

Transparensens betydelse för tillit till AI-baserade beslutsstöd i banker: En kvalitativ studie av bankanställdas perspektiv

Av: Saga Adolphsson & Masah Alkurdi

Handledare: Pär-Ola Zander
Södertörns högskola | Institutionen för naturvetenskap, miljö & teknik
Kandidatuppsats 15 hp
Informatik | Höstterminen 2025



Sammanfattning

Den ökade användningen av artificiell intelligens (AI) som beslutsstöd inom banksektorn har medfört nya möjligheter till effektivisering och datadrivet beslutfattande. Samtidigt har utvecklingen aktualiserat frågor om transparens, ansvar och tillit kopplat till AI som beslutsstöd. Eftersom banker verkar i en reglerad miljö där förtroende utgör en grundläggande förutsättning för verksamheten, blir det centralt att förstå hur bankanställda upplever och förhåller sig till AI-baserade beslutsstöd i sitt dagliga arbete. Mot denna bakgrund är syftet med studien att undersöka hur transparens i AI-baserade beslutsstöd påverkar bankanställdas upplevelse av tillit i beslutfattande. Studien genomförs med en kvalitativ forskningsansats och bygger på semistrukturerade djupintervjuer med bankanställda som använder eller har erfarenhet av AI som stöd i beslutsprocesser. Datamaterialet analyseras genom tematisk analys med deduktiv och induktiv utgångspunkt. Analysen tar sin utgångspunkt i socioteknisk systemteori, arbetsystemteori samt etablerade teorier om tillit och transparens, men även teman som framkommer under empirin. Resultaten visar att tillit till AI-baserade beslutsstöd inte enbart kan förstås utifrån teknikens prestanda, utan formas i ett sociotekniskt samspel mellan tekniska system, mänskliga aktörer och organisatoriska strukturer. Transparens framträder som en central förutsättning för tillit, men tar sig olika uttryck genom (1) insyn i AI:s beslutslogik, (2) förståelse som byggs genom interaktion och (3) organisatorisk tydlighet kring ansvar och användning. Studien visar även att tilliten både är situationsbunden och till viss del erfarenhetsbaserad, samt att mänskligt ansvar och professionellt omdöme utgör en tydlig gräns för hur långt tilliten till AI kan sträcka sig i praktiken.

Nyckelord: Artificiell intelligens, AI-baserade beslutsstöd, Transparens, Tillit, Beslutfattande, Banksektorn.

The role of transparency in trust in AI-based decision support systems in banks: A qualitative study of bank employees' perspective

Abstract

The increased use of artificial intelligence (AI) as decision support within the banking sector has led to new opportunities for efficiency and data-driven decision-making, while simultaneously raising questions regarding AI transparency, how responsibility is structured, and how trust is achieved. As banks operate in a regulated environment where trust constitutes a fundamental prerequisite for their operations, it becomes central to understand how bank employees experience and relate to AI-based decision support in their daily work. Against this background, the aim of the study is to examine how transparency in AI-based decision support affects bank employees' experience of trust in decision-making. The study is conducted using a qualitative research approach and is based on semi-structured in-depth interviews with bank employees who use or have experience of AI as support in decision-making processes. The empirical material is analyzed through thematic analysis with a deductive and inductive point of departure, where the analysis is based on sociotechnical systems theory, work system theory, as well as established theories of trust and transparency, but also themes that emerge from the empirical material. The results show that trust in AI-based decision support cannot be understood solely on the basis of the technology's performance, but is shaped in a sociotechnical interplay between technical systems, human actors, and organizational structures. Transparency emerges as a central prerequisite for trust but takes different forms through (1) insight into AI decision logic, (2) understanding built through interaction, and (3) organizational clarity regarding responsibility and use. The study also shows that trust is both situational and to some extent experience-based, and that human responsibility and professional judgment constitute a clear boundary for how far trust in AI can extend in practice.

Key terms: Artificial Intelligence, AI-based Decision Support, Transparency, Trust, Decision-Making, Banking Sector.

Förord

Vi vill rikta ett särskilt tack till vår handledare Pär-Ola för värdefull vägledning, konstruktiv återkoppling och stöd under arbetets gång. Dina insikter har varit av stor betydelse för uppsatsens kvalitet och utveckling.

Vi vill också tacka våra opponenter samt examinator för insiktsfulla synpunkter som bidragit till att förtydliga och stärka uppsatsen.

Slutligen vill vi även tacka de respondenter som deltagit i studien och generöst delat med sig av sin tid och sina erfarenheter. Er medverkan har varit avgörande för genomförandet av denna studie.

Saga Adolphsson och Masah Alkurdi

Innehållsförteckning

SAMMANFATTNING	2
ABSTRACT	4
FÖRORD	5
INNEHÅLLSFÖRTECKNING	6
DEFINITIONER	9
AKRONYMER	10
1. INTRODUKTION	11
1.1 Bakgrund	11
1.2 Problemformulering	13
1.3 Syfte och forskningsfråga	16
2. TEORI	17
2.1 Socioteknisk systemteori	17
2.2 Tillitsteori	18
2.3 Arbetssystemteorin	19
2.3.1 Arbetssystemramverket	20
2.3.2 Arbetssystemlivscykelmodellen	23
2.4 AI:s framkant	24
2.5 Transparens som villkor för tillit till AI	26
2.5.1 Transparens som förutsättning till tillit	26
2.5.2 Olika dimensioner av transparens	27
2.6 Förklarbar AI	28
2.7 Deduktivt konceptuellt ramverk	30
3. METOD	33
3.1 Metodval	33
3.2 Urval	34

3.3 Datainsamlingsmetod	34
3.4 Analysmetod	37
3.5 Forskningsetiska överväganden	38
3.6 Metodkritik	39
4. DATAINSAMLING	41
4.1 Respondenter	41
4.2 Översikt över datainsamlingens innehåll	43
4.2.1 Socioteknisk systemteori	44
4.2.2 Tillitsteori	45
4.2.3 Arbetssystemteorin	47
4.2.4 AI:s framkant	47
4.2.5 Tillit och transparens	48
4.2.6 Förklarbar AI	50
5. ANALYS	52
5.1 Transparensens betydelse för tillit till AI som beslutsstöd	52
5.1.1 Insyn och transparens i AI:s beslutslogik	53
5.1.2 Transparens genom interaktion med AI	55
5.1.3 Organisationens roll kopplat till transparens och tillit	58
5.2 Erfarenhetsbaserad, situationsbunden och begränsad tillit till AI som beslutsstöd	61
5.2.1 Erfarenhetsbaserad tillit	61
5.2.2 Tillit som situationsbunden och begränsad	63
5.3 Mänskligt ansvar som gräns för tillit till AI som beslutsstöd	66
5.3.1 Professionellt ansvar som grund för beslutsfattande	66
5.3.2 Förklarbarhet som förutsättning för professionellt omdöme	68
6. DISKUSSION	71
7. SLUTSATS	77
8. FRAMTIDA FORSKNING	79
KÄLLFÖRTECKNING	80
BILAGOR	83
Bilaga 1: Intervjuguide	83
Bilaga 2: Deduktiva koder	87

Bilaga 3: Induktiva koder	89
Bilaga 4: Kategorier och teman	90
Bilaga 5: Missivbrev	91

Definitioner

Nedan presenteras centrala begrepp som används i studien. Dessa definieras och avgränsas för att säkerställa en tydlig och konsekvent användning inom studien.

Artificiell intelligens (AI): Stryker och Kavlakoglu (u.å.) beskriver artificiell intelligens som en teknik som gör det möjligt för datorer och maskiner att simulera mänskligt lärande, förståelse, problemlösning, beslutsfattande, kreativitet och autonomi.

AI-baserade beslutsstöd: AI-baserade beslutsstöd syftar till AI som används för att analysera information och generera rekommendationer som stöd i operativa och strategiska beslut inom organisationer (Kovari 2024, s. 1).

Transparens: Transparens innebär att AI-system utformas på ett sätt som möjliggör spårbarhet och förståelse, att människor informeras när de interagerar med AI samt att berörda aktörer får tydlig information om systemens förmågor, begränsningar och sina rättigheter (European Union 2024).

Förklarbar AI (XAI): Förklarbar AI syftar till metoder samt tekniker som gör AI:s beslut begripliga, spårbara och motiverade för användare (Dwivedi et al. 2023, s. 194:1).

Tillit: Tillit definieras som en vilja att vara sårbar inför tekniken baserat på förväntningar om att den agerar på ett tillförlitligt sätt. Tilliten bygger på de tre dimensionerna som är teknikens förmåga, välvilja och integritet (Söllner, Benbasat, Gefen, Leimeister & Pavlou 2016, s. 2).

AI:s framkant: Berente, Gu, Recker och Santhanam (2021, s. 1438) beskriver AI:s framkant (AI-frontier) som den yttersta gränsen för datorsystemets utveckling, där tekniska framsteg gör det möjligt för AI att hantera alltmer komplexa beslutssituationer.

Arbetsystem: Enligt Alter (2013, s. 75) består ett arbetsystem av människor, teknik, information och organisatoriska strukturer som samverkar för att producera värde.

Akronymer

Förkortningar och akronymer som används i studien förklaras för att underlätta förståelsen och säkerställa en konsekvent användning genom hela arbetet.

AI: Artificiell intelligens.

IT: Informationsteknik.

IS: Informationssystem.

XAI: Förklarbar artificiell intelligens (explainable AI).

WST: Arbetsystemteorin (work system theory).

1. Introduktion

Denna studie tar utgångspunkt i hur den ökade användningen av artificiell intelligens (AI) som beslutsstöd påverkar bankanställdas beslutsfattande i en reglerad bransch. Den inledande bakgrunden belyser hur digitalisering och AI successivt integreras i banksektorns arbetsprocesser, vilket aktualiserar frågor om transparens, ansvar och tillit. Därefter formuleras studiens problemområde med fokus på de spänningar som uppstår när AI används som stöd för beslutsfattning inom banker kopplat till transparens och tillit. Följaktligen presenteras studiens syfte samt forskningsfråga och dessa delar utgör tillsammans utgångspunkten för studiens fortsatta teoretiska och empiriska analys.

1.1 Bakgrund

Den snabba digitaliseringen har omformat både samhälle och näringsliv i grunden. Drivkrafter som ökad datatillgång, global uppkoppling samt tekniska framsteg inom molntjänster, stora datamängder (big data) och sakernas internet (internet of things, IoT) har bidragit till nya möjligheter för organisationer gällande att effektivisera processer och fatta mer datadrivna beslut. Utvecklingen har således skapat ett växande behov av mer avancerade verktyg som kan stödja komplexa och snabbföränderliga beslutsprocesser. (Valacich, Schneider & Hashim 2022, s. 41). Samtidigt menar Polinati et al. (2025, s. 322) att artificiell intelligens i allt högre grad används för att förstärka och utveckla traditionella beslutsstödsystem genom att analysera stora datamängder, identifiera mönster och generera relevanta rekommendationer för mänskliga beslutsfattare.

Mot denna bakgrund definierar Valacich, Schneider och Hashim (2022, s. 281) artificiell intelligens som en teknik som kan simulera mänsklig intelligens genom att lära sig, resonera och fatta beslut. Stryker och Kavlakoglu (u.å.) beskriver AI på ett liknande sätt, och framhåller vidare att AI möjliggör för datorer och maskiner att simulera mänskligt lärande, förståelse, problemlösning, beslutsfattande, kreativitet och autonomi. Den ökade användningen av AI innebär därför inte enbart tekniska innovationer, utan även förändringar i arbetsroller, beslutsprocesser samt organisatoriska strukturer, vilket ställer nya krav på ledarskap, etik och digital kompetens. (Valacich, Schneider & Hashim 2022, s. 42-43).

Följaktligen har AI-tillämpningar inom finanssektorn visat sig stärka precisionen i beslutsfattandet, minska beroendet av felaktiga informationskällor och bidra till mer effektiva

beslutsprocesser. AI-drivna beslutsstödsystem är dessutom särskilt värdefulla i branscher där beslut är tidskänsliga och kräver omfattande informationsunderlag - som exempelvis finanssektorn. (Polinati et al. 2025, s. 322). Samtidigt innebär den ökade användningen av AI en förskjutning i arbetsroller och ansvar eftersom mänskligt omdöme i allt högre grad kompletteras av algoritmiska beslutsystem. Denna förändring väcker frågor om tillit till tekniken, särskilt inom banksektorn där beslutsfattande präglas av höga krav på tillförlitlighet och regulatorisk efterlevnad (Islam et al. 2025).

Vidare visar utvecklingen inom den svenska finanssektorn tydligt hur AI håller på att bli en integrerad del av bankernas verksamhet. Enligt Finansinspektionens (2024b) rapport "*Artificiell intelligens i den svenska finanssektorn*" använder, eller planerar en majoritet av de svenska finansinstituten att använda, AI-lösningar inom områden som riskhantering, kreditprövning, bedrägeriövervakning och kundtjänst. Rapporten visar fortsättningsvis att många aktörer ännu befinner sig i ett inledande skede där tekniken används i begränsad omfattning och ofta utan en fullt utvecklad styrnings- och kontrollstruktur (Finansinspektionen 2024b). Finansinspektionen (2024b) betonar därför vikten av att införa robusta processer för datakvalitet, modellriskhantering och intern styrning för att säkerställa ett ansvarsfullt användande av AI. Bristande transparens, otillräcklig dokumentation och svag förståelse för algoritmers beslutslogik kan enligt myndigheten medföra risker som påverkar både konsumentskydd och finansiell stabilitet (Finansinspektionen 2024b).

För att upprätthålla allmänhetens och medarbetarnas förtroende krävs därför tydliga ansvarsfördelningar, etiska riktlinjer och transparens i hur AI används i beslutsprocesser - särskilt inom banker där kundens ekonomiska trygghet står i direkt fokus (Finansinspektionen 2024b). Vidare visar undersökningen från Finansinspektionen (2024b) att 84 % av de svarande finansinstituten uppger att deras anställda använder någon form av generativ AI, medan endast 22 % har AI-system i produktion eller under utvecklingsprocess. Rapporten pekar särskilt på att datakvalitet samt dataskydd utgör betydande risker och att endast 41 % av företagen med AI i produktion har en formellt godkänd policy gällande AI-användning (Finansinspektionen 2024b).

Detta indikerar att tillit till AI inte kan förstås isolerat som en teknisk fråga, utan även är beroende av de organisatoriska och styrningsmässiga förutsättningar som formar hur tekniken används i bankverksamheten. Samtidigt framhåller Alter (2013) att informationssystem utgör

arbetssystem där teknik, människor, information och arbetsprocesser är tätt sammanlänkade och tillsammans bidrar till värdeskapande. Ur detta perspektiv framstår AI-baserade beslutsstöd inte enbart som tekniska verktyg, utan som en del av ett arbetssystem som formas genom organisatoriska strukturer, arbetsroller och styrningsmekanismer. Med andra ord behöver de tekniska lösningarna förstås i relation till mänskliga och organisatoriska sammanhang (Alter 2013).

Samtidigt visar forskning att transparens är nära kopplat till tillit och spelar en avgörande roll för hur människor uppfattar och accepterar teknik. Detta beskriver Yu och Li (2022) som menar att transparens dels kan öka förståelsen för AI och därmed stärka tillit, dels skapa obehag när insynen i tekniken upplevs som överdriven. De betonar därför vikten av en balanserad transparens, där öppenhet kombineras med begriplighet för att främja hållbar tillit till AI. Denna balans är särskilt viktig inom banksektorn, där beslutfattande både måste vara effektivt samt upplevas som rättvist och etiskt. Maier, Menold och McComb (2022) visar vidare att låg transparens i AI-system resulterar i att människor i högre grad förlitar sig på egna antaganden och intuition i stället för på systemets faktiska beslutsgrund. Detta innebär att samma typer av kognitiva snedvridningar som präglar mänskligt beslutsfattande också påverkar hur AI används, vilket understryker vikten av transparens för kritisk granskning av beslutsunderlag och möjliggöra välgrundade bedömningar.

I relation till detta menar Berente et al. (2021) att en av de största utmaningarna för att skapa tillit till AI är användarnas begränsade förståelse för hur AI fattar beslut. När AI som beslutsstöd upplevs som svårt att överblicka, blir transparens och förklarbarhet centrala förutsättningar för att tekniken ska uppfattas som tillförlitlig. Sammantaget belyser detta en socioteknisk utmaning där tekniska system får en allt större roll i beslutsprocesser, vilket aktualiserar frågor om ansvar, förståelse och tillit.

1.2 Problemformulering

Valacich, Schneider och Hashim (2022, s. 38) menar att den ökade digitala tätheten, det vill säga den snabbt växande mängden uppkopplad data per aktivitet, har resulterat i nya affärsmodeller. Denna utveckling blir särskilt tydlig inom finanssektorn, där framväxten av tekniska lösningar som AI används för att automatisera och effektivisera finansiella tjänster (Valacich, Schneider & Hashim 2022, s. 281).

Även Finansinspektionen (2024a;2024b) framhåller att AI i allt högre grad används som ett centralt verktyg inom finanssektorn, främst för att söka eller sammanfatta information, processautomatisering, kundinsikter samt kundsupport, och att ett flertal verksamheter planerar att investera i ytterligare AI-tekniker. I detta sammanhang kan AI förstås som en dynamisk gräns inom databehandling som ständigt flyttas fram i takt med att maskiner tar över alltmer komplexa kognitiva uppgifter som tidigare krävde mänskligt omdöme. Detta innebär att AI inte enbart fungerar som ett informationshanterande verktyg, utan även som ett verktyg i beslutsprocesser, där AI kan påverka hur beslut formas och genomförs. (Berente et al. 2021, s. 1434-1436).

Följaktligen aktualiserar den ökade användningen av AI frågan om hur en trygg och tillförlitlig användning kan säkerställas, då banker ställs inför utmaningar kopplade till exempelvis datakvalitet och dataskydd (Finansinspektionen 2024b, s. 18). Samtidigt konstaterar Finansinspektionen (2024b, s. 3) att befintliga riktlinjer och regleringar riskerar att inte hålla jämn takt med den tekniska utvecklingen. Detta fenomen beskriver även Mittelstadt (2019, s. 2) som identifierar fyra centrala svagheter inom AI:s utveckling, det vill säga brist på gemensamma mål och ansvarsskyldighet, avsaknad av en etablerad yrkestradition, svårigheter att omsätta etiska principer i praktiken samt otillräckliga rättsliga och professionella ansvarsmekanismer. En hållbar tillit till AI kräver därför tydliga regler, fungerande ansvarssystem och en genuin tillämpning av etiska principer (Mittelstadt 2019, s. 9). Om detta uteblir, kan det innebära att berörda aktörer inte kan vara säkra på att utvecklare agerar i deras bästa intresse, vilket i sin tur riskerar att minska tilliten till AI (Mittelstadt 2019, s. 3-4).

Detta är problematiskt eftersom tidigare forskning visar att tillit utgör en grundläggande förutsättning för social interaktion och samarbete mellan människor och teknik (Söllner et al. 2016). Samtidigt är brist på tillit ett av de största hindren för verksamheter när det kommer till att fullt ut kunna dra nytta av AI (Gillath et al. 2020, s. 1). Detta blir extra tydligt inom banksektorn där tillit är särskilt avgörande eftersom banker förvaltar kunders ekonomiska tillgångar, vilket ställer extra höga krav på att AI-system ska vara tillförlitliga, tillgängliga och säkra (Lazo & Ebarido 2023, s. 65).

Mot denna bakgrund blir det nödvändigt att reflektera över vad tillit egentligen innebär och hur en gemensam förståelse för AI skapas. I relation till detta belyser Haresamudram, Larsson

och Heintz (2023, s. 94) att transparens utgör en central komponent i tillförlitlig AI. Betydelsen av transparens förstärks dessutom av den nya EU-förordningen om artificiell intelligens (AI Act), som etablerar ett riskbaserat regelverk för utveckling och användning av AI (European Commission 2025). Enligt dessa riktlinjer klassas AI-system som används för att utvärdera kreditvärdighet och andra finansiella beslutsprocesser som hög risk, vilket medför krav på spårbarhet, förklarbarhet, robust datakvalitet och mänsklig övervakning (European Commission 2025). För banker innebär detta att transparens inte längre enbart är en organisatorisk eller etisk ambition, utan ett krav, samtidigt som Finansinspektionens (2024b, s. 3) rapport visar att många verksamheter ännu inte har tillräckliga strukturer för att uppfylla dessa krav. Detta aktualiserar behovet av att förstå huruvida bankanställda i praktiken upplever transparens och tillit kopplat till AI-baserade beslutsstöd i arbetet.

