskere om hva som er profesjonell skepsis og hvordan det måles, men mange er enige om at det både er et personlighetstrekk hos revisor og noe som avhenger av den revisjonssituasjonen revisor befinner seg i. Hvis profesjonell skepsis er noe som dekkes av revisors personlige egenskaper, tanker og holdninger, kan det også tenkes at dette da inkluderer revisors skepsis til egne metoder. De personlighetstrekkene som blir kartlagt ved for eksempel HPSS skalaen er nettopp personlige egenskaper, uavhengig av revisjonssammenheng. En revisor utøver skepsisen i en profesjonell setting ved å utføre flere revisjonshandlinger, og han opptrer da i en form for profesjonell skepsis. Det ville dermed være rart å diskutere skepsis i en revisjonssammenheng uten å definere det som profesjonell skepsis.

## 6.3 Konklusjon

Oppgaven har som formål å undersøke sammenhengen mellom forståelse for beslutningsprosessen til et verktøy som benytter kunstig intelligens og revisors skeptiske holdning. Vi målte den skeptiske holdningen gjennom hvor pålitelig de vurderte at revisjonsbeviset fra KIT var og om de ville utføre flere revisjonshandlinger. Våre funn indikerer at lavere forståelse vil føre til et høyere nivå av skeptisk holdning. Likevel ga ikke resultatene signifikante verdier, slik at vi ikke kan forkaste nullhypotesene våre. Dermed har vi ikke statistisk grunnlag for å konkludere at det er den manglende forståelsen som gjorde deltakerne mer skeptisk.

I tillegg fant vi at deltakerne anså revisjonsbeviset som mer pålitelig enn tilstrekkelig, dette tilsier at en mindre skeptisk vurdering enn skeptisk handling var tilstede hos deltakerne. Vi fant også at flertallet av deltakerne mente at profesjonell skepsis også omhandler skepsis til revisjonsselskapets arbeidsmetoder, til tross for at dette ikke er spesifisert i standardene.

## 6.4 Begrensninger

Selv om et eksperiment er omtalt som den beste metoden til å teste og forklare sammenhenger, så inneholder et eksperiment også mange begrensninger. En iboende begrensning i et eksperimentelt design er at det nettopp er et eksperiment, og ikke den virkelige verden. Deltakerne kan ha motivasjoner eller insentiver for å svare noe annet enn hva de egentlig mener, eller eksperimentet kan være for dårlig lagt opp slik at man ikke oppnår en forståelse for den helhetlige situasjonen, og at eksperimentet dermed ikke er et godt bilde på virkeligheten. Et eksempel er at deltakerne skulle være revisorer på et oppdrag, men det er ikke mulig å ta i betraktning alle elementer ved et revisjonsoppdrag i eksperimentet. Deltakerne vil ikke kjenne på det tidspresset og kravet til en effektiv revisjon slik revisorer i en revisjonssammenheng vil gjøre, og dermed kan det være lettere for dem å si at de ville utført flere revisjonshandlinger.

En begrensning ved denne undersøkelsen er at den er gjennomført som en masteroppgave. Poenget med oppgaven er å fullføre et eksperiment anvendt på revisjonsbeslutninger, og for å klare det innenfor rammen av en masteroppgave var vi nødt til å kompensere med tanke på antall deltakere og hvilke deltakere som var tilgjengelige. Dette gjør det vanskelig å oppnå signifikante resultater og er dermed en begrensning i seg selv.

En annen begrensning er at vi kun har manipulert én variabel og vi kan ikke utelukke at det er andre variabler som har påvirket resultatene som vi ikke har tatt hensyn til. Et eksempel er at gruppe en fikk mer informasjon, og dermed kan ha vært mer motiverte enn gruppe to. Et annet eksempel er at vi valgte å ikke ta hensyn til skepsis som et personlighetstrekk hos deltakerne. Vi tok heller ikke hensyn til teknologiske ferdigheter eller utdannelsesbakgrunn. Vi vet dermed ikke om deltakerne i den ene gruppen var mer skeptiske av natur eller hadde en lavere grad av teknologikunnskap, og at dette dermed utgjorde den største forskjellen på hvor skeptiske de var, i stedet for forklaringen.