Med andra ord framträder transparens som en central faktor i relation till tillit till AI. Å ena sidan visar Yu och Li (2022) att hög transparens inom AI dels kan öka tilliten genom ökad förståelse och upplevd effektivitet, dels skapa obehag och därmed minska tilliten. Detta bekräftas av Jiang, Li och Liu (2025, s. 13) som framhåller att för mycket transparens kan hämma det kognitiva tänkandet och därigenom ha negativa konsekvenser.

Samtidigt menar Berente et al. (2021, s. 1441-1442) att användningen av AI i verksamheter kräver öppenhet och förklarbarhet för att främja förtroende och användning. Detta bekräftas även av Kovari (2024, s. 1) som beskriver att transparens är avgörande för att användarna ska känna tillit till AI. Rai (2020, s. 137) betonar vidare att bristande förståelse för hur AI fattar beslut kan leda till minskat förtroende, vilket illustrerar de kontrasterande forskningsresultaten om hur transparens påverkar tillit till AI. Samtidigt påpekar Alter (2006, s. 73-74) att många IT-initiativ misslyckas, där förståelse för verksamhetens behov och användarnas förväntningar kan vara en anledning. Otydliga mål, svag kommunikation och avsaknad av gemensam förståelse för hur systemen ska skapa nytta är återkommande orsaker till dessa misslyckanden (Alter 2006, s. 74).

Sammantaget visar detta att transparensens roll i relation till tillit till AI-baserade beslutsstöd är komplex, särskilt ur bankanställdas perspektiv där besluten får konsekvenser för kunder och organisation. Bristande förståelse för hur transparens bör utformas riskerar, i linje med Alter (2006), inte bara att minska användarnas tillit, utan även att äventyra verksamhetens effektivitet och förmåga att dra nytta av tekniken. Detta väcker frågor om hur banker i praktiken

hanterar denna balans och hur anställda upplever samt förhåller sig till AI som beslutsstöd i det dagliga arbetet.

Därmed finns det kontrasterande uppfattningar om hur transparens påverkar tillit, vilket innebär att frågan kring transparens och ansvar är central. Det råder därmed ett forskningsgap kring hur transparens påverkar bankanställdas tillit till AI-baserade beslutsstöd, vilket är problematiskt i en bransch där förtroende utgör en grundläggande förutsättning för verksamheten. Det finns således ett behov av att empiriskt undersöka hur bankanställda upplever transparens och hur de utvecklar samt upplever tillit till AI som beslutsstöd. Då AI används i olika arbetsystem inom bankverksamheten blir det centralt att förstå hur transparens uppfattas i dessa olika kontexter. Studien fokuserar därför på hur transparens och tillit formas inom de specifika arbetssystem där AI fungerar som beslutsstöd, snarare än som autonomt beslutssystem.

1.3 Syfte och forskningsfråga

Syftet med denna studie är att undersöka hur transparens i AI-baserade beslutsstöd påverkar bankanställdas upplevelse av tillit till verktyget. Genom att analysera hur transparens upplevs och tolkas i mötet mellan människa och AI, belyser studien hur graden av insyn i AI-baserade beslutsprocesser påverkar tillit, upplevd trygghet samt ansvar i beslutsfattandet.

Forskningsfrågan för studien är därför:

“Hur uppfattar bankanställda transparens i AI-baserade beslutsstöd och hur påverkar denna upplevda transparens deras tillit till AI?”.

2. Teori

För att förstå hur bankanställda utvecklar tillit till AI-baserade beslutsstöd krävs teoretiska perspektiv som fångar både tekniska egenskaper samt de organisatoriska och professionella sammanhang där besluten fattas. Studien tar därför utgångspunkt i forskning om tillitsteori, transparens, förklarbarhet, socioteknisk systemteori samt arbetssystemteori.

2.1 Socioteknisk systemteori

Den sociotekniska teorin utgår från att teknik, människor och organisationer utgör ömsesidigt beroende delar av ett sammanhängande system (Alter 2013, s. 76). Alter (2013) betonar vidare att informationssystem bör förstås som arbetssystem där människor, tekniska artefakter, information, aktiviteter och organisatoriska processer tillsammans skapar värde. Ett arbetssystem omfattar därmed både tekniska och sociala element vilket innebär att informationssystem aldrig kan reduceras till enbart programvara eller hårdvara (Alter 2013, s. 78-80). Tekniken påverkar hur arbetet utförs, men formas samtidigt av de aktörer, normer, regler och maktrelationer som finns i organisationen (Lamb & Kling 2003). Syftet med ett sociotekniskt synsätt är därför att förstå hur sociala och tekniska komponenter samspelar i praktiken, snarare än att enbart se tekniken som ett verktyg för effektivisering (Alter 2013).

Lamb och Kling (2003) utvecklar denna tanke genom att kritisera den traditionella bilden av "användaren" inom informationssystemforskningen. De menar att användare inte bör förstås som en passiv mottagare av tekniken, utan som en social aktör vars användning av IT formas av yrkesidentitet, organisatoriska roller, institutionella villkor och externa krav (Lamb & Kling 2003, s.197-199). Informationssystem är därmed inte endast tekniska verktyg utan sociotekniska konstruktioner som både speglar och formar organisationens kultur, ansvarsfördelning och maktstruktur (Lamb & Kling 2003, s. 218-219). För att förstå effekterna av digital teknik, som AI, krävs därför en analys av hur tekniken omformar sociala interaktioner samt hur sociala och organisatoriska faktorer påverkar teknikens användning och den upplevd tilliten.

Även Alter (2013, s. 76-77) förespråkar denna grundidé gällande socioteknisk systemteori, där han framhåller att alla arbetssystem i praktiken är sociotekniska system, eftersom tekniska och mänskliga komponenter är sammanflätade. Enligt Alter (2013, s. 91) är det därför inte

tillräckligt att analysera tekniken i sig, utan det krävs också en förståelse för hur den används, övervakas och anpassas i samspel med organisatoriska mål och mänskligt handlande.

Tillämpat på banksektorn innebär detta att AI-baserade beslutsstöd inte enbart ska förstås som tekniska innovationer, utan som sociotekniska förändringar som påverkar hela beslutsmiljön. När AI används som beslutsstöd förändras därför inte bara tekniken inom banken, utan även arbetsroller, ansvar och tillit.

2.2 Tillitsteori

Inom forskningsfältet ledningsinformationssystem (Management Information Systems, MIS) betraktas tillit som en central mekanism för att förstå relationer mellan människor, organisationer och tekniska system (Söllner et al. 2016). Enligt Söllner et al. (2016) definieras tillit som en vilja att vara sårbar inför en annan part, som exempelvis teknik, baserat på en förväntan om att den agerar på ett tillförlitligt sätt. Författarna betonar att tillit inom informationssystem inte enbart är en psykologisk egenskap, utan en nödvändig förutsättning för samverkan mellan människa och teknik (Söllner et al. 2016).

Vidare bygger den teoretiska förståelsen av tillit på tre dimensioner, det vill säga förmåga (ability), välvilja (benevolence) och integritet (integrity), som beskriver användarnas uppfattning om teknikens förmåga, dess intention att stödja snarare än att skada och därmed välvilja, samt dess etiska och pålitliga beteende. Tillsammans förklarar dessa dimensioner varför människor väljer att förlita sig på tekniska system trots inneboende risker och informationsasymmetrier. (Söllner et al. 2016).

Samtidigt visar Tomkins (2001) att tillit fungerar som ett alternativ till detaljerad informationskontroll där graden av tillit påverkar behovet av information i relationer och samarbeten. När tillit ökar, minskar behovet av kontrollinformation, medan brist på tillit skapar ett större behov av övervakning och transparens. Tomkins (2001) framhåller vidare att tillit och informationsflöden är ömsesidigt beroende av varandra, där information både kan bygga och undergräva tillit beroende på hur den används. Sammantaget beskriver dessa teorier tillit som en dynamisk och relationell process som balanserar sårbarhet, information och förtroende.

Inom informationssystemforskning har även begreppet tillit till teknik (trust in technology) fått en framträdande roll. McKnight, Carter, Bennett Thatcher och Clay (2011, s. 12:1) menar att tillit till teknik har en central funktion, eftersom den påverkar människors attityder och beteenden gentemot tekniska system. Tidigare forskning har främst fokuserat på tillit till människor eller organisationer, medan nyare studier visar att även tillit till själva tekniken är avgörande (McKnight et al. 2011, s. 12:2). Denna forskning vidareutvecklas av McKnight et al. (2011, s. 12:15), som framhåller att när tillit etableras, både till människor och till teknik, bedömer individer huruvida den aktör eller det system de är beroende av besitter de egenskaper som krävs för att fungera som förväntat i en given situation.

Författarna beskriver vidare hur erfarenhetsbaserad tillit (knowledge-based trust) utvecklas genom faktisk användning av tekniken (McKnight et al. 2011, s. 12:15). När individen har byggt upp en erfarenhetsbaserad tillit uppstår vad författarna kallar tillit till en specifik teknik (trusting beliefs in a specific technology), vilket innebär att tilliten grundas på upplevda erfarenheter av teknikens funktionalitet samt prestanda, och denna tillit är situationsbunden. Den erfarenhetsbaserade tilliten omfattar tre centrala aspekter, det vill säga att tekniken uppfattas som funktionell (kan utföra uppgifterna), hjälpsam (stödjer individen effektivt) och tillförlitlig (fungerar stabilt och förutsägbart). (McKnight et al. 2011, s. 12:9-12:14).

Tillit utgörs därmed av uppfattningar om teknikens förmåga att leverera utifrån sitt syfte och dessa bedömningar formas inte enbart av teknikens objektiva egenskaper, utan påverkas också av individens erfarenheter och kontexten där tekniken används. I likhet med tillit till människor grundas tillit till teknik således på situationsbundna uppfattningar om dess funktionalitet och tillförlitlighet. (McKnight et al. 2011, s. 12:15). Tillsammans bidrar dessa perspektiv till en helhetsförståelse av tillit som både ett socialt och tekniskt fenomen inom informationssystem. Vidare tyder detta på att användarnas uppfattningar om AI:s kompetens, intentioner och pålitlighet är avgörande för hur tekniken accepteras och används i praktiken, men även för hur behovet av transparens ser ut.

2.3 Arbetssystemteorin

Enligt Alter (2013) bidrar teorin gällande arbetssystem (work system theory) med ett ramverk för att förstå hur människor och teknik samverkar inom organisatoriska system. Alter (2013, s. 75) beskriver att ett arbetssystem är ett system där mänskliga aktörer samt maskiner utför

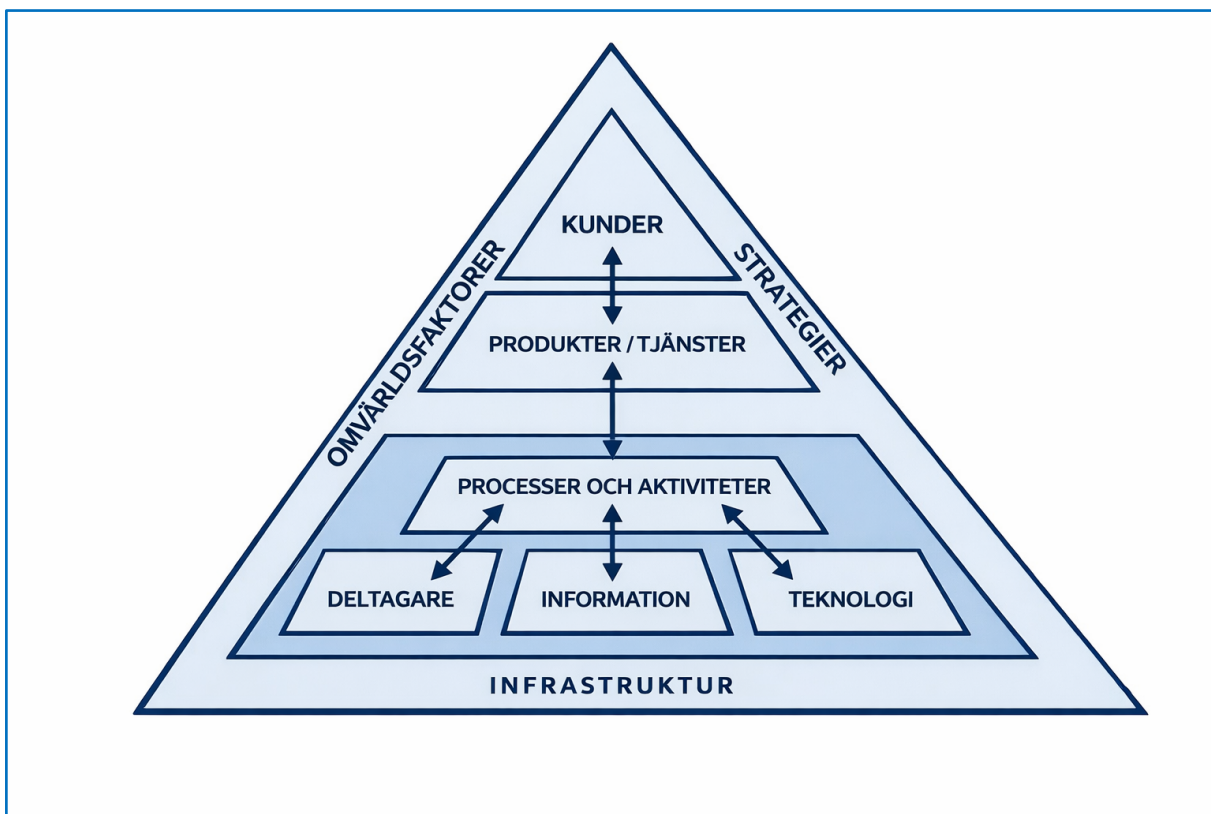
arbete, i form av processer eller aktiviteter, med hjälp av information, teknik eller andra resurser där syftet är att producera specifika produkter eller tjänster till interna samt externa mottagare.

Vidare hävdar Alter (2013) att arbetssystem betraktas som sociotekniska system där människor och teknik samverkar i organisatoriska processer. Detta synsätt skiljer sig från den traditionella förståelsen inom IT, där system tidigare främst betraktats som tekniska verktyg, snarare än som en del av en social helhet (Alter 2013, s. 76). Således innebär detta att arbetet inte enbart kan förstås som en teknisk process, utan måste ses i relation till samspelet mellan människor, teknik, information och organisatoriska mål. Alter (2013) beskriver att alla delar i arbetssystemet (människor, processer, teknik och information) behöver vara i balans och anpassade till målen. Om detta inte sker, kan det i stället resultera i ineffektivitet och motstånd till tekniken (Alter 2013, s. 76).

Följaktligen beskriver Alter (2013, s. 77) att arbetssystemteorin inkluderar både ett statiskt och ett dynamiskt perspektiv på arbetssystem och dessa är arbetssystemramverket samt arbetssystemslivecykelmodellen. Arbetssystemteorin används i denna studie för att analysera hur AI-baserade beslutsstöd påverkar samspelet mellan deltagare, processer, teknik och information. Inom detta ramverk kan transparens förstås som en egenskap hos informationskomponenten (se tabell 1) det vill säga hur begriplig och spårbar systemets presenterade information är. Samtidigt visar McKnight et al. (2011, s. 12:1) att tillit till tekniken kan påverka deltagarnas interaktion med tekniska lösningar. Med andra ord kan detta få konsekvenser för hur arbetssystemet fungerar i praktiken som konsekvens av användarnas förmåga att förlita sig på, ifrågasätta eller komplettera AI:s rekommendationer.

2.3.1 Arbetssystemramverket

Alter (2013, s. 77) beskriver att arbetssystemramverket (work system framework) (se figur 1), som bidrar med det statiska perspektivet, visar hur arbetssystemet ser ut under en stabil period och hur dess delar samverkar. Detta ramverk bidrar med förståelse kring hur arbete faktiskt utförs i en organisation, särskilt när det bygger på teknik som stödjer olika processer (Alter 2013, s. 79). Modellen visar att ett arbetssystem består av flera delar, det vill säga människor, teknik, information och aktiviteter, som tillsammans skapar produkter eller tjänster till kunder (Alter 2013, s. 79).



Figur 1: Egen bearbetning av arbetssystemramverket baserad på Alter (2013, s. 79).

Följaktligen behöver samtliga delar i arbetssystemet hänga ihop och vara i balans. Deltagarnas kunskap, motivation och roller behöver vara anpassade till de arbetsprocesser samt de tekniska verktyg som används. Som resultat av detta kan förändringar i någon av komponenterna medföra behov av justeringar i arbetssätt, kompetens eller ansvarsfördelning. (Alter 2013, s. 79).

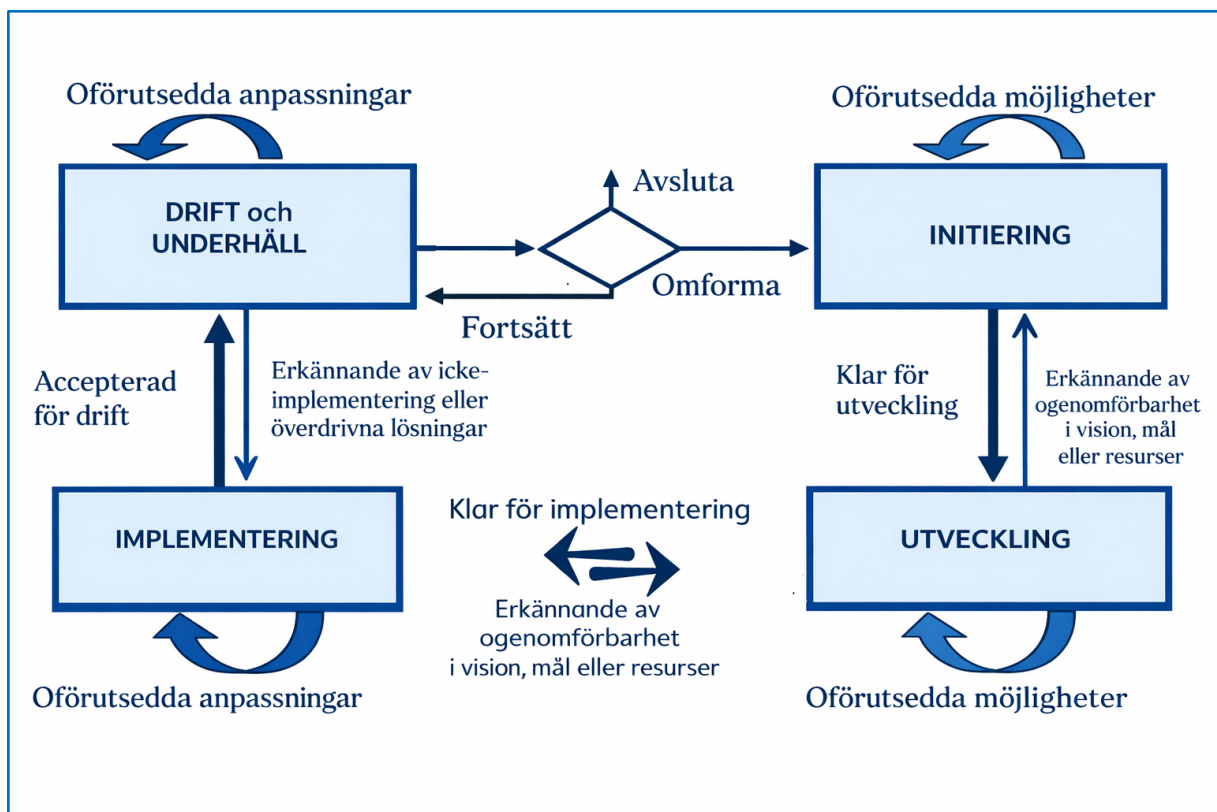
Vidare tydliggör tabell 1 de element som ingår i arbetssystemramverket enligt Alter (2013) och anledningen till att dessa element ingår, samt visar hur de tillsammans utgör den organisatoriska kontext inom vilket arbetet utförs i.

Tabell 1: Fri sammanfattning av tabell enligt Alter (2013, s. 80-81) med element som är relevanta för arbets-systemramverket.

Element	Anledning till inkludering i arbets-systemet
Processer och aktiviteter	De arbetsmoment som utförs för att producera produkter eller tjänster.
Deltagare	Personer som utför arbete inom systemet, både de som interagerar med IT och de som inte gör det.
Information	Den data och kunskap som används, skapas eller delas inom systemet, både digitalt och muntligt.
Teknik	Verktyg och automatiserade system som används i arbetet, inklusive AI-lösningar.
Produkter och tjänster	Resultatet av arbetssystemets aktiviteter, det som produceras för externa kunder eller interna mottagare.
Kunder	Mottagare av systemets produkter/tjänster, både interna (anställda) och externa (kunder).
Miljö	De kulturella, organisatoriska, tekniska och rättsliga förutsättningarna som påverkar systemets funktion.
Infrastruktur	Resurser, som kan vara mänskliga, tekniska, informativa, och som stödjer arbetssystemet.
Strategi	Organisationens övergripande mål, prioriteringar och riktning.

2.3.2 Arbetssystemlivscykelmodellen

Arbetssystemlivscykelmodellen (work system life cycle model) (se figur 2) bidrar med ett dynamiskt perspektiv på hur arbetssystem utvecklas och förändras över tid genom en kombination av planerade och oplanerade aktiviteter. Modellen (se figur 2) visar att förändring inte enbart är ett resultat av formella utvecklingsinsatser, utan även formas av löpande anpassningar i det dagliga arbetet. De planerade förändringarna sker i form av projekt som omfattar tre huvudsakliga faser, det vill säga initiering, utveckling och implementering. (Alter 2013, s. 81).



Figur 2: Egen bearbetning av arbetssystemlivscykelmodellen baserad på Alter (2013, s. 78).

Under utvecklingsfasen skapas eller införskaffas de resurser som krävs för att genomföra förändringar i organisationen, exempelvis genom utveckling eller inköp av programvara, konfiguration av system samt framtagande av nya rutiner, dokumentation och utbildningsmaterial (Alter 2013, s. 81). Dessa åtgärder syftar till att möjliggöra en ny version av arbetssystemet och skapa förutsättningar för att förändringar ska kunna implementeras i praktiken. Samtidigt betonar modellen att arbetssystem även kan förändras genom informella anpassningar, där användningen av teknik successivt formas i vardagliga arbetspraktiker snarare än genom enbart planerade projekt. (Alter 2013, s. 82).

Arbetsystemlivscykelmodellen illustrerar därmed att förändring i organisationer även sker genom framväxande förändringar (emergent change), vilka uppstår när anställda anpassar, improviserar eller utvecklar egna lösningar i arbetet (Alter 2013, s. 82). Dessa förändringar visualiseras i figur 2 genom de inåtriktade pilarna, som representerar pågående och ofta informella anpassningar i arbetssystemet. Anpassningarna kan vara både kortsiktiga lösningar (workarounds) och mer långsiktiga förändringar i arbetsrutiner och mål, särskilt i drift- och underhållsfasen och sker utan att initieras genom formella projekt eller särskild tilldelning av projektresurser (Alter 2013, s. 82).

2.4 AI:s framkant

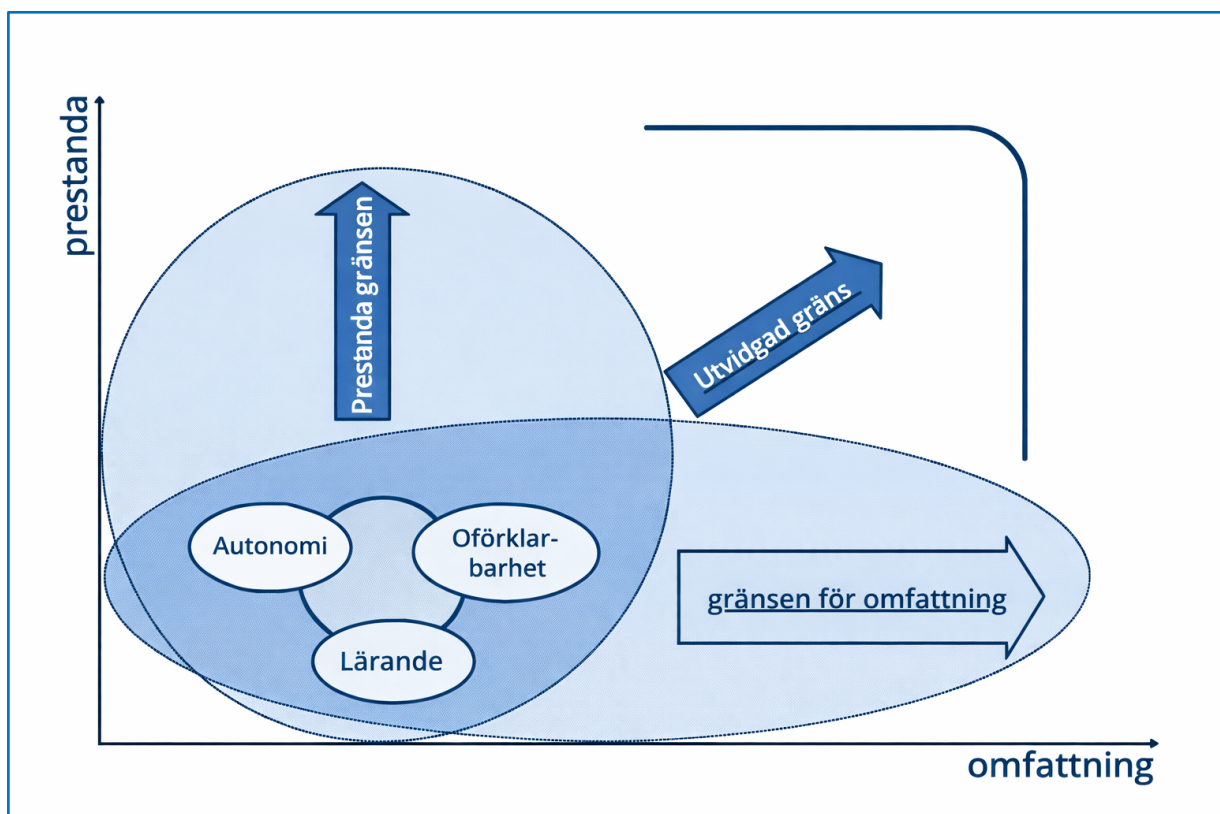
AI:s framkant utvecklades av Berente et al. (2021) som ett sätt att förstå AI som ett dynamiskt gränsområde inom databehandling snarare än som en fast avgränsad teknik. AI beskrivs som en del av den tekniska utvecklingen där beräkningssystem i allt högre grad närmar sig mänsklig intelligens i hanteringen av mer komplexa beslutsproblem. Detta innebär att AI representerar den yttersta nivån av datorers kognitiva kapacitet i relation till mänsklig intelligens, där gränsen ständigt förskjuts i takt med att nya innovationer inom automatisering, maskininläring och datadriven beslutsanalys utvecklas. (Berente et al. 2021, s. 1438).

AI:s framkant-ramverk bygger på tre centrala egenskaper, det vill säga autonomi, lärande och oförklarbarhet (inscrutability), som tillsammans fångar de kvalitativa egenskaper som särskiljer AI från tidigare informationssystem. Autonomi syftar på AI:s förmåga att agera självständigt utan direkt mänsklig inblandning, vilket innebär att beslutsfattandet delvis flyttas från människor till algoritmer. Lärande handlar om att AI kontinuerligt förbättras genom erfarenhet, data och feedback, vilket gör tekniken adaptiv och svår att förutsäga. Oförklarbarhet beskriver den begränsade förståelsen av hur AI-system fattar beslut, vilket skapar osäkerhet och utmanar principer om ansvar och spårbarhet. (Berente et al. 2021, s. 1438-1440).

Författarna menar vidare att dessa tre dimensioner medför nya typer av lednings- och styrningsproblem. Eftersom AI-system har förmågan att utvecklas bortom mänsklig förståelse, måste organisationer etablera strukturer för kontroll, tillsyn och ansvarstagande. Samtidigt uppstår spänningar mellan mänsklig expertis och algoritmisk logik, vilket innebär att roller, beslutsprocesser och maktförhållanden förändras. AI:s framkant utmanar därför traditionella uppfattningar om teknikstyrning genom att framhäva behovet av en socioteknisk förståelse,

där både tekniska och mänskliga faktorer samverkar i skapandet av organisatoriskt värde. Ur detta perspektiv blir AI inte bara ett verktyg, utan en transformativ kraft som omdefinierar hur organisationer fattar beslut och agerar. (Berente et al. 2021, s. 1442-1444).

Figur 3 illustrerar AI:s framkant längs de två övergripande dimensionerna prestanda och omfattning. Prestanda avser hur väl AI-system kan utföra en given uppgift, exempelvis i termer av precision, effektivitet och tillförlitlighet, medan omfattning syftar på hur brett AI tillämpas över olika uppgifter, processer och organisatoriska sammanhang.



Figur 3: Egen bearbetning baserad på AI:s framkant av Berente et al. (2021, s. 1438).

I figur 3 representerar prestandagränsen den nivå där AI:s kapacitet når dess maximala prestationsförmåga inom ett givet användningsområde, medan gränsen för omfattning visar hur långt AI har spridits till nya typer av uppgifter och beslutsprocesser. (Berente et al. 2021, s. 1437). Den utvidgade gränsen illustrerar hur AI:s framkant kontinuerligt förskjuts i takt med att systemen blir mer autonoma, lärande och oförklarbara. När AI-system utvecklas genom kontinuerligt lärande kan det både förbättra sin prestanda och användas inom allt fler områden, vilket successivt förskjuter både prestandagränsen och gränsen för omfattning. Samtidigt

bidrar denna utveckling till ökad oförklarbarhet, vilket innebär att beslutslogiken blir svårare att tolka och kontrollera. (Berente et al. 2021, s. 1437).

I en bankkontext innebär detta att AI-baserade beslutsstöd kan öka effektiviteten, men att det samtidigt kräver nya former av tillit, transparens och ansvarsfördelning mellan människa och system. När AI-system uppvisar hög autonomi och oförklarbarhet innebär det ett ökat gap mellan AI:s kapacitet och människors förståelse, vilket gör transparens till ett centralt villkor för att minska risken för att anställda upplever AI som oförklarbart och opålitligt.

2.5 Transparens som villkor för tillit till AI

Tillitsteori inom informationssystemforskning pekar på transparens som en central förutsättning för hur tillit till tekniska system, som AI-baserade beslutsstöd, utvecklas.

2.5.1 Transparens som förutsättning till tillit

Forskning kring tillit och transparens har haft kontrasterande resultat, vilket Yu och Li (2022, s. 1) belyser eftersom de hävdar att transparens både kan stärka och försvaga tilliten beroende på hur den uppfattas av den berörda aktören. Högre transparens kan öka förståelsen och därigenom stärka tilliten genom ökad upplevd effektivitet, men kan samtidigt försvaga tilliten genom att skapa obehag när den ökade insynen upplevs som belastande och i vissa situationer väger tyngre än effektiviteten (Yu & Li 2022, s. 12). Detta belyser även Jiang, Li och Liu (2025, s. 13) som visar att transparensdimensionen kan hämma kognitiva processer genom ökad kognitiv belastning. Detta indikerar att ökad transparens inte per automatik leder till bättre kognitiva utfall, utan kräver en kontextanpassad utformning.

Å andra sidan hävdar Maier, Menold och McComb (2022, s. 15) att människors förtroende för AI i finansiella sammanhang påverkas av hur transparent systemet är. Maier, Menold och McComb (2022, s. 15) visar i studien att tillit till AI inte enbart påverkas av systemets eller människans faktiska prestation, utan snarare av skillnaden i deras prestation. När AI presterade bättre än deltagaren ökade tilliten, samtidigt som tilliten i stället minskade när AI presterade sämre. Vid likvärdiga prestationer använde deltagarna båda resultaten som referens för att bedöma den framtida tilliten till systemet och författarna konstaterar därmed att tillit formas i samspelet mellan mänskliga och tekniska faktorer (Maier, Menold & McComb 2022, s. 15).

När transparensen är låg tenderar människor att reproducera samma kognitiva snedvridningar som i mänskliga beslutssituationer, vilket kan leda till felaktiga bedömningar och minskat förtroende. Det finns därmed en central utmaning kring att förstå hur AI-system fattar beslut, vilket understryker behovet av transparens och förklarbarhet för att användarna ska uppleva tillit till AI som beslutsstöd (Berente et al. 2021, s. 1441-1442).

2.5.2 Olika dimensioner av transparens

Forskningen om transparens inom AI präglas enligt Haresamudram, Larsson och Heintz (2023, s. 93) av en splittrad förståelse, där fokus generellt begränsas till den tekniska eller algoritmiska dimensionen. I stället hävdar författarna att transparens bör förstås som ett flerdimensionellt fenomen som verkar på tre olika nivåer, det vill säga algoritmisk, interaktionell och social transparens, som tillsammans utgör grunden för att skapa tillit till AI (Haresamudram, Larsson & Heintz 2023, s. 93).

Den algoritmiska transparensen syftar till förståelse för hur AI fungerar, vilket inkluderar förståelse för AI:s behandling av data och beslutsfattning (Haresamudram, Larsson & Heintz 2023, s. 96). Eftersom många AI-modeller är komplexa är deras beslutslogik ofta svår att förstå, och för att hantera dessa utmaningar har området förklarbar AI (explainable AI, XAI) utvecklats, med syfte att göra algoritmiska beslut mer begripliga genom olika typer av förklaringar (Haresamudram, Larsson & Heintz 2023, s. 96).

Samtidigt konstaterar Haresamudram, Larsson och Heintz (2023, s. 97) att tekniska förklaringar ofta är anpassade för experter och därför kan vara svåra att tolka för människor utan teknisk bakgrund. Detta innebär att algoritmisk transparens inte enbart handlar om teknisk insyn i modellens uppbyggnad, utan om att möjliggöra en praktiskt användbar förståelse av hur och varför ett visst beslut har fattats (Haresamudram, Larsson & Heintz 2023, s. 96). Algoritmisk transparens går även att koppla till frågor om ansvar och insyn, där möjligheten att granska koder, träningsdata och modellantaganden ses som centralt för att säkerställa ansvarsstagande och tillförlitlighet (Haresamudram, Larsson & Heintz 2023, s. 96).

Interaktionell transparens handlar om att transparens i AI uppstår genom interaktion mellan människa och system, inte bara genom tekniska förklaringar (Haresamudram, Larsson &

Heintz 2023, s. 97). Transparens uppstår därmed när den berörda aktören, genom interaktion med AI, successivt bygger en förståelse för dess resonemang, begränsningar och användbarhet i relation till den aktuella uppgiften (Haresamudram, Larsson & Heintz 2023, s. 97).

När AI blir mer avancerad skapas nya sätt för människor att förstå och lita på tekniken genom praktisk, materiell samt situationsbunden interaktion, och transparens blir därmed ett resultat av samspel samt delad erfarenhet mellan människa och AI (Haresamudram, Larsson & Heintz 2023, s. 97). Det är därmed inte graden av teknisk förklarbarhet som är avgörande, utan huruvida interaktionen möjliggör en praktisk förståelse för AI:s styrkor, begränsningar och relevans i situationen, vilket i sin tur påverkar användarnas tillit (Haresamudram, Larsson & Heintz 2023, s. 98).

Social transparens handlar i sin tur om hur AI-system förstås samt styrs i organisatoriska och samhällseliga sammanhang (Haresamudram, Larsson & Heintz 2023, s. 98). Detta inkluderar ansvar, etik och reglering samt hur utvecklare, den som interagerar med AI-systemet och myndigheter samverkar kring AI (Haresamudram, Larsson & Heintz 2023, s. 98). Enligt Haresamudram, Larsson och Heintz (2023, s. 98) går det att konstatera att tillit till AI ofta formas genom institutionella strukturer, som lagstiftning, interna riktlinjer och organisatoriska värderingar, snarare än genom individens tekniska förståelse. Mot denna bakgrund kan transparens i relation till AI analyseras som ett flerdimensionellt fenomen som formas av AI:s förmåga, mänsklig förståelse, interaktion samt organisatoriskt ansvar och reglering.

2.6 Förklarbar AI

Det går att konstatera att den bristande insynen i hur AI fattar beslut är ett centralt problem, då många modeller fungerar som svarta lådor (black boxes) där beslutsprocessen är svår att tolka för människor (Rai 2020, s. 137). Förklarbar AI (XAI), utvecklades som tidigare nämnt som ett svar på detta, och syftar till att skapa metoder som gör AI-system begripliga för användarna och därigenom stärka både transparens och tillit (Dwivedi et al. 2023, s. 194:1).

Förklarbar AI är också en central etisk komponent, eftersom ökad insyn i beslutsprocesser gör det möjligt att upptäcka och motverka orättvisor, snedvridningar och andra oetiska konsekvenser gällande AI som beslutsfattare (Ali et al. 2023, s. 8). Oförklarbarhet inom AI kan försvåra upptäckten av exempelvis diskriminerande mönster och leda till att de berörda aktörerna

känner osäkerhet kring att förlita sig på systemen och för att motverka detta behöver AI-modeller vara begripliga och kunna förklara sina beslut, vilket på så sätt kan stärka förtroendet (Rai 2020, s. 137-139). Genom XAI blir svarta lådor därmed mer transparenta, vilket innebär att beslutsprocesser blir synliga, spårbara och möjliga att förstå även för berörda aktörer utan teknisk expertis (Rai 2020, s. 138-139). Dwivedi et al. (2023, s. 194:3) menar att traditionella maskininlärningsprocesser främst fokuserar på korrekta prediktioner och därför saknar två centrala komponenter, det vill säga förståelse under modellens utveckling samt förklaring under användning av modellen.

För att beskriva hur förklarbarhet kan uppnås presenterar Dwivedi et al. (2023, s. 194:5) en taxonomi över olika tekniker. Även här används exemplet med vita lådor, som är transparenta till sin konstruktion, och svarta lådor, som är mer komplexa och kräver metoder för att bli begripliga (Dwivedi et al. 2023, s. 194:5-194:6). De skiljer även mellan modellspecifika tekniker, som är anpassade till en viss algoritm, och modelloberoende tekniker som kan tillämpas på alla modeller (Dwivedi et al. 2023, s. 194:7). Slutligen lyfter de skillnaden mellan global förklarbarhet, som förklarar hur modellen fungerar i stort, och lokal förklarbarhet, som visar varför ett specifikt beslut togs (Dwivedi et al. 2023, s. 194:8).

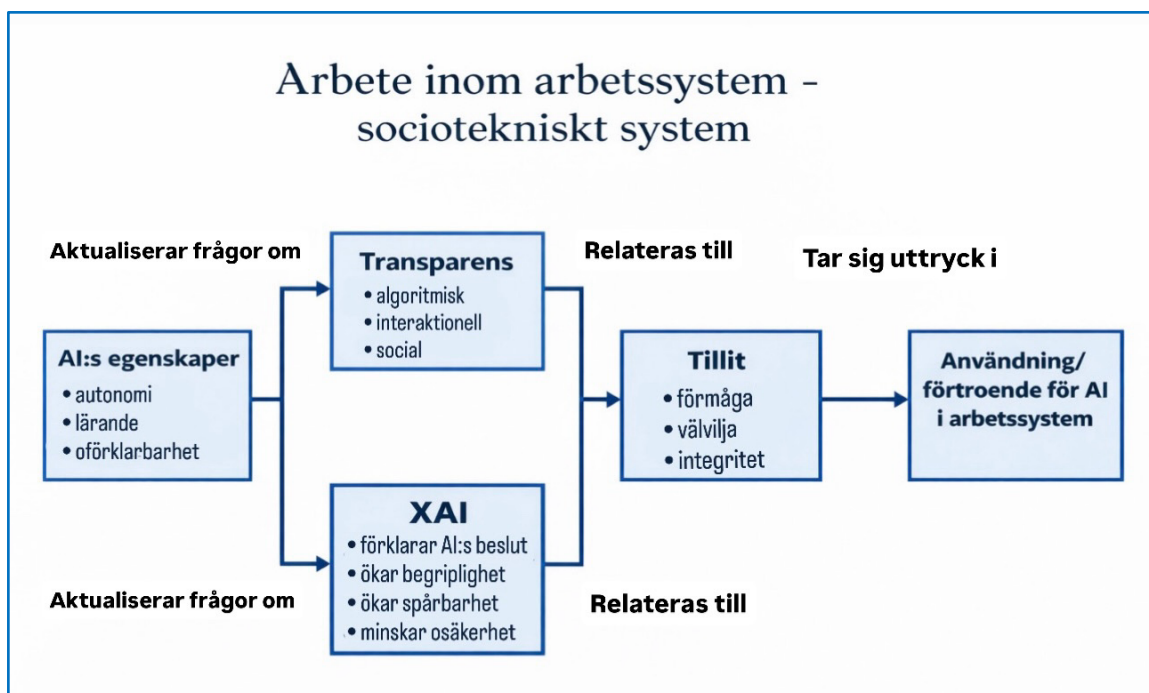
Samtidigt menar Rai (2020, s. 141) att förklarbarhet i AI-system inte bör förstås som ett krav på maximal transparens. Även om detaljerade förklaringar kan vara värdefulla för tekniskt insatta aktörer, som utvecklare och dataexperter, är slutanvändare ofta mindre vana vid hög teknisk transparens. Komplexa förklaringar och omfattande insyn i AI-modeller kan därför i stället medföra ökad kognitiv belastning, informationsöverflöd och frustration hos människor (Rai 2020, s. 141). Detta visar vikten av att anpassa graden av förklarbarhet och transparens efter den professionella aktörens roll och behov, där förenklade och tillgängliga förklaringar ofta är tillräckliga för att stödja förståelse och användning - i stället för överbelastning (Rai 2020, s. 141).