Det var 56 personer som åpnet undersøkelsen vår, men vi oppnådde kun et utvalg på 22 deltakere på grunn av høyt frafall. Det lille utvalget gjorde det vanskeligere å oppnå signifikante resultater, og dermed si noe sikkert. I tillegg vil en lav svarprosent gjøre at vi kan ha fått skjeve resultater, og dermed har vi ikke grunnlag for å si at utvalget er representativt. En lav svarprosent og et lite utvalg gjør at det er høyere sannsynlighet for at de forskjellene vi fant mellom gruppene er tilfeldigheter, og ikke reelle forskjeller forårsaket av den uavhengige variabelen.

Det var 50% som svarte feil på manipulasjonssjekken. Det svekker den interne validiteten til eksperimentet vårt, og er en stor begrensning til resultatene i analysen. Grunnen til at så mange svarte feil kan være at spørsmålet var for detaljert, slik at deltakerne misforsto det, men det kan også være at de ikke mottok manipulasjonen. Uansett grunn gjør den mislykkede manipulasjonssjekken at vi ikke kan si at gruppe en mottok manipulasjonen, og dermed har vi heller ikke grunnlag for å si at det var forståelse som utgjorde forskjellen mellom gruppene.

## 6.5 Anbefalinger til praksis og videre forskningstema

Vi anbefaler at det utføres mer forskning omkring hvordan forklaring fra verktøy som har kunstig intelligens påvirker revisor i beslutningssammenhenger. For eksempel kan det være nyttig å se på hva som er beste mengde og type informasjon fra verktøyet. Det bør også forskes mer på hvordan alder og arbeidserfaring påvirker revisors holdning til den nye teknologien. Med tanke på at det nå skjer et generasjonsskifte i revisjonsbransjen, hvor mange revisorer vil ha vokst opp med mer teknologi, kan det være interessant å se nærmere på effekten dette har.

Vi vil anbefale bruk av forklaring i praksis. Vi vet at det er sannsynlig at kunstig intelligens vil bli tatt mer i bruk i revisjonsbransjen, og vi mener at forklaring kan være viktig når det

kommer til å få til en effektiv revisjon og best mulig samarbeid mellom revisor og maskin. Vi anbefaler også at revisjonsstandardene blir omgjort for å legge mer til rette for bruk av verktøy som inneholder kunstig intelligens. Hvis det i fremtiden blir maskinens skjønn som tar avgjørelser så må standardene være lagt opp til at dette blir gjort på en effektiv måte, samtidig som det sikres at revisjonskvaliteten blir opprettholdt.

Videre bør det forskes på om det er forskjell på revisors skepsis til klienten, og revisors skepsis til egne metoder og verktøyer. Dette kan gi en indikasjon på om revisors skepsis til metoder også er omfavnet av profesjonell skepsis. Standardsetterne bør også presisere hvordan revisor skal forholde seg til revisjonsselskapets verktøy. Dagens løsning med å ikke nevne dette direkte i standardene mener vi at etterlater et stort hull.