Tillsammans visar dessa distinktioner att XAI kan operationaliseras på olika sätt beroende på bankanställdas behov, modellens komplexitet och kraven på insyn. Genom att synliggöra hur beslut genereras och vilka faktorer som påverkar modellens rekommendationer kan XAI minska den osäkerhet som uppstår när system fungerar som ”svarta lådor”. Upplevelsen av förklarbarhet och begriplighet påverkar i sin tur tillit, eftersom förklarbarhet stärker användarnas uppfattning om AI:s förmåga, integritet och pålitlighet. XAI blir därmed ett teoretiskt

verktyg för att analysera hur transparens formas i praktiken och hur denna transparens påverkar bankanställdas vilja att förlita sig på AI i sitt beslutsfattande.

2.7 Deduktivt konceptuellt ramverk

Enligt Hennink, Hutter och Bailey (2020, s. 36) används ett konceptuellt ramverk för att organisera studiens centrala begrepp och visa hur de relaterar till varandra. Vid en deduktiv ansats fungerar ramverket som ett analytiskt verktyg som strukturerar hur det empiriska materialet tolkas (Hennink, Hutter & Bailey 2022, s. 36). Det deduktiva konceptuella ramverket (se figur 4) utgår därför från studiens syfte, det vill säga att förstå hur transparens i AI-baserade beslutsstöd påverkar bankanställdas tillit, och syftar till att förklara hur teorier samt begrepp relaterar till varandra.



Figur 4: Deduktivt konceptuellt ramverk för studien (egen figur).

Som figur 4 visar, används arbetssystemteorin och det sociotekniska perspektivet som grund för det deduktiva konceptuella ramverket, som beskriver hur beslutsfattande sker i ett sociotekniskt sammanhang där människor, teknik, information och organisatoriska processer samspelar (Alter 2013). AI utgör därmed en teknisk komponent vars utformning och begränsningar aktualiserar frågor om hur de bankanställda uppfattar och värderar dess förmågor som beslutsstöd, vilket i sin tur påverkar hur AI används i beslutsprocesser.

Följaktligen appliceras teorier enligt Berente et al. (2021, s. 1438–1442) gällande att AI:s egenskaper, i form av autonomi, lärande och oförklarbarhet, skapar ett ökat behov av transparens, eftersom systemen blir allt svårare att förstå, förutse och kontrollera. Samtidigt går det att konstatera att forskning visar att transparens både kan stärka och hämma tillit eftersom den ökar förståelsen, men även skapa kognitiv belastning och oro (Yu & Li 2022, s. 12; Maier, Menold & McComb 2022, s. 15). Transparens har således en positiv effekt först när den gör AI begriplig samt hanterbar och dessa perspektiv bidrar med kunskap kring olika effekter av transparens.

För att analysera hur transparens uppstår appliceras Haresamudram, Larsson och Heintz (2023, s. 93) teoretiska ramverk, där transparens förstås som ett flerdimensionellt fenomen bestående av algoritmisk, interaktionell och social transparens. Dessa dimensioner belyser insynen i hur AI-system behandlar data, hur förståelse utvecklas genom interaktion samt hur ansvar och styrning formar användningen av AI i organisatoriska sammanhang (Haresamudram, Larsson & Heintz 2023, s. 96–98). Förklarbar AI integreras även som en central mekanism för att operationalisera transparens genom att synliggöra AI-modellers beslutslogik (Dwivedi et al. 2023, s. 194:5-194:8). Genom att synliggöra AI-modellers beslutslogik bidrar XAI till att göra systemen mer begripliga och spårbara (Dwivedi et al. 2023, s. 194:5-194:8). Samtidigt påpekar författarna att den bakomliggande komplexiteten i AI-modeller kan leda till orättvisa eller svårförklarliga beslut, vilket tydliggör hur transparens blir avgörande för hur tillit till AI-system utvecklas (Haresamudram, Larsson & Heintz 2023, s. 96).

Vidare appliceras teori enligt Söllner et al. (2016), för att få en ökad förståelse kring transparens och hur tillit formas, där utgångspunkten är att tillit bygger på tre dimensioner, det vill säga förmåga, välvilja och integritet. McKnight et al. (2011, s. 12:15) kompletterar denna förståelse med teorin om erfarenhetsbaserad tillit, det vill säga tillit som utvecklas genom praktisk interaktion med tekniken, och som formas av hur funktionell, hjälpsam och tillförlitlig tekniken upplevs över tid.

Det konceptuella ramverket visar därför att tillit till AI i bankernas beslutsstöd formas i samspillet mellan AI:s egenskaper (autonomi, lärande & oförklarbarhet), transparensens olika dimensioner (algoritmisk, interaktionell & social) samt användarnas erfarenhetsbaserade bedömning av dess förmåga, välvilja och integritet, vilket genom arbetssystemteorin kan operationaliseras. Därmed kan det konstateras att tillit till AI-baserade beslutsstöd utgör den

beroende variabeln i ramverket, medan de oberoende variablerna syftar till AI:s egenskaper, transparensens dimensioner (algoritmisk, interaktionell & social) samt graden av förklarbarhet (XAI). Sammantaget påverkar dessa faktorer bankanställdas bedömning av AI:s förmåga, välvilja och integritet, vilka enligt tillitsteorin utgör centrala förutsättningar för att tillit ska kunna utvecklas och som i sin tur tar sig uttryck i hur AI-baserade beslutsstöd används i praktiken.

3. Metod

Metodkapitlet redogör för de metoder som ligger till grund för studiens genomförande och beskriver den kvalitativa forskningsstrategin, urvalet av deltagare samt hur datainsamlingen genomförts genom semistrukturerade djupintervjuer. Vidare förklaras den analytiska ansatsen och hur det empiriska materialet har bearbetats för att möjliggöra en tolkning av bankanställdas upplevelser av transparens och tillit kopplat till AI-baserat beslutsfattande. Syftet är att ge en tydlig och transparent redogörelse för forskningsprocessen, så att det blir möjligt att bedöma studiens kvalitet och reproducerbarhet.

3.1 Metodval

Denna studie genomförs med hjälp av en kvalitativ forskningsstrategi, eftersom syftet med studien är att undersöka hur transparens i AI-baserade beslutsstöd påverkar bankanställdas upplevelse av tillit i sitt beslutsfattande. Kvalitativ forskning används för att få en djup förståelse av människors beteenden, upplevelser och den kontext de verkar i (Hennink, Hutter & Bailey 2020, s. 11). Den lämpar sig särskilt för att utforska komplexa eller känsliga fenomen samt ge röst åt deltagarnas egna perspektiv (Hennink, Hutter & Bailey 2020, s. 11). Syftet är att tolka den mening människor ger sina erfarenheter samt att skapa djup och nyans i forskningen (Hennink, Hutter & Bailey 2020, s. 11). Mot bakgrund av dessa insikter bedöms en kvalitativ forskningsstrategi vara lämplig för att fånga bankanställdas subjektiva upplevelser av tillit till AI som beslutsstöd, snarare än att mäta tillit till AI i kvantitativa termer.

Den kvalitativa metoden möjliggör dessutom en djupare förståelse av de sociala och organisatoriska faktorer som påverkar hur tekniken uppfattas och används i praktiken. Genom att utgå från deltagarnas egna berättelser och erfarenheter kan forskningen belysa de nyanser och mot-sägelser som präglar deras relation till AI, exempelvis hur tillit, transparens och ansvarsfördelning uppfattas i beslutsprocesser. Enligt Hennink, Hutter och Bailey (2020, s. 10) ger en kvalitativ ansats forskaren möjlighet att studera människor i deras naturliga miljö och tolka hur deras handlingar formas av kontextuella faktorer som organisatorisk kultur, teknikens roll och sociala normer. Detta är särskilt relevant i banksektorn, där införandet av AI medför både möjligheter och etiska utmaningar, och där förståelse för anställdas perspektiv framstår som en viktig förutsättning för hur tillit mellan människa och teknik utvecklas.

3.2 Urval

Enligt Hennink, Hutter och Bailey (2020, s. 92) används ett strategiskt urval inom kvalitativ undersökning. Studiepopulationen bestäms utifrån forskningsfrågan eftersom den anger vilka individer som kan bidra med relevant och detaljerad information om ämnet (Hennink, Hutter & Bailey 2020, s. 92). I denna studie definierades studiepopulationen utifrån behovet av att förstå hur transparens i AI-baserade beslutsstöd påverkar bankanställdas upplevelse av tillit i beslutsfattandet. För att möjliggöra detta formulerades därför följande två kriterier:

- Deltagaren ska vara anställd inom en bank.
- Deltagaren ska använda, eller ha använt, AI som ett stöd i arbetsrelaterade beslutsprocesser.

Dessa kriterier motiveras av att tillit och transparens endast kan studeras hos personer som i praktiken interagerar med tekniken och därmed kan beskriva sina erfarenheter och upplevelser av tekniken i den kontext där besluten faktiskt fattas. Rekryteringen genomfördes genom ett strategiskt urval, kombinerat med ett snöbollsurval i vissa fall, där en initial kontaktperson inom banksektorn förmedlade kontakt till ytterligare en deltagare som uppfyllde kriterierna. Enligt Hennink, Hutter och Bailey (2020, s. 104) är snöbollsurval särskilt användbart när forskaren söker personer med specifik erfarenhet, eftersom befintliga deltagare kan identifiera relevanta individer som annars hade varit svåra att nå.

3.3 Datainsamlingsmetod

För att samla in data till denna studie används djupintervjuer då denna typ av intervju är lämplig för att fånga individens erfarenheter, upplevelser och åsikter kring ett specifikt ämne (Hennink, Hutter & Bailey 2020, s. 117). Djupintervjuer används därför för att fånga hur bankanställda förstår och tolkar transparens i AI-baserade beslutsstöd samt hur denna förståelse påverkar deras upplevelse av tillit. Den semistrukturerade ansatsen som används för studien skapar en balans mellan förutbestämda teman och flexibilitet att följa upp deltagarnas egna beskrivningar, vilket gör det möjligt att få fram nyanserade och detaljerade berättelser (Hennink, Hutter & Bailey 2020, s. 118-120).

Vidare betonar Hennink, Hutter och Bailey (2020, s. 116) att djupintervjuer bygger på en dialogisk interaktion där forskare skapar en förtroendefull relation präglad av lyhördhet och

empati. Dessa aspekter är centrala för att deltagarna ska känna sig trygga att dela sina erfarenheter och reflektioner, vilket i sin tur bidrar till ett rikare och mer tillförlitligt datamaterial (Hennink, Hutter & Bailey 2020, s. 117). Metoden blir därför särskilt värdefull i studier som syftar till att förstå subjektiva upplevelser och de meningsskapande processer som ligger till grund för deltagarnas handlande. Genom att låta deltagarna beskriva sina erfarenheter i relation till deras yrkesroll och arbetsmiljö belyser intervjuerna de praktiska situationer där mänskliga samt tekniska komponenter möts. Metoden gör det därmed möjligt att synliggöra hur uppfattningar om transparens formas i vardagliga beslutprocesser, vilket ligger i linje med kvalitativa forskningens mål, det vill säga att belysa individers perspektiv i deras naturliga sammanhang (Hennink, Hutter & Bailey 2020, s. 11).

Intervjuerna genomfördes huvudsakligen med en respondent i taget, med undantag för ett intervjutillfälle där två respondenter från samma företag deltog samtidigt. Wilson, Onwuegbuzie och Manning (2016, s. 1549) beskriver parade djupintervjuer som en kvalitativ metod, där två personer intervjuas samtidigt för att belysa hur de tillsammans förstår och beskriver ett gemensamt fenomen. Metoden bygger på att deltagarna interagerar med varandra, vilket kan ge mer nyanserade data än enskilda intervjuer (Wilson, Onwuegbuzie & Manning 2016, s. 1549). Genom att observera hur paret kompletterar varandras berättelser eller uttrycker olika perspektiv kan forskaren få djupare insikt i deras delade erfarenheter. Denna metod fungerar särskilt väl när deltagarna har en relation eller gemensam arbetskontext, eftersom detta ofta leder till en mer naturlig och reflekterande diskussion. (Wilson, Onwuegbuzie & Manning 2016, s. 1554).

Samtidigt går det att konstatera att metoden har vissa begränsningar, som exempelvis risken att en respondent dominerar samtalet eller att deltagarna anpassar deras utsagor för att framstå som eniga. Metoden är heller inte lämplig om studien undersöker ämnen som kan vara känsliga. (Wilson, Onwuegbuzie & Manning 2016, s. 1555). Denna metodologiska begränsning bedöms dock som hanterbar i studien, eftersom den parade intervjun utgjorde ett enskilt inslag i datainsamlingen och behandlade arbetsrelaterade erfarenheter i en gemensam professionell kontext. För att ytterligare minska risken för att utsagor påverkades av den gemensamma intervjusituationen genomfördes även en uppföljande kontakt med respondent 3, där vissa resonemang kunde förtydligas individuellt.

Intervjuguiden (se bilaga 1) är i denna studie utformad enligt rekommendationer av Hennink, Hutter samt Bailey (2020) och fungerar som ett flexibelt stöd för intervjun snarare än ett strikt frågeformulär. Guiden består av öppna frågor som bjuder in deltagare att beskriva sina erfarenheter med egna ord, genom en struktur med en inledande introduktion med bakgrundsfrågor, huvudfrågor och avslutande frågor (Hennink, Hutter & Bailey 2020, s. 143). Detta möjliggör ett naturligt samtalsflöde och skapar den trygghet som krävs för att få fram rika och detaljerade berättelser. Följaktligen är nyckelfrågorna i intervjuguiden deduktiva och kopplade till studiens forskningsfråga och undersöker centrala begrepp som transparens och tillit till AI-baserade beslutsstöd. I linje med Hennink, Hutter och Bailey (2020, s. 124) inkluderar guiden även följdfrågor (topical probes) som hjälpmedel för att fördjupa svaren vid behov utan att styra respondenterna för mycket. Intervjuerna genomförs som resultat av detta med flexibilitet, samtidigt som de har en tydlig teoretisk förankring, för att säkerställa relevans för studiens syfte.

Majoriteten av intervjuerna genomfördes genom videointervjuer via Microsoft Teams. Denna form av digital datainsamling är enligt Bryman, Bell och Harley (2024, s. 407) ett etablerat alternativ till fysiska intervjuer som erbjuder flera praktiska fördelar. Bland annat försvinner geografiska begränsningar, vilket gör det enklare att inkludera deltagare oavsett plats, samtidigt som tidsåtgången minskar då inga resor krävs (Bryman, Bell & Harley 2024, s. 407). Bryman, Bell och Harley (2024, s. 407) menar dessutom att videobaserade intervjuer inte nödvändigtvis försämrar möjligheten att skapa en bra relation till deltagaren, då både verbal och icke verbal kommunikation fortfarande kan fångas. Mot denna bakgrund bedömdes videointervjuer vara ett ändamålsenligt sätt för datainsamling vid de tillfällen då respondenterna föredrog detta framför fysiska intervjuer.

En svårighet med djupintervjuer är generellt att förutspå hur deltagarna kommer att reagera på frågorna och det är därför lämpligt att pilottesta frågorna innan intervjuerna äger rum (Hennink, Hutter & Bailey 2020, s. 143). Under pilottesterna är det relevant att undersöka huruvida respondenterna förstår frågorna, om frågorna behöver omformuleras och om de är i en logisk ordning (Hennink, Hutter & Bailey 2020, s. 125). Som resultat av dessa insikter genomfördes en pilottestning av intervjuguiden på en person för att säkerställa att frågorna var begripliga, relevanta och gav det djup av svar som eftersträvades. Utifrån testet justerades formuleringar och följdfrågor för att skapa ett tydligare och mer användbart intervjumaterial.

3.4 Analysmetod

I denna studie används tematisk analys som metod för att systematiskt identifiera, analysera och tolka återkommande mönster i det empiriska materialet. Enligt Hennink, Hutter och Bailey (2020, s. 6) är kodning, jämförelse, kategorisering och konceptualisering centrala moment. Analysarbetet inleddes därför med en kombination av deduktiv och induktiv kodning, där den deduktiva kodningen användes för att förankra analysen i studiens teoretiska ramverk, medan den induktiva kodningen innebar att det var möjligt att identifiera nya ämnen och låta datan från respondenterna tala (Hennink, Hutter & Bailey 2020, s. 221-222).

För att säkerställa att både de deduktiva och induktiva koderna var robusta och analytiskt användbara användes kriterierna "7R" enligt Hennink, Hutter och Bailey (2020, s. 224). Dessa kriterier fungerar som en struktur för att bedöma kodernas kvalitet och huruvida de faktiskt stödjer studiens forskningsfråga. Enligt 7R-kriterierna bör koder vara relevanta för studiens ämne, representera det analyserade fenomenet på ett adekvat sätt, vara identifierbara i datamaterialet, återkomma hos flera deltagare, uttryckas av respondenterna själva, godkännas av forskningsgruppen samt kunna kopplas till tydliga textsegment (Hennink, Hutter & Bailey 2020, s. 224).

Dessa kriterier tillämpades både på de deduktiva koderna och på de induktiva koderna som växte fram under analysen. Detta innebar att även de teoretiskt förankrade koderna kontrollerades mot datamaterialet för att säkerställa att de faktiskt förekom och var meningsfulla i respondenternas utsagor. På så sätt var det möjligt att konstatera att kodschemat var empiriskt grundat, konsekvent och användbart för att analysera hur transparens i AI-system påverkar bankanställdas tillit.

I denna studie var det två personer som kodade datan och därför diskuterades olika strategier för att säkerställa en konsekvent kodning, enligt metoder av Hennink, Hutter och Bailey (2020, s. 227). Inledningsvis genomfördes en gemensam genomgång av kodschemat där centrala koder definierades och avgränsades för att skapa en gemensam förståelse. Därefter kodades ett urval av intervjumaterialet parallellt av båda forskarna, som sedan jämfördes. Skillnader i tolkning användes som underlag för att förtydliga kodernas definitioner och justera gränsdragningarna mellan närliggande koder, vilket Hennink, Hutter och Bailey (2020, s. 226-227) rekommenderar.

Därefter jämfördes deltagarnas utsagor för att identifiera likheter och skillnader, där Hennink, Hutter och Bailey (2020, s. 245) menar att beskrivning av olika ämnen bidrar med detaljer, som jämförelsen i sin tur kan fördjupa. Denna jämförande analys låg till grund för en initial kategorisering, där närliggande koder grupperades utifrån gemensamma betydelser och analytiska fokusområden. Kategorisering innebär att koder med liknande attribut placeras i kategorier som speglar studiens forskningsfråga. Genom att samla koder med liknande betydelse skapas mer övergripande begrepp som gör datan mer strukturerad och hanterbar. Kategorier kan utvecklas både induktivt utifrån materialet och deduktivt utifrån teori, men måste alltid valideras mot datan för att vara hållbara (Hennink, Hutter & Bailey 2020, s. 247).

Följaktligen konceptualiserades datan, vilket innebär att kategorierna kopplades samman och lyftes till en mer abstrakt nivå. Detta gjorde det möjligt att identifiera mönster, samband och övergripande strukturer som förklarar hur och varför något sker i materialet. Genom konceptualisering utvecklas en teoretisk förståelseram som sträcker analysen bortom beskrivning till förklaring. (Hennink, Hutter & Bailey 2020, s. 247-248). På så sätt skapades en struktur som möjliggjorde en mer systematisk och sammanhållen analys av materialet i relation till studiens forskningsfrågor och denna process låg till grund för de teman som utvecklas i analysen, fördjupas i diskussionen och sammanfattas i studiens slutsats.