---

## Litteraturliste

- Abou-El-Sood, H., Kotb, A., & Allam, A. (2015). Exploring auditors' perceptions of the usage and importance of audit information technology. *International journal of auditing*, 19(3), ss. 252-266. doi:10.1111/ijau.12039
- AICPA. (2015). *Audit analytics and continuous audit: Looking toward the future*. New York: American Institute of certified public accountants.
- Andersen, L. Ø., & Jakobsen, K. U. (2004). *Introduksjon til statistikk med SPSS*. Frederiksberg: Samfundslitteratur.
- Appelbaum, D. (2016). Securing big data provenance for auditors: the big data provenance black box as reliable evidence. *Journal of emerging technologies in accounting*, 13(1), ss. 19-21. doi:10.2308/jeta-51473
- Asklund, A. (2016). Digger digitaliseringen. *Regnskap og revisjon*, 7, ss. 12-14.
- Aurstad, T. (2017). Revisjonsbevis i en digital hverdag. *Regnskap og revisjon*, 7, ss. 26-27.
- Bai, G. h. (2017). Research on the application and influence of auditing artificial intelligence. *International conference on education innovation and economic management*, ss. 245-249. doi:10.12783/dtssehs/eiem2017/16097
- Baldwin, A. A., Brown, C. E., & Trinkle, B. S. (2006). Opportunities for artificial intelligence development in the accounting domain: The case for auditing. *Intelligent systems in accounting, finance and management*, 14, ss. 77-86. doi:10.1002/isaf.277
- Broberg, P., Tagesson, T., Argento, D., Gyllengahm, N., & Mårtensson, O. (2017). Explaining the influence on time budget pressure on audit quality in Sweden. *Journal of management and governance*, 21(2), ss. 331-350. doi:10.1007/s10997-016-9346-4
- Buchholz, V., Kulms, P., & Kopp, S. (2017). It's (not) your fault! Blame and trust repair in human-agent cooperation. *DuEPublico*. doi:10.17185/duepublico/44538
- Davenport, T. H. (2016). The power of advanced audit analytics. Hentet fra <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/deloitte-analytics/us-da-advanced-audit-analytics.pdf>

- 
- Elson, J. S., Derrick, D., & Ligon, G. (2018). Examining trust and reliance in collaborations between humans and automated agents. *In Proceedings of the 51st Hawaii International Conference on System Sciences*, ss. 430-439. doi:10.24251/HICSS.2018.056
- Fischer, M. J. (1996). "Real-izing" the benefits of new technologies as a source of audit evidence: An interpretive field study. *Accounting, organizations and society*, 21(2-3), ss. 219-242. doi:10.1016/0361-3682(95)00005-4
- Fjørtoft, L. E. (2018). Digitalisering og disrupsjon i revisjonsbransjen. *Regnskap og revisjon*, 1, ss. 24-26.
- Francis, J. R. (2011). A framework for understanding and researching audit quality. *Auditing: a journal of practice and theory*, 30(2), ss. 125-143. doi:10.2308/ajpt-50006
- Gerber, A. S., & Green, D. P. (2012). *Field experiments. Design, analysis and interpretation*. New York: W.W. Norton.
- Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2017). When will AI exceed human performance? Evidence from AI experts. *arXiv:1705.08807*.
- Gray, G. L., Chiu, V., Liu, Q., & Li, P. (2014). The expert systems life cycle in AIS research: What does it mean for future AIS research? *International Journal of Accounting Information Systems*, 15, ss. 423-451. doi:10.1016/j.accinf.2014.06.001
- Greenman, C. (2017). Exploring the impact of artificial intelligence on the accounting profession. *Journal of research in business, economics and management*, 8(3), ss. 1451-1454.
- Grenness, T. (2004). *Hvordan kan du vite at noe er sant?* Oslo: Cappelen Damm Akademisk.
- Gulden, B. P. (2016). *Revisjon. Teori og metode*. Oslo: Cappelen Damm.
- Halvorsen, K. (2006). *Å forske på samfunnet: en innføring i samfunnsvitenskapelig metode*. Oslo: Cappelen Damm Akademisk.

- 
- Hampton, C. (2005). Determinants of reliance: An empirical test of the theory of technology dominance. *International Journal of Accounting Information Systems*, ss. 217-240. doi:10.1016/j.accinf.2005.10.001
- Hurttt, R. K. (2010). Development of a scale to measure professional skepticism. *Auditing: a journal of practice & theory*, 29(1), ss. 149-156. doi:10.2308/aud.2010.29.1.149
- Hurttt, R., Brown-Liburd, H., Earley, C. E., & Krishnamoorthy, G. (2013). Research on auditor professional skepticism: literature synthesis and opportunities for future research. *Auditing: a journal of practice & theory*, 32(1), ss. 45-79. doi:10.2308/ajpt-50361
- IAASB. (2009, a). ISA 200 Overordnede mål for den uavhengige revisor og gjennomføringen av en revisjon i samsvar med de internasjonale revisjonsstandardene.
- IAASB. (2009, b). ISA 501 Revisjonsbevis - særlige hensyn knyttet til utvalgte poster.
- Issa, H., Sun, T., & Vasarhelyi, M. A. (2016). Research ideas for artificial intelligence in auditing: The formalization of audit and workforce supplementation. *Journal of Emerging Technologies in Accounting*, 13(2), ss. 1-20. doi:10.2308/jeta-10511
- Jacobsen, D. I. (2005). *Hvordan gjennomføre undersøkelser?* Oslo: Høyskoleforlaget.
- Jensen, M. L., Lowry, P. B., Burgoon, J. K., & Nunamaker, J. F. (2010). Technology dominance in complex decision making: The case of aided credibility assessment. *Journal of management information systems*, 27(1), ss. 175-210. doi:10.2753/MIS0742-1222270108
- Johannessen, A. (2009). *Introduksjon til SPSS*. Oslo: Abstrakt Forlag.
- Kelly, J. E. (2015). Computing, cognition and the future of knowing. Hentet fra [http://www.visionquestit.com/content/aggregate/articles/Computing\\_Cognition\\_WhitePaper.pdf](http://www.visionquestit.com/content/aggregate/articles/Computing_Cognition_WhitePaper.pdf)
- Kokina, J., & Davenport, T. H. (2017). The emergence of artificial intelligence: how automation is changing auditing. *Journal of Emerging Technologies in Accounting*, 14(1), ss. 115-122. doi:10.2308/jeta-51730

- 
- KPMG. (2016). Game changer: The impact of cognitive technology on business and financial reporting. Hentet fra <https://assets.kpmg.com/content/dam/kpmg/pdf/2016/05/game-changer-impact-of-cognitive-technology.pdf>
- Lewis, M., Sycara, K., & Walker, P. (2018). The role of trust in human-robot interaction. *Foundations of trusted anatomy*, 117, ss. 135-159. doi:10.1007/978-3-319-64816-3\_8
- Libby, R. (1981). *Accounting and human information processing: Theory and applications*. Englewood Cliffs, N.J: Prentice-Hall.
- Messier, W. F., & Hansen, J. V. (1987). Expert systems in auditing: The state of the art. *Auditing: A journal of practice & theory*, 7(1), ss. 94-105.
- Morris, M. G., & Venkatesh, V. (2000). Age difference in technology adoption decisions: Implications for a changing work force. *Personnel psychology*, 53(2), ss. 375-403. doi:10.1111/j.1744-6570.2000.tb00206.x
- Mortensen , T., Fisher, R., & Wines , G. (2012). Students as surrogates for practicing accountants: further evidence. *Accounting forum*, 36(4), ss. 251-264. doi:10.1016/j.accfor.2012.06.003
- Nelson, M. W. (2009). A model and literature review of professional skepticism in auditing. *Auditing: a journal of practice & theory*, 28(2), ss. 1-23. doi:10.2308/aud.2009.28.2.1
- Nolder, C. J., & Kadous, K. (2014). *Grounding the professional skepticism construct in mindset and attitude theory: a way forward*. Hentet fra [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2524573](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2524573)
- Nunes, I., & Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27(3-5), ss. 393-444. doi:10.1007/s11257-017-9195-0
- Olsen, C. (2014). Hva er profesjonell skepsis? *Revisjon og regnskap*, 2014(3), ss. 32-37.
- Omoteso, K. (2012). The application of artificial intelligence in auditing: looking back to the future. *Expert systems with applications*, 39(9), ss. 8490-8493. doi:10.1016/j.eswa.2012.01.098