3.5 Forskningsetiska överväganden

Hennink, Hutter och Bailey (2020, s. 71) beskriver att viktiga etiska aspekter inom kvalitativa studier är informerat samtycke (informed consent), självbestämmande (self-determination), skademinimering (minimization of harm), anonymitet (anonymity) och konfidentialitet (confidentiality). Eftersom kvalitativa studier i många fall undersöker känsliga ämnen och människors individuella erfarenheter är det av stor vikt att säkerställa att informationen behandlas med omsorg och att deltagarnas integritet skyddas (Hennink, Hutter & Bailey 2020, s. 71-72).

Tydliga etiska riktlinjer samt att säkerställa att informanterna inte känner sig exponerade på ett negativt sätt är därmed viktigt under studiens gång (Rienecker & Stray Jørgensen 2025, s. 239). Alla deltagare erhöll därför ett skriftligt missivbrev (se bilaga 5) inför intervjun som redogjorde för studiens syfte, deltagandets innebörd, att deltagandet var frivilligt samt att samtycke kunde återkallas, det vill säga informerat samtycke och självbestämmande. Följaktligen

informerades deltagarna om att deras identitet skyddas genom anonymisering samt konfidentialitet och samtycke inhämtades även innan intervjuerna. Detta upprepades även muntligt innan intervjuerna startade och finns inkluderat i intervjuguiden (se bilaga 1).

3.6 Metodkritik

En kritik mot studien, och kvalitativ forskning generellt, är den begränsade generaliserbarheten, eftersom resultaten inte bygger på sannolikhetsurval och därmed inte kan anses vara representativa för en större population (Bryman, Bell & Harley 2024, s. 348). Samtidigt menar Bryman, Bell och Harley (2024, s. 348) att detta inte heller är syftet med kvalitativa studier, vilket även gäller denna studie, där målet är att pröva och utveckla teorier baserat på respondenternas erfarenheter.

En annan begränsning i studien rör genomförandet av djupintervjuerna eftersom majoriteten av dessa skedde på distans. Enligt Hennink, Hutter och Bailey (2020, s. 125) genomförs djupintervjuer idealiskt i en för deltagaren välbekant miljö, exempelvis på den egna arbetsplatsen, eftersom detta kan stärka trygghet och ge forskaren en mer kontextuell förståelse. Författarna påpekar samtidigt att digitala intervjuer har blivit allt vanligare när deltagare befinner sig på annan plats eller har begränsad tillgänglighet, även om de fortsatt betonar att djupintervjuer helst bör ske ansikte mot ansikte (Hennink, Hutter & Bailey 2020, s. 133).

I denna studie genomfördes 8 av 10 intervjuer digitalt, i enlighet med deltagarnas önskemål och tidsmässiga förutsättningar. Detta kan ha begränsat möjligheten till att observera deltagarnas kroppsspråk, fysiska uttryck och arbetsmiljö, vilket potentiellt kan ha påverkat den kontextuella förståelsen. För att minska denna svaghet användes videolänk med kameran påslagen, vilket möjliggjorde ögonkontakt och viss observation av deltagarens kroppsspråk trots avståndet. Samtidigt som videointervjuer erbjuder praktiska fördelar finns det begränsningar att beakta. Bryman, Bell och Harley (2024, s. 407) framhåller att tekniska problem som instabil internetuppkoppling eller bristande ljud- och bildkvalitet kan störa intervjusituationen och påverka samtalet. Innan varje intervju gjordes ett test för att säkerställa att ljud och bild fungerade och vid tillfälliga störningar pausades intervjun tills stabil uppkoppling återupprättats.

Följaktligen innebär digitala intervjuer att det finns en risk för exkludering av personer som saknar tillgång till nödvändig teknik eller inte känner sig bekväma med att använda digitala

plattformar, vilket kan påverka urvalets bredd (Bryman, Bell & Harley 2024, s. 407). Eftersom ambitionen initialt var att genomföra intervjuerna på plats utgjorde exkludering ingen risk för studien, då det i stället var deltagarna som själva uttryckte en önskan om att medverka via distansintervjuer genom Microsoft Teams. Flera av deltagarna föredrog digitala intervjuer eftersom det sparar tid och gör det möjligt att delta utan att behöva avsätta tid för resor. Dessutom arbetar flera av deltagarna i en hybrid arbetsform, där arbete på kontoret kombineras med distansarbete, vilket innebär en etablerad vana vid digitala arbetsmiljöer.

Respondent 1 beskrev exempelvis hur arbetet präglas av växling mellan olika kontor och distansarbete, samt hur logistiska faktorer, som svårigheter att boka mötesrum, påverkar valet av mötesformer, och uttryckte detta på följande sätt:

“Jag jobbar ganska mycket där jag växlar mellan kontor [...] och sen så jobbar jag hemma och nu är jag faktiskt hemma [...] det är lite fram och tillbaka [...] det är lite svårt det där med att få mötesrum [...] Så det är otroligt praktiskt det här med Teams.”

Eftersom studien riktar sig mot anställda inom banksektorn omfattas datainsamlingen av säkerhets- och integritetskrav. I flera fall krävdes därför internt godkännande inom organisationen för att intervjuerna skulle få ske, vilket i sig innebär en tidskrävande intern process. Dessa regelverk påverkade även vissa potentiella respondenters möjlighet och vilja att delta i studien, vilket märktes eftersom några potentiella respondenter avböjde medverkan som resultat av detta, vilket kan ha begränsat urvalet.

En annan möjlig begränsning är det ett faktum att intervjuerna var avsedda att genomföras som längre djupintervjuer, men att vissa intervjuer blev kortare än planerat, vilket i några av fallen berodde på respondenternas tillgänglighet. Detta bedömdes i ett fall ha begränsat möjligheten att fördjupa resonemangen ytterligare eller utforska alternativa perspektiv, varför en uppföljande intervju genomfördes med respondent 3.

Avslutningsvis tog studien initialt sin utgångspunkt i AI-baserade beslutsstöd i bred bemärkelse för att inte i förväg begränsa förståelsen av hur AI används i olika roller inom bankverksamheten. I praktiken kom dock generativ AI att dominera respondenternas beskrivningar av den faktiska användningen. Det empiriska materialet och den efterföljande analysen fokuserade därför främst på de AI-former som är mest synliga och interaktiva i respondenternas vardag, samtidigt som andra former av AI, främst prediktiv AI, i vissa fall också diskuterades.

4. Datainsamling

Det empiriska materialet i studien består av intervjuer med bankanställda som i sitt dagliga arbete använder, eller har erfarenhet av, AI-baserade beslutsstöd. Det insamlade materialet redovisas med utgångspunkt i studiens teoretiska referensram, utan att resultaten tolkas eller värderas i detta avsnitt, och illustrerar hur de anställda uppfattar användningen av AI i beslutsprocesser samt hur frågor om transparens samt tillit kommer till uttryck i deras arbete.

4.1 Respondenter

Datainsamlingen i denna studie omfattar 10 intervjuer med deltagare i olika roller från totalt fem banker. Tabell 2 presenterar en sammanställning av respondenterna, inklusive deras arbetsroller, datum för intervjun, anonymiserad organisationstillhörighet, intervjulängd samt om intervjun genomfördes på plats eller digitalt.

Tabell 2: Sammanställning av studiens respondenter och information om intervjuerna

Respondent	Datum	Företag	Intervjulängd (minuter)	Arbetsroll	Metod
Respondent 1	2025-11-27	Företag 1	63	Verksamhetsanalytiker	Digitalt möte
Respondent 2	2025-12-09	Företag 2	58	Kontorschef	Fysiskt möte
Respondent 3 & 4	2025-12-10	Företag 3	45	Kredithandläggare	Digitalt möte
Respondent 5	2025-12-11	Företag 3	61	Senior privatrådgivare	Digitalt möte
Respondent 6	2025-12-15	Företag 4	40	AML-specialist	Digitalt möte
Respondent 7	2025-12-17	Företag 2	67	Privatrådgivare	Digitalt möte
Respondent 8	2025-12-18	Företag 3	67	Företagsrådgivare	Fysiskt möte
Respondent 9	2025-12-19	Företag 2	45	Marknadsoperationsanalytiker	Digitalt möte
Respondent 10	2025-12-21	Företag 5	65	Produktchef för krediter	Digitalt möte
Respondent 3	2026-01-04	Företag 3	32	Kredithandläggare	Digitalt Möte

Som tabellen visar genomfördes majoriteten av intervjuerna individuellt. I ett fall deltog dock två respondenter från samma organisation som tidigare nämnts i samma intervju, vilket skedde på deras egen begäran. För att säkerställa tillräckligt djup i datamaterialet genomfördes även en kompletterande uppföljningsintervju med respondent 3, då den första intervjun var relativt kort. Detta möjliggjorde en fördjupning av centrala teman i materialet.

4.2 Översikt över datainsamlingens innehåll

I det empiriska materialet framkommer att respondenterna använder AI som stöd i beslutsfattande processer snarare än som ett självständigt beslutsverktyg, även om beskrivningarna varierade beroende på arbetsroll och typ av ärenden. Redovisningen av respondenternas svar följer studiens teoretiska ramverk, då intervjufrågorna utformades med utgångspunkt i dessa teorier. Empirin presenteras därför strukturerad efter respektive teoretiskt perspektiv nedan. Tabell 3 ger en översikt över intervjuens tematiska fokus i relation till respektive teori, medan Bilaga 1 redovisar den fullständiga intervjuguiden och de specifika frågor som ställts i relation till teorin.

Tabell 3: Intervjuens tematiska fokus i relation till studiens teoretiska ramverk.

Avsnitt i kapitlet	Fokus i intervjuerna
4.2.1 Socioteknisk systemteori	Hur AI integreras i det dagliga arbetet och hur samspelet mellan människa, teknik och organisation tar sig uttryck.
4.2.2 Tillitsteori	Hur och i vilka situationer AI upplevs som tillförlitlig med fokus på dimensionerna förmåga, välvilja och integritet, samt hur erfarenhet och ansvar påverkar tillit.
4.2.3 Arbetsystemteorin	Hur arbetsprocesser, roller och ansvar är organiserade kring användningen av AI som beslutsstöd.
4.2.4 AI:s framkant	Hur respondenterna uppfattar AI:s utveckling över tid och vilka implikationer detta har för användning och tillit.
4.2.5 Tillit och transparens	Hur olika former av transparens – algoritmisk, interaktionell och social – kommer till uttryck i användningen av AI.
4.2.6 Förklarbar AI	Hur respondenterna hanterar och förhåller sig till förklarbarhet i AI-systemens rekommendationer.

4.2.1 Socioteknisk systemteori

Respondenternas utsagor baseras på frågor som behandlar deras yrkesroll, beslutsfattande, användning av AI samt hur AI-baserade beslutsstöd påverkat arbetsprocesser, roller och ansvar i organisationen (bilaga 1 visar de frågor som ställdes). Respondenterna menar att AI används i det dagliga arbetet inom bankerna, främst i uppgifter som rör informationssökning, sammanställning av material, analysförberedelser, beräkningar i exempelvis Excel samt formulering och omarbetning av texter.

AI används även för att effektivisera administrativa moment, som schemaläggning eller strukturering av innehåll. Respondent 1 beskriver vidare användningen av AI som ett samspel mellan människa och teknik och menar att beslutsstödet inte alltid ger fullständiga svar utan kräver kompletterande frågor:

“Den säger inte fel, men man behöver addera till kompletterande frågor och då får man det svaret också. Men det får man inte initialt. Så jag tror att det är en kombination av människa och maskin fortfarande lite.”

På liknande sätt beskriver respondent 2 hur AI används som ett informations- och analysstöd i beslutsprocesser:

“Det är på flera sätt kan man väl säga. Mycket kopplat till analys där man behöver inhämta information, att man vill få information om en viss bransch, liksom ett bolag eller en viss arbetsrättslig fråga. En arbetsmiljörättslig fråga kanske där jag behöver stöd som jag kanske inte inhämtar från lokalavdelningen. Då kan jag enkelt ställa frågan i Co-Pilot och få svar, källhänvisningar och liknande. Så mycket som ett beslutsstöd, eller informationsstöd, skulle jag säga i dagsläget.”

I samtliga intervjuer framkommer att AI främst används i mer repetitiva och standardiserade arbetsuppgifter, medan mer komplexa moment fortsatt hanteras manuellt av de anställda. Alla respondenter menar med andra ord att AI fungerar som ett komplement till befintliga arbetsprocesser snarare än som ett verktyg som ersätter mänskligt arbete helt.

Respondent 2 lyfter exempelvis hur AI bidrar till effektivisering utan att förändra roller i grunden:

“Men vi använder ju AI framgångsrikt i organisationen för att effektivisera oss. För oss här i kontorsrörelsen än så länge så skulle jag inte säga att rollerna som sådana har dramatiskt

förändrats, däremot så har vi ett bra hjälpmedel i att vi sparar tid i när vi använder det. Vi skapar tid i liksom att förut satt vi och googlande grejer för att få in och veta på information.”.

Sammantaget visar detta att AI integreras i arbetet som ett stödjande element i ett samspel mellan teknik och mänskligt arbete, där effektivisering och tidsbesparing lyfts fram som centrala aspekter och de vanligast förekommande anledningarna till användningen av AI.

4.2.2 Tillitsteori

Genom frågor (se bilaga 1) om AI:s förmåga, välvilja och pålitlighet undersöktes hur respondenterna utvecklar tillit till AI och hur denna tillit påverkar deras användning av AI-baserade beslutsstöd. Respondenterna uttrycker tillit till AI på olika sätt och ofta i relation till den egna erfarenheten samt kompetensen. Majoriteten av respondenterna menar att AI upplevs fungera väl i enklare och mer standardiserade arbetsuppgifter, medan tilliten minskar i mer komplexa eller ansvarstygda beslutssituationer.

Detta beskriver även respondent 5, som menar att AI fungerar väl i standardiserade kreditärenden som tydligt faller inom gällande regelverk, men upplevs som mindre användbart i mer komplexa gränsfall där mänsklig bedömning blir nödvändig.

Följaktligen hävdar respondent 2 följande om AI:s förmåga och den egna användningen:

“[...] jag tycker att det är konsekvent bra på att leverera det man är ute efter. Återigen, jag har använt det på en relativ grundnivå tror jag, utifrån, jämfört med många andra. Men jag tycker ju att det är pålitlig data jag har fått.”.

Samtidigt menar respondent 5 att AI:s rekommendationer vägs mot den egna expertisen och att den mänskliga bedömningen fortsatt är central:

“Det är väl klart att man alltid tänker att jag är bättre, min mänskliga hjärna är bättre än den här roboten. Just för att man har den känslomässiga biten till det. Sen kan det ju såklart vara både positivt och negativt. För man behöver ju vara väldigt neutral i en kreditbedömning helt enkelt.”.

Om AI uppger något som skiljer sig från den egna bedömningen, menar respondent 9 att detta sker:

“Så skulle den säga en sak och jag skulle uppfatta en sak. [...] så diskuterar AI fram en lösning och jag diskuterar fram en lösning. Det kommer alltid finnas någonstans där vi har likheterna. Så i stället för att fokusera på det som är olikheter, så försöker man få fram det som är likheterna då, och få fram det bästa beslutet därifrån.”

Respondentens beskrivning visar att beslutsfattandet sker genom en sammanvägning av AI:s förslag och den egna professionella bedömningen.

När det gäller etiska och normativa aspekter kring användningen av AI som beslutsstöd betona respondenterna vikten av den egna moraliska kompassen samt gällande regelverk. Respondent 4 menar att bedömningen av om AI:s rekommendationer är etiskt hållbara grundas i den egna kunskapen och organisationens regler:

“Jag skulle säga att det har enbart att göra med min kunskap själv. [...] Jag skulle säga att vi alla har en ganska stark moralisk kompass. Och sen har vi såklart vilka regler ur lagperspektiv och allt möjligt som vi måste följa.”

Respondent 6 uttrycker ett liknande förhållningssätt och framhåller att AI inte används okritiskt:

“Eftersom den är baserad på våra regelverk så tar man ju för givet att den ska vara etisk. Men man ska inte lita på den liksom med blinda ögat så. Och det är just därför vi dubbelkollar allt. Eller oftast allt den säger eller ger för svar bara för att säkerställa att det här är rätt.”

Respondent 9 nämner att det kan vara svårt att veta på vilka grunder AI-beslutsstödet agerar, särskilt eftersom tjänsten tillhandahålls av en extern leverantör. För att skapa en ökad förståelse för hur verktyget hanterar etiskt känsliga situationer uppger respondent 9 att denne personligen har testat AI genom att ställa frågor:

“Men jag har faktiskt ställt några etiska brottsfrågor för att testa AI. Och jag kan säga att den har klarat sig alldeles utmärkt.”

Enligt respondent 9 svarade AI att det inte kunde bistå i ärenden som bedömdes som oetiska.

4.2.3 Arbetssystemteorin

Samtliga respondenter uttrycker sina arbetsroller och beslutsprocesser som strukturerade kring tydliga ansvarsområden, mandat och organisatoriska rutiner (bilaga 1 visar de frågor som ställdes till respondenterna. Arbetet präglas vidare av ett samspel mellan olika aktörer, där beslut sällan enbart fattas av en enskild individ utan ofta involverar kollegor, chefer eller formella beslutsinstanser som kreditkommittéer. AI beskrivs i detta sammanhang som ett stöd i de delar av arbetssystemet som rör informationsinhämtning, analysförberedelser och dokumentation. Samtliga respondenter menar vidare att AI används för att strukturera underlag, formulera beslutsmotiveringar samt effektivisera administrativa moment.

Samtidigt menar alla respondenter att AI inte fattar självständiga beslut inom dessa processer i nuläget. Ansvar för besluten ligger därför fortsatt hos de anställda, oavsett om AI har använts som stöd i processen.

Flera respondenter betonade vikten av att kunna stå för besluten gentemot både organisationen och kunden, vilket i sin tur påverkar hur och i vilken utsträckning AI används i arbetet. I detta sammanhang uttrycker respondent 7 följande:

“För det är ofta så att när ett beslut tas, är det jag som sätter mitt namn på det. Det är mitt namn som står på beslutet, och det är jag som har tagit det. Då vill jag inte känna att det egentligen är någon annan som har fattat beslutet och att jag bara sätter mitt namn på det. Även om jag förstår AI:s rekommendation, upplever jag att vi har varit delaktiga och att beslutet är fattat av oss själva.”

Detta perspektiv återkommer även i respondent 10:s beskrivning av organisationens roll i AI-användningen. Respondenten uppger att organisationen främst har tillhandahållit AI-verktyget, medan riktlinjer och uppföljning kring hur verktyget ska användas i det dagliga arbetet framstår som begränsade, vilket innebär att användningen i stor utsträckning formas av individens eget omdöme. Användningen av AI beskrivs därmed som beroende av enskilda medarbetares kompetens och initiativ snarare än av tydligt etablerade organisatoriska rutiner.

4.2.4 AI:s framkant

I den insamlade empirin reflekterar respondenterna kring hur de uppfattar AI:s centrala egenskaper, det vill säga autonomi, lärande och beslutslogik, samt vilka utmaningar dessa egenskaper medför i det dagliga arbetet, inklusive upplevd förståelse, kontroll och

användningsvilja (se bilaga 1 för att se vilka frågor som ställdes). Respondent 3 beskriver exempelvis hur användningen successivt har ökat i takt med att AI-beslutsstödet inte uppvisat några uppenbara fel, vilket enligt respondenten kan bidra till att AI får en mer central funktion framöver:

“På senaste tiden tror jag att vi har börjat förlita oss på den lite mer och mer. Så som du säger historiskt, om vi ser att misstag inte sker och det funkar som det ska. Då tror jag att den kommer ha en mycket större roll i framtiden.”

Gällande situationer där AI:s beslutslogik har varit svår att förstå menar respondent 7 att:

“Det som jag vet om AI är ju att man behöver prata mycket med den och träna den. Så det är ju när man kanske ställer en dålig prompt. Då blir det att jag inte får fram det jag vill och då kommer inte den fatta heller vad jag vill och då blir det ju inte bra på något sätt liksom. Så att det är ju när man inte lägger en bra prompt.”

Vidare beskriver respondenterna olika förhållningssätt till AI-systemens lärande över tid. Exempelvis har respondent 10 aktivt begränsat AI:s möjlighet att anpassa sig, genom att stänga av personalisering i verktyget. Detta har respondenten gjort för att få neutrala svar som inte är påverkade av individens tidigare användning. Samtidigt upplever respondent 8 ingen tydlig förändring i AI:s prestanda över tid, utan menar att AI:s ökade träffsäkerhet snarare beror på egen tillvänjning och ökad vana vid verktyget. Respondenten betonar att det är det egna arbetssättet som har förändrats, snarare än att AI har anpassat sig.

4.2.5 Tillit och transparens

Respondenterna beskriver deras bedömningar av AI:s förmåga, välvilja och pålitlighet i beslutsstöd och bilaga 1 visar de frågor som ställdes till respondenterna. Gällande insyn i AI:s beslutsprocess och i vilka situationer insynen är tillräcklig samt otillräcklig, menar respondent 3 följande:

“[...] om det är ett svar som man känner igen och man tänker att den har bekräftat min tanke. Då skulle jag lita på den. Men ifall den säger något helt annat. [...] Då skulle jag ifrågasätta.”

I relation till insynen i beslutsprocesserna hos AI beskriver respondent 2 detta:

“Alltså jag skulle säga att insynen i beslutsprocessen är nog obefintlig där vi är.”

Respondenten beskriver vidare att om AI används i högre grad kring beslut, skulle ökad insyn i verktyget behövas:

“Det är ju väldigt många parametrar som vi skulle behöva känna oss väldigt trygga med att den tar hänsyn till. Så då skulle jag behöva ett väldigt stort krav på insyn eller i alla fall att jag får det bekräftat från organisationen internt att nu har AI co-pilot koll på liksom vår värdegrund och hur vi bedriver vår verksamhet och hur vi ser på våra kunder, och det känns avlägset. Där tror jag verkligen att människor som vi ser på det i alla fall behövs för att själva ställningstagandet.”

Vidare beskriver respondent 8 att om AI svarar otydligt, så brukar respondenten ställa följdfrågor:

“Jag ställer mycket följdfrågor. Framför allt är det kanske någonting man reagerar på i svaret och då vill man kanske vidareutveckla vad faktiskt chatten, eller vad man ska säga botten den menar. Så jo, det gör jag absolut.”

Till skillnad från detta, menar Respondent 5 följande:

“Jag har tyvärr inte så bra tålamod. Så det blir bara att om det till exempel är en text eller en som kanske får ett faktafel eller någonting sådant. Då bara ändrar jag det själv i stället för att typ så här. Man skulle ju kunna egentligen bara fortsätta prata med AI och försöka få den att typ så här fatta själv. Men det har jag typ inte tålamod till.”

Vidare belyser respondent 5 även detta:

“[...] tyvärr så är ju AI-systemet typ det långsammaste och trögaste som finns så att det tar, även om man skulle liksom upptäcka någonting, och man feedbackar det, så tar det lite tid innan det blir åtgärdat.”

Med andra ord påverkar AI:s upplevda effektivitet användningen av det. Vidare reflekterar Respondent 7 kring hur tydligt det är vilka regler, rutiner och ansvar som gäller kring AI inom organisationen på följande sätt:

“Jag upplever att det är ganska ... Än så länge otydligt och att vi inte har haft någon större diskussion kring det. I alla fall inte på min enhet eller mitt kontor som vi sitter på. Och det är nog därför jag inte använder det så mycket heller än så länge i mitt arbete.”

På frågan hur detta påverkar tilliten till AI svarar respondenten följande:

“Sen huruvida jag känner mig trygg att använda det. Alltså, det är klart att man blir ju lite mer mån om vad man i så fall skriver till en AI. Och på så sätt kan det bli svårare att få bra svar om jag känner mig begränsad.”

Med andra ord beskriver respondenten att avsaknaden av organisatoriska riktlinjer påverkar respondentens användning av AI som beslutsstöd.

4.2.6 Förklarbar AI

Respondenterna beskriver hur AI:s förklaringar presenteras och upplevs i praktiken, samt hur graden av förklarbarhet påverkar förståelse, beslutsfattande och fortsatt användning av AI-baserade beslutsstöd. Bilaga 1 visar de frågor som respondenterna besvarade.

Empirin visar att bankanställda tillämpar olika strategier för att hantera situationer där AI:s rekommendationer saknar tydliga förklaringar. I flera fall påverkar avsaknaden av förklarbarhet inte det fortsatta arbetet i sig, utan innebär i stället att rekommendationerna inte betraktades som tillräckliga för det slutgiltiga beslutsunderlaget. Respondent 6 säger följande gällande avsaknaden av förklaring från AI och hur det påverkar den fortsatta beslutsfattningen:

“Alltså jag skulle nog inte säga att det gör det eftersom vi inte litar på den fullt ut än. Just för att den är så pass ny.”

Vidare menar respondenten även att det i en sådan situation sker en fortsatt konversation med AI där syftet är att få den att förklara tydligare:

“[...] Då pratar jag med den igen. Alltså då ber jag den att förklara tydligare liksom. Ja, det är så jag skulle göra. Jag skulle inte ta det som att det är det slutgiltiga utan då hade jag ju försökt få fram mer information från den liksom.”

Samtidigt menar respondent 7 att det är svårt att bedöma AI:s faktiska förmåga, då kunskapen om hur långt AI-utvecklingen har kommit upplevs som begränsad:

“För att jag, och som sagt jag har ju inte så bra koll på hur långt AI-utvecklingen har kommit och hur duktig den är egentligen.”

Gällande förståelse för hur AI behandlar information och kommer fram till förslag säger respondent 8 att det är svårt, vilket går att se på följande uttalande:

“[...] det är klart att det är lite som att förstå hur stort universum är. Men det är väl klart att man förstår att ställer jag frågan tio gånger så kommer givetvis svaret bli lite närmare sanningen kanske.”

Sammantaget visar empirin att AI i bankernas beslutsprocesser främst används som ett stödjande verktyg snarare än som en självständig beslutsfattare. Respondenternas beskrivningar pekar på återkommande teman som begränsad insyn i AI:s beslutslogik, situationsbunden tillit, otydliga organisatoriska riktlinjer samt ett starkt fokus på mänskligt ansvar.

5. Analys

Studiens empiriska material analyseras genom en kombination av induktiv och deduktiv ansats. Analysen tar sin utgångspunkt i det empiriska materialet för att identifiera återkommande mönster och teman, samtidigt som resultaten tolkas i relation till det teoretiska ramverket och det deduktiva konceptuella ramverk som utvecklats för studien. På så sätt möjliggörs en analys där empiri och teori samspelar och där samband mellan tekniska, organisatoriska och mänskliga aspekter av AI-användning kan belysas.

Analysarbetet har strukturerats med stöd i den upprättade kodningsboken, där återkommande mönster i materialet identifierats genom både induktiv (se bilaga 3) och deduktiv kodning (se bilaga 2). Dessa mönster har därefter samlats i kategorier och sammanförts till övergripande teman (se bilaga 4), vilka utgör grunden för analysen och dess inkluderade delar. De teman som identifierats presenteras i avsnitt 5.1-5.3 och rör transparensens betydelse för tillit till AI som beslutsstöd, tilliten som situationsbunden, begränsad och erfarenhetsbaserad samt det mänskliga ansvaret som gräns för tillit. Kategorierna som ligger till grund för dessa teman presenteras i respektive avsnitt.

5.1 Transparensens betydelse för tillit till AI som beslutsstöd

Empirin visar att transparens utgör ett återkommande tema i respondenternas beskrivningar av AI som beslutsstöd. De anställda betonar i avsnitt 4.2.5 genomgående behovet av insyn i hur AI-system tar fram rekommendationer, särskilt i relation till ansvar, bedömning och möjligheten till efterhandsgranskning av beslut. Transparens framträder dock inte enbart som en teknisk egenskap hos AI-systemen, utan som ett fenomen som tar form i samspelet mellan teknik, människa och organisatoriska strukturer.

Följaktligen är kategorierna som ligger till grund för detta tema:

- Förklarbar AI som praktisk strategi, inte tekniskt krav.
- Transparens genom interaktion för att förstå och värdera AI:s rekommendationer.
- Organisatoriska strukturer och begränsningar som formar tillit och transparens

Vidare analyseras kategorierna i följande avsnitt där 5.1.1 analyserar behovet av insyn i AI:s beslutslogik, 5.1.2 fokuserar på hur transparens skapas genom interaktion med AI och 5.1.3 belyser organisationens roll i att möjliggöra eller begränsa transparens. Tillsammans visar

dessas avsnitt hur transparens, i olika former, påverkar bankanställdas tillit till AI i beslutsfattande processer.

5.1.1 Insyn och transparens i AI:s beslutslogik

Majoriteten av respondenterna betonar vikten av att förstå varför AI föreslår en viss lösning eller rekommendation, vilka faktorer som har vägts in samt vilka källor eller underlag som ligger till grund för resultatet. Exempelvis beskriver respondent 6 vikten av att AI hänvisar till korrekta källor för att rekommendationerna ska kunna användas i arbetet. Denna efterfrågan på insyn kan förstås som ett uttryck för algoritmisk transparens, där användarna eftersöker en begriplig och praktiskt användbar förståelse av hur AI bearbetar information och genererar rekommendationer.

Samtidigt beskriver respondent 2 att insynen i AI:s beslutsprocess upplevs som obefintlig och att de anställda i nuläget saknar förståelse för hur AI genererar beslutsstöd. Detta leder till att AI inte används som självständig beslutsfattare, utan snarare betraktas som ett begränsat stöd i arbetet. Respondenten menar vidare att en ökad delaktighet från AI:s sida inom beslutsfattande skulle kräva en betydligt större insyn i verktyget än vad som finns i nuläget, för att möjliggöra både förståelse och ansvarstagande. Detta överensstämmer med Maier, Menold och McComb (2022) som menar att användarnas förtroende för AI i finansiella sammanhang påverkas av hur transparent AI upplevs vara.

Enligt Berente et al. (2021) kan detta förstås mot bakgrund av att AI:s ökade autonomi, kontinuerliga lärande och upplevda oförklarbarhet skapar ett växande avstånd mellan AI:s kapacitet och användarnas förståelse. Författarna beskriver AI som ett dynamiskt gränsområde där beslutslogiken successivt rör sig bortom användarnas överblick. När AI-system hanterar alltmer komplexa beslutsproblem ökar därmed avståndet mellan vad systemen kan göra och användarnas möjlighet att förstå hur rekommendationer genereras. I denna studie tar detta sig uttryck i respondenternas uttalade behov av transparens som ett sätt att hantera osäkerhet kring AI:s resonemang och begränsningar. Respondent 7, 8 och 10 beskriver att rekommendationer som saknar tydliga förklaringar, källhänvisningar eller insyn i regel inte används som beslutsunderlag, utan kräver komplettering. I sådana situationer uppger respondenterna att de i stället vänder sig till kollegor snarare än till AI-beslutsstödet.

I denna kontext framstår algoritmisk transparens som en nödvändig mekanism för att minska osäkerheten kring användningen av AI som beslutsstöd. När insynen i AI-beslutsstödet brister påverkas tilliten negativt, eftersom användarna inte upplever sig kunna bedöma dess förmåga och därmed inte fullt ut kan stå bakom beslutsunderlaget, vilket enligt Söllner et al. (2016) är centralt för att tillit till tekniska system ska utvecklas. Respondenternas efterfrågan på källhänvisningar och insyn i beslutslogiken visar därmed att tillit till AI inte enbart grundas i de rekommendationer som tillhandahålls, utan i möjligheten att förstå, granska och ta ansvar för det beslutsunderlag som presenteras. Detta understryks ytterligare av att respondenterna i första hand efterfrågar ökad insyn och förklaringar, snarare än mer precisa eller avancerade rekommendationer från AI. I linje med Söllner et al. (2016) kan detta förstås som att bristande algoritmisk transparens begränsar användarnas vilja att göra sig sårbara inför tekniken, vilket i praktiken påverkar vilka typer av uppgifter AI används som beslutsstöd i.

Bristande algoritmisk transparens påverkar därmed inte enbart tilliten till AI, utan fungerar även som en begränsande faktor för hur AI faktiskt integreras och används i arbetssystemets beslutsprocesser. Detta tydliggörs eftersom samtliga respondenter beskriver att AI som beslutsstöd främst används i situationer där det visat sig fungera tidigare, vilket är processer som anses vara standardiserade och enkla. Exempelvis beskriver respondent 6 en selektiv användning av AI som beslutsstöd, där tekniken används för avgränsade och enklare uppgifter, men där användningen avslutas när verktyget upplevs som ineffektivt eller tillhandahåller felaktiga rekommendationer. I sådana situationer väljer respondenten att manuellt korrigera resultatet i stället för att fortsätta dialogen med AI, vilket motiveras av tidsbrist och bristande tålamod.

Liknande mönster framkommer även hos flera andra respondenter, där respondent 7 menar att otydliga organisatoriska riktlinjer kring AI bidrar till att verktyget används i mindre utsträckning, trots att AI som beslutsstöd har en potentiell nytta i arbetet. Även respondent 5 och 8 beskriver olika anpassningar som begränsar AI:s roll i beslutsprocessen. Dessa anpassningar innebär att AI:s rekommendationer korrigeras eller kompletteras manuellt och i vissa fall att dialogen avslutas när svaren upplevs som otydliga eller felaktiga. Respondent 6 betonar på liknande sätt att AI:s rekommendationer regelbundet kontrolleras för att säkerställa deras korrekthet.

Samtidigt beskriver Alter (2013) arbetssystem som helheter där deltagare använder information och teknik för att utföra arbetsuppgifter. I arbetsuppgifter som är kopplade till ansvar kan detta förstås som att informationen behöver vara begriplig för den som utför arbetet. När AI används som beslutsstöd blir insyn i dess informationsunderlag därmed central för att användarna ska kunna stå bakom resultatet av arbetet. Bristande begriplighet riskerar att skapa obalans mellan arbetssystemets olika delar, vilket Alter (2013) menar kan leda till motstånd mot tekniken. Respondenternas upplevelse av begränsad insyn i AI:s beslutslogik indikerar att informationselementet (se tabell 1) inte fullt ut stödjer beslutsprocessen, vilket i sin tur påverkar tilliten till AI. Ur ett arbetssystemperspektiv innebär detta att den information som AI genererar inte upplevs som tillräcklig begriplig eller spårbar för att möjliggöra fullt förtroende.

Således går det att tolka det som att respondenterna undviker att göra sig sårbara inför tekniken, vilket enligt Söllner et al. (2016) kan tolkas som ett uttryck för begränsad tillit till att systemet agerar på ett tillförlitligt sätt. Samtidigt menar Alter (2013) att dessa mönster kan förstås som framväxande anpassningar till en teknik som ännu inte är fullt integrerad i systemets ansvarsfördelning och informationsflöden. Snarare än att avvisa AI helt, justerar respondenterna hur och när tekniken används. Detta innebär att AI blir ett stödjande verktyg som enbart används i vissa situationer där erfarenheten säger att det fungerar som resultat av den bristande insynen, snarare än en central komponent i beslutsfattandet. Sammantaget visar detta hur bristande algoritmisk transparens kan leda till ett praktiskt motstånd som påverkar de anställdas tillit och faktiska användning av AI som beslutsstöd.

5.1.2 Transparens genom interaktion med AI

Utöver teknisk insyn beskriver flera respondenter hur transparens i praktiken skapas genom interaktion med AI, där otydliga svar från AI hanteras genom fortsatt dialog i form av följdfrågor och omformuleringar. Respondent 10 framhåller att transparens ibland skapas genom interaktion, där respondenten aktivt arbetar med AI:s svar för att uppnå en mer ändamålsenlig förståelse. Samtidigt framgår det att denna interaktion är selektiv och beroende av uppgiftens karaktär, där respondenten främst väljer att fortsätta dialogen när AI kan bidra med en grundläggande förståelse som respondenten saknar. Respondent 10 uppger även att interaktionen med AI avslutas när det inte bedöms vara tillräckligt användbart eller kompetent i relation till uppgiften.

Å andra sidan beskriver Respondent 1 hur transparens uppstår genom aktiv interaktion med AI, där följdfrågor och ett gemensamt resonemang används för att utveckla ett situationsanpassat svar. Detta beskriver även Respondent 7, som även hävdar att bristande tydlighet från AI hanteras genom aktivt ifrågasättande av dess svar. Interaktionen används i detta fall som ett sätt att testa AI:s förmåga för att kunna avgöra om rekommendationerna är tillräckligt tillförlitliga för att användas. När denna dialog inte längre ger önskad klarhet eller kvalitet väljer även denna respondent att avbryta interaktionen och i stället använda ett alternativt tillvägagångssätt vid beslutsfattandet.

Genom att ifrågasätta AI:s resonemang och samtidigt tillföra egen förståelse och kontext, skapas en dialog där både den bankanställdes och systemets perspektiv vävs samman. Detta kan enligt Haresamudram, Larsson och Heintz (2023) förstås som ett tydligt exempel på interaktionell transparens, där förståelsen för AI:s rekommendationer inte är given från början, utan formas successivt i samspel mellan människa och system, och där tillit därefter kan skapas beroende på hur AI-systemet hanterar följdfrågorna.

Samtidigt visar dessa beskrivningar från respondenterna hur transparens inte enbart handlar om tekniska förklaringar, utan om möjligheten att genom interaktion utforska AI:s resonemang, begränsningar och relevans i den aktuella situationen. Med andra ord framstår transparens i dessa situationer inte enbart som en egenskap hos AI, utan något som skapas och utvecklas i interaktioner mellan respondenten och AI. Genom dialogen ges användarna möjlighet att successivt bygga förståelse för AI:s bedömning och därigenom värdera dess rekommendationer.

Vidare beskriver, som tidigare konstaterat, några av respondenterna (5, 6 & 8) att interaktionen med AI avslutas när rekommendationerna upplevs som felaktiga, otydliga eller ineffektiva. Respondenterna kopplar även avbruten interaktion till organisatoriska och tekniska förutsättningar, där brister i bankens AI-system upplevs påverka arbetsflödet negativt. Detta tydliggör respondent 6 som uttrycker en frustration över organisationens AI-system, som beskrivs som långsamt och trögt, samt att förbättringar inom systemet tar lång tid för organisationen att genomföra. Dessa beskrivningar kan tolkas som ett motstånd till tekniken som yttrar sig i vissa situationer. Enligt Alter (2013, s. 77) uppstår ett sådant motstånd när det finns en obalans mellan arbetssystemets olika delar, som exempelvis när tekniska lösningar inte är anpassade till befintliga arbetsprocesser eller deltagarnas roller och kompetens. I detta fall kan

motståndet förstås som ett uttryck för att AI-systemet inte fullt ut stödjer arbetets krav, vilket leder till att användarna korrigerar, kompletterar eller avstår från att använda beslutsstödet rekommendationer.

Dessa beskrivningar om ineffektiva system ligger även i linje med Lamb och Kling (2003) som beskriver användaren som en social aktör vars användning av teknik formas av organisatoriska resurser, roller och institutionella villkor snarare än enbart individuella attityder. När dessa resurser inte stödjer smidig och effektiv interaktion begränsas möjligheten att utveckla interaktionell transparens genom dialog, vilket i sin tur bidrar till att användningen av AI förblir selektiv och situationsbunden. I förlängningen innebär detta att tilliten till AI som beslutsstöd blir villkorad, snarare än stabil, vilket McKnight et al. (2011) beskriver.

Samtidigt beskriver respondent 3 att organisationen troligtvis har svårigheter kring att införa AI som resultat av bankens storlek. Detta indikerar att interaktionell transparens, trots sin potential att minska osäkerhet kring AI:s rekommendationer som resultat av den ökade insynen, är beroende av ett arbetssystem som möjliggör smidig och effektiv interaktion. När sådana förutsättningar saknas, i form av tekniska begränsningar eller utebliven kommunikation, minskar möjligheten till interaktionell transparens eftersom respondenterna upplever svårigheter kring att interagera med AI.

Detta perspektiv lyfts genom respondent 2, som beskriver att möjligheten att interagera med AI genom följdfrågor inte har varit något som aktivt har övervägts i användningen av verktyget. I relation till detta beskriver Haresamudram, Larsson och Heintz (2023) att interaktionell transparens förutsätter att människor engagerar sig i dialog med systemet för att successivt skapa begriplighet kring AI:s resonemang och rekommendationer. När denna medvetenhet saknas begränsas möjligheten att utveckla transparens genom interaktion, trots att tekniken formellt är implementerad i organisationen. Detta indikerar att interaktionell transparens inte enbart är beroende av systemets tekniska utformning, utan även av användarnas förståelse för hur AI kan användas.