- 
- Pajarinen, M., Rouvinen, P., & Ekeland, A. (2015). Computerization threatens one-third of finnish and norwegian employment. *ETLA Brief*, ss. 1-8.
- Peecher, M. E., & Solomon, I. (2001). Theory and experimentation in studies of audit judgments and decisions: avoiding common research traps. *International journal of auditing*, 5(3), ss. 193-203. doi:10.1111/1099-1123.00335
- PwC. (2018, april 2). Ny revisjons-teknologi forutser fremtiden for bedrifter. Hentet fra <https://www.pwc.no/no/pwc-aktuelt/ny-revisjonsteknologi-forutser-fremtiden-for-bedrifter-.html>
- Quadackers, L., Groot, T., & Wright, A. (2014). Auditors' professional skepticism: neutrality versus presumptive doubt. *Contemporary accounting research*, 31(4), ss. 639-644. doi:10.1111/1911-3846.12052
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ss. 1135-1144. doi:10.1145/2939672.2939778
- Samuel, J. (2018). KPMG applying IBM artificial intelligence to help businesses efficiently meet IFRS 16 lease accounting requirements. Hentet fra <https://home.kpmg.com/xx/en/home/media/press-releases/2018/03/kpmg-applying-ibm-ai-to-help-businesses-meet-ifs-16.html>
- Saunders, M., Lewis, P., & Thornhill, A. (2016). *Research methods for business students*. Harlow: Pearson Education Limited.
- Sutton, S. G., Holt, M., & Arnold, V. (2016). "The reports of my death are greatly exaggerated" - Artificial intelligence research in accounting. *International Journal of Accounting Information System*, 22, ss. 60-73. doi:10.1016/j.accinf.2016.07.005
- Trochim, W. M., Donnelly, J. P., & Arora, K. (2016). *Research methods: the essential knowledge base*. Boston: Cengage learning.
- Trotman, K. T. (2001). Design issues in audit JDM experiments. *International journal of auditing*, ss. 181-192.

- 
- van Dongen, K., & van Maanen, P. P. (2013). A framework for explaining reliance on decision aids. *International Journal of Human-Computer Studies*, 71(4), ss. 410-424. doi:10.1016/j.ijhcs.2012.10.018
- Wang, N., Pynadath, D. V., & Hill, S. G. (2016). Trust calibration within a human-robot team: Comparing automatically generated explanations. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, ss. 109-116.
- Whitecotton, S. M. (1996). The effects of experience and confidence on decision aid reliance: A casual model. *Behavioral research in accounting*, 8, ss. 194-216. doi:10.1006/obhd.1996.0042
- Ye, R., & Johnson, P. E. (1995). The impact of explanation facilities on user acceptance of expert system advice. *MIS Quarterly*, 19(2), ss. 157-172. doi:10.2307/249686

---

## Vedlegg 1 Casetekst og spørsmål

### Vedlegg 1.1 Casetekst

Takk for at du deltar på vår undersøkelse. Den er utarbeidet i forbindelse med vår masterutredning ved Norges Handelshøyskole. Alle besvarelser er anonyme og data vil bli behandlet konfidensielt. Undersøkelsen tar ca. 7 minutter å gjennomføre. Vi setter pris på at du tar deg tid til å lese hele caset og svare på alle spørsmålene.

Året er 2033 og du er revisjonsmedarbeider på revisjonsoppdraget av et selskap med et stort sentrallager. Du er i gang med revisjonen og er i prosessen med å vurdere varelageret. Dette er en vesentlig post i selskapets balanse. En av handlingene du utfører i revisjonen av varelageret er å bruke verktøyet KIT.