Enligt arbetssystemteorin av Alter (2013) visar detta hur organisationen påverkar hur tekniken används och förstås i praktiken. Om organisationen inte kommunicerar tillräckligt kring hur AI kan, och får, användas som beslutsstöd begränsas även användarnas möjlighet att utveckla transparens genom interaktion. Detta beskriver även respondent 1, som menar att

interaktionen med AI förutsätter en viss grundläggande kunskap hos den anställda gällande både beslutssituationen samt AI:s begränsningar för att rekommendationerna ska bli korrekta. Detta indikerar att interaktionell transparens inte är tillgänglig på lika villkor för alla anställda, utan är beroende av den individuella kompetensen, erfarenheten och förmågan att aktivt delta i dialogen med systemet. Även om Alter (2013) genom arbetssystemlivscykelmodellen betonar betydelsen av framväxande förändringar i det dagliga arbetet, indikerar empirin att sådana anpassningar i detta fall hämmas av bristande organisatoriska ramar. Med andra ord kan detta begränsa användarnas möjlighet att experimentera samt utveckla förståelse genom interaktion, vilket i sin tur kan påverka den interaktionella transparensen.

När transparensen i interaktionen brister får användarna det svårare att bedöma AI:s förmåga, förstå dess rekommendationer och avgöra om det agerar i linje med organisatoriska värderingar. Enligt Söllner et al. (2016) är dessa dimensioner, förmåga, välvilja och integritet, avgörande för att tillit till ett tekniskt system ska utvecklas. I enlighet med det deduktiva konceptuella ramverket visar resultaten att interaktionell transparens påverkar hur tillit till AI formas i praktiken, snarare än att tillit enbart kan härledas till systemets tekniska egenskaper, vilket Haresamudram, Larsson och Heintz (2023) förespråkar.

5.1.3 Organisationens roll kopplat till transparens och tillit

Empirin visar att det råder osäkerhet bland de anställda kring hur AI får användas som beslutsstöd inom organisationen. Detta framkommer bland annat från respondent 7 och 8, som beskriver att otydlighet kring vad som är tillåtet leder till en mer restriktiv och selektiv användning av AI-verktyget. Denna osäkerhet kan förstås som ett uttryck för bristande social transparens, där avsaknaden av tydliga organisatoriska ramar skapar oklarhet kring ansvar, mandat och användningsområden för tekniken.

Följaktligen beskriver samtliga respondenter, med undantag för respondent 10, att en tydligare och mer positiv organisatorisk inställning till AI, genom uttalat stöd och godkännande av användning i flera arbetsprocesser, bidrar till ökad tillit och trygghet i användningen. Respondent 5 framhåller att organisationens positiva kommunikation kring AI främjar tillit och att användningen sannolikt hade sett annorlunda ut om organisationen inte uppmuntrat till användning av AI som beslutsstöd. Samtidigt menar Respondent 7 att både tillit och upplevd

trygghet skulle öka om organisationen tydligare tillät hantering av mer känslig information, som kunduppgifter, i verktyget.

Även respondent 10 framhåller att ett tydligare organisatoriskt och regulatoriskt godkännande skulle kunna öka användningen av AI, eftersom det skulle innebära att det yttersta ansvaret i högre grad upplevs ligga hos organisationen snarare än hos den enskilda medarbetaren. Å andra sidan tydliggör respondenten att organisationen inte har en påverkan på respondentens tillit till AI. Detta resonemang visar på en tydlig åtskillnad mellan organisatoriskt stöd för användning av AI och individuell tillit till beslutsstödet. Beskrivningen av respondent 10 illustrerar därmed att tillit till AI inte kan institutionaliseras i vissa fall, utan att den formas genom den faktiska användningen av systemet, vilket förklaras genom McKnight et al. (2011) teori om erfarenhetsbaserad tillit som utvecklas genom faktisk användning och upplevd funktionalitet.

Ur ett arbetssystemperspektiv kan detta förstås som ett uttryck för samspelet mellan tekniska system, mänskliga aktörer och organisatoriska strukturer (Alter 2013, s. 76). När transparens och tydlighet saknas på organisatorisk nivå försvåras integreringen av AI som en fungerande komponent i arbetssystemet, vilket påverkar hur verktyget används och upplevs. I linje med Alter (2013, s. 77), som menar att obalans mellan arbetssystemets olika delar kan leda till motstånd mot tekniken, beskriver respondent 7 att AI inte används i den utsträckning det hade kunnat om organisationen varit tydligare i kommunikationen kring AI.

Bristande kommunikation från organisationens sida blir även tydligt i respondents 2 beskrivning som tidigare nämnts, där möjligheten att interagera med AI genom följdfrågor inte har varit något som aktivt har övervägts i användningen av verktyget. I detta fall kan avsaknaden av interaktionell transparens inte per automatik förstås som ett resultat av AI:s tekniska begränsningar, utan snarare som en konsekvens av bristande social transparens. När organisationen inte tydliggör hur AI kan användas som ett interaktivt verktyg och beslutsstöd, begränsas användarnas förståelse för vilka interaktionsmöjligheter som faktiskt finns. I enlighet med Haresamudram, Larsson och Heintz (2023) indikerar detta att interaktionell transparens förutsätter organisatoriska ramar som gör användarna medvetna om hur transparens kan skapas i praktiken. Det går därmed att tolka det som att avsaknaden av sådan vägledning från organisationen resulterar i att användningen av AI blir mer begränsad.

Även om AI-tekniken har implementerats formellt, indikerar empirin att organisatoriska riktlinjer, ansvarsfördelning och utbildningsinsatser ännu inte är fullt utvecklade. Som en följd av detta uppstår utrymme för framväxande förändringar, där anställda anpassar och i vissa fall begränsar sin användning av AI utifrån egen bedömning snarare än utifrån etablerade rutiner. Enligt Alter (2013) kan sådana individuella anpassningar påverka hur organisatoriska rutiner utvecklas över tid.

Detta fenomen illustreras av respondent 10, som på eget initiativ har informerat kollegor om hur AI-verktyget kan användas, vilket i sin tur har bidragit till en ökad användning av verktyget på detta sätt inom organisationen. På liknande sätt framhåller respondent 3 att kunskap om AI-användning sprids i informella sammanhang, exempelvis genom gruppchattar där medarbetare utbyter erfarenheter, tips och praktiska lösningar.

Sammantaget visar detta hur kunskap om AI-användning i hög grad sprids genom informella och individdrivna initiativ snarare än genom formella strukturer. I linje med arbetssystemlivscykelmodellen enligt Alter (2013) (se figur 2) kan detta förstås som en framväxande förändring (*emergent change*), där lokala anpassningar utvecklas i avsaknad av tydliga organisatoriska riktlinjer. Trots att AI-verktyget är formellt implementerat formas dess användning i praktiken genom informella rutiner, manuella kontroller och selektiv användning, snarare än genom standardiserade arbetssätt.

Detta blir även tydligt eftersom respondent 10 uppger att organisationen enbart gav de anställda AI-verktyget utan att tillhandahålla tydliga ramar, strukturer eller uppföljning kring hur verktyget förväntas användas i det dagliga arbetet. Samtidigt innebär en sådan utveckling att användningen av AI blir beroende av enskilda medarbetares kompetens och initiativ, vilket kan leda till variation i hur tekniken används inom organisationen och därmed begränsa AI:s integrering i arbetssystemets kärnprocesser.

Följaktligen innebär detta att social transparens inte enbart påverkar tilliten till AI, utan även hur verktyget används i praktiken. När organisatoriska riktlinjer och ansvarsfördelning är otydliga begränsas användningen och transparens fungerar därmed som en avgörande faktor för hur och om de anställda använder AI.

5.2 Erfarenhetsbaserad, situationsbunden och begränsad tillit till AI som beslutsstöd

Respondenternas erfarenheter visar, som delvis presenterats i avsnitt 4.2.2, att tilliten till AI som beslutsstöd inte är generell eller stabil, utan varierar beroende på arbetsuppgiftens karaktär, kontext och de anställdas tidigare erfarenheter av AI-beslutsstödet. Tilliten framträder som högre i enklare och mer standardiserade uppgifter, samtidigt som den begränsas i mer komplexa beslutsprocesser. I enlighet med McKnight et al. (2011) kan tilliten därför i flera fall förstås som situationsbunden, men även till viss del erfarenhetsbaserad genom återkommande användning och möjlighet till verifiering av AI:s rekommendationer.

Vidare är kategorierna som ligger till grund för detta tema:

- Tillit till AI som växer genom erfarenhet och upplevd förmåga.
- Tillit till AI som situationsbunden och begränsad.

Dessa former av tillit analyseras vidare i avsnitten om erfarenhetsbaserad tillit (5.2.1) respektive tillit som situationsbunden och begränsad (5.2.2).

5.2.1 Erfarenhetsbaserad tillit

Empirin visar att tilliten till AI ökar i takt med att respondenterna (2, 5 & 8) ser att beslutsstödet fungerar i praktiken. Detta ligger i linje med McKnight et al. (2011), som beskriver hur tillit utvecklas genom faktisk användning och formas av hur funktionell, hjälpsam och tillförlitlig tekniken upplevs över tid. När de bankanställda ges möjlighet att återkommande interagera med AI och verifiera dess rekommendationer skapas därmed förutsättningar för en mer stabil och erfarenhetsbaserad tillit, som grundas i konkreta erfarenheter snarare än generella antaganden om tekniken.

När interaktionen med AI däremot begränsas till vissa typer av uppgifter, eller avslutas i ett tidigt skede, minskar möjligheten att bygga upp den erfarenhet som krävs för att tilliten ska utvecklas över tid. Detta framkommer exempelvis när respondenterna 7 och 10 beskriver hur otydliga organisatoriska ramar leder till en mer restriktiv användning av AI. På liknande sätt menar respondenterna 5, 6 och 8 som tidigare nämnt (se avsnitt 5.1.2) att upplevd ineffektivitet hos AI innebär att de väljer att avsluta interaktionen och i stället korrigera manuellt.

Ur ett arbetssystemperspektiv (se tabell 1) kan detta förstås som en begränsning i systemets miljö, där otydliga organisatoriska ramar påverkar hur och i vilken omfattning AI integreras i det dagliga arbetet, men även som en brist i infrastrukturen (Alter 2013). Detta indikerar en obalans mellan elementen i arbetssystemet (se tabell 1), där systemets tekniska komponent inte är i linje med dess miljö och infrastruktur, vilket enligt Alter (2013) hämmar systemets funktion. När användningen inte tydligt stöds organisatoriskt begränsas möjligheten till återkommande interaktion, vilket i sin tur hämmar utvecklingen av erfarenhetsbaserad tillit. Tilliten riskerar därmed att förbli situationsbunden snarare än generell, som tidigare nämnts, där AI uppfattas som tillförlitlig i vissa avgränsade sammanhang men inte utvecklas till ett mer etablerat och integrerat beslutsstöd i arbetet.

Samtidigt visar empirin att när organisatoriska förutsättningar möjliggör återkommande och effektiv användning, utvecklas tillit till AI gradvis genom faktisk erfarenhet. Respondenterna 2, 3 och 7 beskriver hur positiva och konsekventa erfarenheter av AI bidrar till en ökad tillit över tid, även om de menar att tilliten inte är total. Denna utveckling påverkas även av organisationens kommunikation kring AI, där tydligare stöd och uttalat godkännande skapar trygghet i användningen och därmed underlättar uppbyggnaden av erfarenhetsbaserad tillit.

Detta resonemang nyanseras dock av respondent 10, som menar att organisationens inställning inte har någon direkt påverkan på den individuella tilliten till AI. Respondenten framhåller att tilliten hade påverkats i högre grad om externa aktörer, som Finansinspektionen, tydligare uppmuntrade användning av AI och beskrev tekniken som ett pålitligt verktyg. Detta indikerar att även erfarenhetsbaserad tillit kan vara villkorad i en reglerad kontext, där individuella erfarenheter av systemets funktionalitet behöver kompletteras med externt och institutionellt godkännande för att tilliten ska kunna bli mer generell.

I linje med Söllner et al. (2016) kan denna erfarenhet av respondent 10 förstås som att bedömningen av AI:s integritet i högre grad baseras på externa regelverk och legitimerade aktörer, än på den interna kommunikationen i organisationen. Ur ett arbetssystemperspektiv beskriver Alter (2013) hur arbetssystemets miljö, i form av regulatoriska krav och tillsyn, påverkar hur teknik uppfattas och används i praktiken. Resultatet indikerar därmed att organisatoriskt stöd kan underlätta användning, medan institutionellt godkännande är avgörande för att tillit ska utvecklas i mer ansvarstygda beslutssituationer.

Samtidigt menar McKnight et al. (2011) att frånvaron av långsiktig erfarenhet begränsar möjligheten att utveckla stabil tillit, vilket innebär att de som interagerar med systemet i högre grad förlitar sig på verifiering och kontroll av dess rekommendationer. Detta återspeglas i respondenternas beskrivningar av en kvarstående försiktighet, särskilt eftersom AI uppfattas som en relativt ny teknik inom bankernas beslutsprocesser, vilket respondent 1, 8 och 10 belyser. Denna försiktighet går återigen att relatera till Tomkins (2001), som beskriver ett samband mellan detaljerad informationsinsamling, kontroll och tillit. När tilliten är begränsad ökar behovet av information, insyn och verifiering, medan högre tillit minskar behovet av omfattande kontrollmekanismer. Denna form av tillit kan förstås som situationsbunden, då den är knuten till det specifika beslutsögonblicket och den aktuella arbetsuppgiften snarare än till en generaliserad tillit till tekniken.

5.2.2 Tillit som situationsbunden och begränsad

Empirin visar att tilliten till AI som beslutsstöd inte är konstant, utan varierar beroende på situation, arbetsuppgift och de anställdas tidigare erfarenheter av systemet. Detta går att se eftersom samtliga respondenter beskriver att AI upplevs fungera väl i enklare och mer repetitiva arbetsmoment, som informationssökning, sammanställning av underlag och språklig bearbetning av texter. I dessa sammanhang uttrycks en högre grad av tillit, eftersom AI:s förmåga upplevs som tillräcklig för uppgiften och det faktum att resultaten och rekommendationerna relativt enkelt kan kontrolleras eller verifieras av beslutsfattaren.

Samtidigt framkommer att samtliga respondenter menar att ökad insyn och mer information om AI:s beslutsprocess skulle öka både användningen och tilliten till verktyget. Enligt Tomkins (2001) kan detta tolkas som ett uttryck för låg tillit, då bristande tillit skapar ett ökat behov av information, övervakning och transparens. Om respondenterna i stället hade haft en hög grad av tillit till AI, hade kraven på insyn och information inte varit lika framträdande.

Å andra sidan kan detta behov av information även förstås som en konsekvens av att flera respondenter (3, 4, 6 & 10) framhåller att AI som beslutsstöd upplevs som en relativt ny teknik i organisationen. Utifrån arbetssystemlivscykelmodellen (se figur 2), menar Alter (2013) att införandet av ny teknik sker som en process över tid, där ett arbetssystem gradvis formas genom planerade faser av initiering, utveckling och implementering. Under denna process etableras tekniska lösningar, rutiner, ansvarsfördelning och stödjande resurser som tillsammans

möjliggör den nya versionen av arbetssystemet. Med andra ord kan respondenternas situationsbundna tillit förstås som ett resultat av att AI-baserade beslutsstöd fortfarande befinner sig i en tidig eller ofullständigt genomförd förändringsfas. Även om tekniken har implementerats formellt, tyder empirin på att utvecklings- och implementeringsfaserna ännu inte fullt ut har inneburit tydliga rutiner, utbildning och ansvarsfördelning.

Detta tydliggörs även som resultat av att majoriteten av respondenterna hävdar att utbildningarna kring AI inte är obligatoriska och att de inte vet vem som ansvarar om incidenter sker kopplat till användningen AI, trots att några av respondenterna menar att information gällande detta finns att hämta på bankens intranät. När dessa delar saknas eller är otydliga skapas osäkerhet kring hur AI ska användas, vilket bidrar till att tilliten blir begränsad och knuten till specifika arbetsuppgifter snarare än generell.

Det faktum att AI-baserade beslutsstöd fortfarande befinner sig i en implementeringsfas framkommer genom att majoriteten av respondenterna beskriver tekniken som relativt ny i organisationen. Samtliga respondenter uttrycker däremot en förväntan om att AI kommer att användas i allt större utsträckning framöver. Detta förstärks av respondent 1, 8 och 10, som menar att deras erfarenheter av AI sannolikt hade sett annorlunda ut om några år, i takt med att tekniken blir mer etablerad och integrerad. Enligt McKnight et al. (2011) kan den tillit som respondenterna uttrycker förstås som en tidig, situationsbunden form av tillit, snarare än som erfarenhetsbaserad tillit.

Avsaknaden av långvarig erfarenhet av teknikens funktionalitet och prestanda innebär att respondenterna ännu inte har kunnat utveckla stabila, erfarenhetsbaserade uppfattningar gällande tillit till AI som beslutsstöd. I stället är tilliten beroende av tillgång till information, möjlighet till verifiering samt upplevd insyn, och dessa fungerar som centrala mekanismer för att möjliggöra tillit i det aktuella beslutsögonblicket.

I linje med detta kan respondenternas beskrivningar förstås som att tilliten till AI är högre i enkla och mer repetitiva arbetsuppgifter, där resultaten är lättare att utvärdera, medan tilliten begränsas i mer komplexa beslutsprocesser. I komplexa beslutssituationer, där bedömningar kräver kontextuell förståelse, etiska överväganden eller professionellt omdöme, framträder en mer försiktig och begränsad tillit. Respondenterna beskriver att AI:s rekommendationer i dessa fall används som ett stöd bland flera andra, snarare än som ett avgörande

beslutsunderlag. Detta bekräftas även av respondent 5, som menar att AI generellt är pålitlig i ärenden som håller sig inom tydliga ramar, men att tilliten minskar avsevärt när ärendena blir mer komplexa, vilket i vissa fall beror på det faktum att AI saknar mänskliga känslor.

Denna upplevelse kan relateras till Berente et al. (2021) som beskriver AI som ett dynamiskt gränsområde där systemen i ökande grad närmar sig mänsklig intelligens i hanteringen av komplexa beslutsproblem, utan att därmed besitta mänskliga egenskaper som känslor, omdöme eller moralisk förståelse. I detta perspektiv blir respondentens resonemang inte nödvändigtvis ett uttryck för ett misslyckande hos AI, utan snarare en markering av var gränsen för AI:s förmåga upplevs gå i praktiken. När beslutsfattandet kräver emotionell förståelse, etiska avvägningar eller situationsbunden tolkning, framträder skillnaden mellan mänskligt omdöme och AI:s resonemang tydligare. Å andra sidan kan detta förklaras genom Maier, Menold och McComb (2022) teori eftersom de hävdar att tillit till AI främst formas i skillnaden mellan AI-verktyget och människans prestation. När AI uppfattas prestera bättre än människor tenderar tilliten att öka, medan den minskar i situationer där den mänskliga bedömningen upplevs vara överlägsen. Samtidigt visar empirin att tillit inte enbart tar formen av följande eller avvisande, utan även kan uttryckas genom samspel.

Respondent 9 beskriver exempelvis hur olika bedömningar från AI och människa i vissa situationer jämförs och vägs mot varandra för att nå ett gemensamt ställningstagande, vilket uppfattas leda till ett bättre beslut. I linje med Maier, Menold och McComb (2022) kan detta förstås som att i situationer där människa och AI uppfattas prestera på likvärdig nivå, används båda resultaten som referenspunkter i beslutsprocessen. Tilliten tar här formen av ett dialogiskt beslutsstöd, snarare än okritiskt följande av AI:s rekommendationer.

Sammantaget visar detta att tilliten till AI är tydligt villkorad av det mänskliga ansvar som följer med beslutsfattandet. Även i situationer där AI uppfattas som tekniskt kompetent begränsas tilliten när besluten kräver mänskligt omdöme, etiska överväganden eller professionellt ansvar. Tillit till AI framträder därmed inte enbart som en fråga om AI:s prestation i sig, utan som en konsekvens av de krav på ansvarstagande som är knutna till den mänskliga rollen i beslutsprocessen.

5.3 Mänskligt ansvar som gräns för tillit till AI som beslutsstöd

Empirin visar i avsnitt 4.2.3 att mänskligt ansvar och professionellt omdöme utgör ett återkommande tema i respondenternas beskrivningar av AI som beslutsstöd. Respondenterna betonar genomgående att det slutliga ansvaret för beslut ligger hos den enskilda medarbetaren, vilket påverkar hur AI:s rekommendationer används och värderas i det dagliga arbetet.