KIT er et digitalt verktøy som i samarbeid med en drone bruker kunstig intelligens teknologi til å vurdere varelager. Dette tok revisjonsselskapet ditt i bruk etter at standarden åpnet opp for at det ble tillatt å bruke drone i revisjonen. KIT bruker en kombinasjon av flere teknologier, som maskinlæring, nevralt nettverk, naturlig språk prosessering og dyp læring. Den forstår kontekst og er autonom. Systemet er i en kontinuerlig læringsprosess hvor hver interaksjon og hvert resultat hjelper den til å forbedre seg og lære, og bygger opp en kunnskapsdatabase. Den kan også tolke og organisere tekst, bilder og video.

Dronen er fysisk tilstede ved varetellingen og KIT bruker informasjon fra denne, i tillegg til inputen har du lagt inn, for å vurdere varelageret. For at KIT skal kunne ta en beslutning på riktig grunnlag har du tatt stikkprøver av inputene og kan med betryggende sikkerhet si at den ikke inneholder vesentlig feil.

Etter at varetellingen er over og KIT har gjort en samlet vurdering av posten varelager i regnskapet til enheten sender den dokumentet under til deg. Dette dokumentet er et av revisjonsbevisene du skal ta med i beregningen når du vurderer varelageret.

## Vedlegg 1.2 Revisjonsbevis fra KIT

### Revisjon av varelager

20.01.2033

Basert på en vurdering av eksistens, fullstendig, rettigheter og forpliktelser og verdsettelse har jeg kommet frem til konklusjonen at det er betryggende sikkerhet for at varelageret er uten vesentlige feil.

Av en rekke handlinger jeg har utført i revisjonen av varelageret er dette de faktorene som var av størst betydning.

	Problem	Handling	Resultat
1	Ved observasjon av varetelling gjennom dronen merket jeg at et parti varer ble flyttet	Jeg foretok en kontrolltelling av disse varene, ved bruk av dronen	Kontrolltellingen viste ingen avvik
2	Enheten har fått nye leverandører	Jeg kunne ikke bruke det samme sammenligningsgrunnlaget som tidligere år og valgte derfor å bruke ekstra tid på å gjøre nye analyser	Analysene viste ingen tegn til signifikante feil
3	Ved analyse av video av varetellingen fra dronen fant jeg et vareparti som hadde mulige mangler	Jeg sendte en forespørsel til ledelsen og etterlyste en forklaring til avviket	Ledelsen sendte en forklaring som jeg anser som tilfredsstillende

### Revisjon av varelager

20.01.2033

Basert på en vurdering av eksistens, fullstendig, rettigheter og forpliktelser og verdsettelse har jeg kommet frem til konklusjonen at det er betryggende sikkerhet for at varelageret er uten vesentlige feil.

## Vedlegg 1.3 Spørsmål

### I hvilken grad

	Svært liten grad	Liten grad	Middels	Stor grad	Svært stor grad
Vurderer du revisjonsbeviset som pålitelig?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Har du tillit til at revisjonsbeviset er korrekt?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Er du bekymret for at det kan være vesentlige feil i varelageret?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vektlegger du revisjonsbeviset i din konklusjon om varelageret?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vurderer du revisjonsbeviset som tilstrekkelig?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Anser du det som sannsynlig at du ville utført flere revisjonshandlinger	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### I hvilken grad

	Svært liten grad	Liten grad	Middels	Stor grad	Svært stor grad
Anser du KIT som pålitelig?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Synes du bruk av KIT gir tilstrekkelig bevis?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vurderer du bruk av KIT som en hensiktsmessig revisjonshandling?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Anser du skepsis til revisjonsselskapets arbeidsmetoder som en del av profesjonell skepsis?

- Ja
- Nei

Dokumentet fra KIT viser en tabell over faktorene som var av størst betydning i dens revisjon av varelageret

- Riktig
- Galt

Hvor gammel er du?

- Under 20 år
- 20-24
- 25-29
- 30-34
- 35-40
- Over 40 år

Har du tatt fag innen revisjon? (for eksempel MRR411/BUS426 Revisjon eller lignende)

- Ja
- Nei

Har du arbeidserfaring innen revisjon?

- Nei
- Mindre enn 1 år
- 1-2 år
- 3-4 år
- 5+ år