Mänskligt ansvar framträder därmed som en tydlig gräns för hur tillit till AI-baserade beslutsstöd kan utvecklas i praktiken. I detta avsnitt analyseras hur ansvar och professionellt omdöme påverkar användningen av AI, samt vilken roll förklarbarhet spelar för att möjliggöra förståelse, granskning och ansvarstagande i beslutsprocesser.

Följaktligen är kategorierna som ligger till grund för detta tema:

- AI som stöd, men inte ersättning för mänskligt beslutsfattande.
- Individuellt ansvar gällande användning av AI.

Dessa aspekter behandlas i de följande avsnitten genom analys av förklarbarhet som stöd för professionellt beslutsfattande (5.3.1) och gränsdragningen mellan AI som beslutsstöd och mänskligt beslutsfattande (5.3.2).

5.3.1 Professionellt ansvar som grund för beslutsfattande

Under datainsamlingen betonar respondenterna att det slutgiltiga ansvaret för beslut fortsatt ligger hos den mänskliga aktören, det vill säga de anställda på bankerna. I relation till detta uttrycker respondent 7 och 10 en oro för att AI kan generera felaktiga rekommendationer eller beslut, särskilt eftersom det slutliga beslutet vilar på den anställda och det faktum att det är respondenterna själva som förknippas med beslutet. Detta perspektiv lyfts även av respondent 8, som framhåller att det är centralt att beslutsformuleringar och ställningstaganden kommer från den anställde själv, snarare än från AI-systemet, för att besluten ska uppfattas som korrekta.

Respondenternas ansvarstagande tar sig även uttryck eftersom de konsekvent beskriver hur AI:s rekommendationer kontrolleras, omformuleras eller används som ett av flera underlag snarare än som ett färdigt beslut. Respondent 6 beskriver exempelvis hur AI:s svar sällan accepteras okritiskt, utan att innehållet kontrolleras mot regelverk eller andra källor innan det används. På liknande sätt framhåller respondent 7 att det är avgörande att känna sig delaktig i

beslutsprocessen, snarare än att uppleva att beslutet kommer från AI enbart. I relation till arbetssystemets deltagare (se tabell 1) enligt Alter (2013) framträder det mänskliga ansvaret därför som centralt, eftersom det är deltagarna som förväntas tolka information, genomföra arbetssystemets processer och stå till svars för utfallet av de beslut som fattas.

I samtliga fall används AI även främst för att stödja och effektivisera beslutsfattandet och inte för att ersätta mänskligt ansvar eller professionellt omdöme. Det går därmed att tolka tilliten till AI som villkorad snarare än total. Även om flera respondenter menar att AI i vissa situationer kan ge felfria rekommendationer, är tilliten alltid beroende av de bankanställdas möjlighet att förstå och granska dessa, samt det egna ansvarstagandet där det anses vara viktigt att kunna stå bakom beslutet.

Även respondent 3 hävdar detta och beskriver att användningen av AI som beslutsstöd har ökat i takt med att det upplevts ge färre felaktiga rekommendationer, vilket kan bidra till att AI får en mer central funktion i arbetet över tid. Samtidigt framhåller respondenten att AI:s rekommendationer fortsatt behöver kontrolleras och inte kan användas okritiskt. Detta tydliggörs genom AI:s framkant (se figur 3) enligt Berente et al. (2021), där denna utveckling kan förstås som att AI-systemens ökade kapacitet inte nödvändigtvis följs av en motsvarande ökning i användarnas förståelse för hur rekommendationerna genereras. Empirin indikerar därmed att även när AI får en mer framträdande roll i beslutsprocesser kvarstår behovet av att positionera tekniken som beslutsstöd, där det slutliga ansvaret fortsatt ligger hos den anställde.

Enligt Söllner et al. (2016) kan detta förstås som att användarna i viss utsträckning kan bedöma AI-systemets förmåga, medan tillitsdimensionen integritet är svårare att tillskriva tekniken. Detta beror på det faktum att användarna saknar möjlighet att fullt ut bedöma AI:s ansvarstagande och etiska tillförlitlighet utan egen granskning, vilket framkommer i empirin, där respondent 7 beskriver att etiska överväganden i högre grad hanteras genom dialog med kollegor snarare än genom dialog med AI-verktyget. På liknande sätt uttrycker respondent 8 att det inte går att veta huruvida AI agerar etiskt i alla situationer, vilket medför ett behov av extra eftertanke och mänsklig kontroll. Även respondent 1 betonar vikten av emotionell och etisk förståelse i beslutsprocesser och lyfter risken för att AI saknar förmåga att hantera etiska aspekter, som exempelvis situationer som rör arbetssökande.

Sammantaget beskriver majoriteten av respondenterna ett behov av att själva granska och bedöma huruvida AI:s rekommendationer är etiskt hållbara, vilket indikerar att tilliten till systemets integritet är begränsad och förutsätter aktiv mänsklig kontroll. I linje med definitionen av Söllner et al. (2016) gällande tillit som en vilja att göra sig sårbar inför en annan part kan detta tolkas som att respondenterna inte är beredda att överlåta beslutsansvaret till tekniken och vara sårbara. Eftersom besluten behöver kunna motiveras och försvaras i efterhand, och respondenterna uttrycker vikten av att själva stå bakom beslutet, begränsas tilliten även i situationer där AI upplevs som kompetent och inte riskerar att äventyra beslutsfattandet.

Tilliten till AI blir därmed villkorad och situationsbunden, då det yttersta ansvaret inte kan delegeras till tekniken utan fortsatt ligger hos den anställde, eftersom det finns svårigheter kring att vara sårbar vid användningen av AI som beslutsstöd. Detta innebär att även när AI upplevs som kompetent begränsas tilliten av de krav som följer med professionellt ansvar och en reglerad verksamhet. Tillit till AI kan därmed inte förstås som ett substitut för mänskligt omdöme i nuläget enligt respondenterna, utan som ett villkorat stöd inom vissa processer, och som alltid förutsätter mänsklig kontroll.

Ur ett arbetssystemperspektiv enligt Alter (2013) kan det tolkas som att AI i dessa fall har integrerats i arbetssystemet med syfte att stödja informationsinhämtning och analys, snarare än att omfördela ansvar eller beslutsmandat, eftersom det endast används som ett beslutsstöd. Att de anställda fortsatt bär det slutliga beslutsansvaret innebär att AI inte integreras som en självständig aktör i arbetssystemet, utan som ett stödjande verktyg inom befintliga processer och aktiviteter. Detta bidrar till att AI accepteras som ett effektiviserande verktyg, vilket samtliga respondenter belyser är den huvudsakliga anledningen till användning av AI, samtidigt som professionellt omdöme och mänskligt ansvar fungerar som tydliga gränser för hur långt tilliten till tekniken sträcker sig.

5.3.2 Förklarbarhet som förutsättning för professionellt omdöme

Respondenterna beskriver att behovet av utförliga förklaringar från AI är nära kopplat till det personliga ansvar som följer med beslutsfattandet. Förklarbar AI framstår därmed som en förutsättning för att den anställde ska kunna förstå hela beslutsprocessen och göra en egen professionell avvägning innan det slutgiltiga beslutet fattas. Detta överensstämmer med Rai (2020, s. 137-139) som menar att förklaringar bidrar till ökat förtroende eftersom AI:s

resonemang blir mer begripliga för de bankanställda. Respondent 7 betonar exempelvis vikten av att AI redogör för hela beslutslogiken för att möjliggöra en helhetsförståelse av beslutet. Samtidigt framhåller även respondent 6 betydelsen av att AI hänvisar till korrekta källor för att rekommendationerna ska kunna användas i arbetet. Detta förklaras av Dwivedi et al. (2023, s. 194:1) som menar att bristande insyn i hur AI fattar beslut resulterar i att beslutsprocessen blir svår att tolka.

Vidare belyser tidigare Rai (2020, s. 141) att förklarbarheten inom AI inte nödvändigtvis bör maximeras, då alltför omfattande eller komplexa förklaringar kan leda till informationsöverbelastning och minskad användbarhet för användarna. Detta illustreras exempelvis av respondent 3, som beskriver att det inte är nödvändigt att förstå varje teknisk detalj i AI-systemets beräkningar, utan att det centrala är att kunna följa de huvudsakliga antaganden och faktorer som ligger till grund för rekommendationen. Enligt respondenten riskerar alltför detaljerade förklaringar att ta fokus från själva bedömningsarbetet samt minska effektiviteten och därmed minska AI:s praktiska värde i beslutsprocessen.

Det går därför att konstatera att det finns en skillnad i behovet av förklarbarhet enligt respondent 3 och 7. Medan respondent 7 betonar vikten av att kunna följa hela beslutslogiken för att känna sig trygg i det professionella ansvaret, framhåller respondent 3 att allt för omfattande förklaringar riskerar att bli ett hinder i arbetet. Detta indikerar att förklarbarhet i praktiken behöver vara funktionell och anpassad efter de anställdas behov, snarare än fullständig, för att stödja effektiv användning av AI som beslutsstöd, vilket stämmer väl överens med Rai (2020) teorier där för mycket förklaringar snarare kan innebära minskad användning av systemet.

Följaktligen beskriver respondent 6 att bristen på förklaringar inte nödvändigtvis påverkar den fortsatta beslutsfattningen, eftersom det från början inte finns en tillit till AI-verktyget då det är så pass nytt. Däremot menar respondenten att mer utförliga förklaringar som uppstår genom följdfrågor från respondenten, kan öka tilliten. Teorin gällande distinktionen mellan global och lokal förklarbarhet av Dwivedi et al. (2023) kan förklara respondentens resonemang, där avsaknaden av transparens inte upplevs som avgörande. I stället beskriver respondenten att lokal förklarbarhet, som uppstår genom följdfrågor kring enskilda beslut, kan vara tillräcklig för att öka tilliten i det fortsatta beslutsfattandet. Även detta resonemang visar vikten av att förklaringarna är anpassade efter individernas behov, där Rai (2020) menar att förenklade förklaringar i många fall är tillräckliga för att stödja förståelse och användning.

Ur ett arbetssystemperspektiv enligt Alter (2013) kan detta förstås som ett krav på att informationskomponenten ska vara ändamålsenligt utformad i relation till deltagarnas ansvar och arbetsprocesser. När förklaringar antingen är otillräckliga eller för omfattande uppstår en obalans i arbetssystemet, där informationen inte längre stödjer det professionella beslutsfattandet på ett effektivt sätt.

6. Diskussion

Denna studie har fokuserat på att undersöka hur bankanställda upplever transparens i AI-baserade beslutsstöd och hur den upplevda transparensen påverkar deras tillit till AI i det dagliga arbetet. Även om studiens analys huvudsakligen baseras på bankanställdas erfarenheter av generativ AI som beslutsstöd, förekommer i empirin även resonemang som rör andra typer av AI-tillämpningar inom bankverksamheten. Resultaten bör därför förstås som främst giltiga för generativ AI, samtidigt som flera av de identifierade mönstren kring transparens, tillit och ansvar även kan vara relevanta i relation till andra AI-baserade beslutsstöd.

Bredare samhällseliga aspekter som allmänhetens syn på AI, mediala diskussioner eller generella samhällsnormer framträder däremot inte som centrala i respondenternas beskrivningar av hur tillit formas i det dagliga beslutsarbetet och hur detta kan ha påverkat dem. Mot denna bakgrund bedöms ett samhällseligt perspektiv, det vill säga om samhället kan påverka respondenternas tillit till AI-baserade beslutsstöd, ligga utanför studiens empiriska fokus och analytiska räckvidd. Ur ett metodologiskt perspektiv är avgränsningen också motiverad. Kvalitativ forskning syftar i första hand till att skapa en djup och kontextuell förståelse av ett fenomen, snarare än att uppnå statistisk generaliserbarhet (Hennink, Hutter & Bailey 2020, s. 315-316). För att upprätthålla analytisk stringens bör slutsatser förankras i det insamlade materialet, eftersom teorier som ligger för långt från datan riskerar att leda till tolkningar som blir svåra att belägga (Hennink, Hutter & Bailey 2020, s. 263).

Dessutom har tidigare forskning visat att banker verkar inom en starkt reglerad institutionell kontext, där beslut om att införa nya tekniska lösningar i första hand formas av regulatoriska krav, tillsynsprocesser och strävan efter organisatoriskt rättfärdigande (Crawford 2017, s. 20). Mot denna bakgrund framstår ett fokus på bredare samhällseliga attityder till AI som mindre relevant för studiens analytiska syfte. Eftersom bankanställdas användning av och förhållningssätt till AI i hög grad formas av interna riktlinjer, regulatoriska krav och organisatoriska ansvarsförhållanden, är deras handlingsutrymme begränsat oberoende av hur tekniken uppfattas i samhället i stort.

Följaktligen visar analysen att transparens fungerar som en central förutsättning för att överhuvudtaget kunna utveckla och upprätthålla tillit i beslutsprocesser där AI används. Samtidigt framkommer att transparens inte enbart kan förstås som en teknisk egenskap hos AI-systemet,

utan snarare som något som formas i samspelet mellan människa, teknik och organisatoriska ramar. Ett genomgående resultat gällande transparens är att det uppfattas som nödvändigt för att AI:s rekommendationer ska kunna användas på ett sätt som är förenligt med de anställdas professionella ansvarstagande.

Datainsamlingen visar vidare att de bankanställda i begränsad utsträckning efterfrågar transparens i form av detaljerad teknisk insyn i algoritmernas konstruktion eller funktion. I stället handlar deras efterfrågan av transparens främst om spårbarhet, tydliga motiveringar och möjligheter att förstå vilka faktorer som ligger till grund för AI-systemets rekommendationer. En tolkning av detta är därmed att transparens fyller en funktion i form av att det möjliggör förståelse och granskning för den anställde som ansvarar för beslutet.

Utifrån tillitsteorin av Söllner et al. (2016) tydliggörs detta. När transparens är otillräcklig begränsas användarnas möjligheter gällande att bedöma AI:s förmåga, det vill säga i vilken utsträckning det kan bidra med tillförlitliga rekommendationer i den aktuella arbetsuppgiften. Samtidigt försvåras bedömningen av AI:s integritet, eftersom användarna i viss mån saknar underlag för att avgöra om rekommendationerna följer regelverk, organisatoriska riktlinjer och etiska normer. Därmed påverkas även upplevelsen av AI:s välvilja, då bristande insyn gör det svårare att avgöra om det faktiskt stödjer användarnas professionella mål och ansvar i arbetet. Transparens fungerar därmed som ett centralt bedömningsunderlag för tillit, snarare än enbart en teknisk egenskap hos AI.

Samtidigt nyanseras detta behov av transparens i empirin eftersom respondent 3 lyfter att alltför omfattande eller detaljerade förklaringar från AI kan leda till förvirring snarare än ökad förståelse. Detta sker särskilt när förklaringarna blir tekniskt komplexa, tidskrävande eller svåra att relatera till den konkreta arbetsuppgiften. Således uppfattas transparens inte som ett ideal där mer förklaringar alltid är bättre, utan som något som behöver vara funktionellt samt individ- och situationsanpassad. Förklaringar som inte stödjer den professionella bedömningen riskerar i stället att skapa kognitiv belastning och försvåra beslutsfattandet, vilket kan minska AI:s praktiska värde som beslutsstöd. Algoritmisk transparens och förklarbarhet framstår därmed som meningsfull först när den är anpassad till individernas behov, kompetens och ansvar i den aktuella situationen.

Följaktligen visar analysen att transparens inte enbart skapas genom insyn i AI:s beslutslogik, utan även genom hur de bankanställda interagerar med beslutsstödet i praktiken. Vissa respondenter beskriver hur förståelse successivt byggs genom dialog, följdfrågor och prövande användning. Denna typ av interaktion kan förstås som en form av interaktionell transparens, där tillit utvecklas genom erfarenhet snarare än genom fullständig information. Detta stämmer överens med McKnight et al. (2011) som menar att tillit till tekniska system i hög grad är erfarenhetsbaserad och formas genom upprepad användning, där systemet upplevs som funktionellt och tillförlitligt. Å andra sidan går det att konstatera att empirin visar att respondent två inte interagerar med AI, vilket enligt teori om den interaktionella transparensen bidrar till att förståelsen för systemets resonemang blir begränsad. Detta kan inte enbart förstås som ett individuellt användarbeteende, utan kan också ses i relation till organisatoriska förutsättningar.

Således tyder detta på bristande social transparens, där avsaknaden av organisatorisk vägledning och ansvarstagande gällande effektiv teknik påverkar hur de tekniska möjligheterna gällande insyn och dialog faktiskt tas i bruk. Ur ett sociotekniskt perspektiv kan transparens inte reduceras till en egenskap hos tekniken, utan förstås som ett resultat av samspelet mellan människa, teknik och organisation. Interaktionell transparens förutsätter således sociala och organisatoriska ramar, som ansvarsfördelning och organisatoriskt stöd, för att främja de anställdas användning av tekniken.

Samtidigt bekräftar detta teorin om användaren som social aktör, där användning av teknik formas av roll, ansvar och institutionella villkor snarare än enbart av individuell attityd eller teknisk tilltro (Lamb & Kling 2003). Detta tydliggörs eftersom empirin visar att respondenterna i vissa fall avslutar interaktionen med AI-verktyget när det upplevs vara ineffektivt som resultat av bristande resurser från organisationens sida. Men också eftersom samtliga respondenter är positiva till användningen av AI och trots det väljer att avsluta interaktionen och användningen på grund av organisatoriska faktorer som exempelvis ineffektiva AI-system. Således framstår de institutionella villkoren som mer avgörande för hur AI används i praktiken än de enskilda användarnas attityder till tekniken.

Å andra sidan framkom ett resonemang som kan tolkas som avvikande från Lamb och Kling (2003) teori om användaren som social aktör, vilket är det faktum att respondent 10 menar att organisationens inställning inte påverkar den individuella tilliten till AI. Respondenten menar i stället att tydligare signaler från externa aktörer, som Finansinspektionen, skulle kunna

stärka tilliten. Detta indikerar att tillit i en reglerad kontext delvis formas av institutionella ramar utanför den enskilda organisationen, där tilliten till organisationen inte är lika stor om till externa aktörer som Finansinspektionen. I linje med McKnight et al. (2011) kan detta förstås som att bedömningen av AI:s integritet i sådana sammanhang inte enbart grundas i användarnas erfarenheter av systemet, utan även formas av institutionella och regulatoriska strukturer som omger användningen av tekniken. Ur ett arbetssystemperspektiv enligt Alter (2013) belyser detta hur systemets miljö påverkar hur AI uppfattas och används i praktiken.

Social transparens, policy, riktlinjer, organisatoriska värderingar och ansvarsfördelning framstår som särskilt avgörande eftersom dessa enligt analysen begränsar, eller uppmuntrar, till användning av AI som beslutsstöd. En central tolkning utifrån respondenternas erfarenheter är att social transparens fungerar som ett institutionellt skyddsnät - om organisationen tydligt kommunicerar vad AI får användas till, hur beslut ska dokumenteras och vem som ansvarar för vad, skapas förutsättningar för tryggare och mer enhetlig användning.

Följaktligen stämmer detta överens med studiens teoretiska ramverk där Haresamudram, Larsson och Heintz (2023, s. 93) menar att forskning om transparens inom AI tidigare främst begränsats till den tekniska eller algoritmiska dimensionen, där författarna anser att transparens i stället bör förstås som ett flerdimensionellt fenomen. Resultaten bekräftar att algoritmisk transparens samspelar med interaktionell och social transparens, snarare än att fungera som en fristående förutsättning för tillit. Studien bidrar därmed genom att empiriskt stödja antagandet att transparens i praktiken formas i samspelet mellan tekniska förklaringar, de anställdas interaktion med AI och de organisatoriska ramar inom vilka AI används.

Utan tydliga organisatoriska ramar riskerar användningen av AI att bli individbaserad, fragmenterad och präglad av försiktighet, vilket kan hämma utvecklingen av erfarenhetsbaserad tillit och samtidigt skapa obalans i arbetssystemet. Detta blev tydligt under datainsamlingen, där flera respondenter framhöll att användningen av AI i hög grad förutsätter personlig kompetens och individuellt omdöme, snarare än gemensamma riktlinjer eller etablerade arbetsätt. Ur ett arbetssystemperspektiv innebär detta att transparens inte enbart kan förstås som tillgång till information, utan som ett resultat av samspelet mellan deltagare, processer, teknik och organisatorisk miljö. När policys och riktlinjer är otydliga uppstår ett glapp mellan AI-verktygens möjligheter och arbetssystemets krav på kvalitet, spårbarhet och ansvarstagande.

Vidare framkom även att respondenterna inte strävar efter att överlåta beslutsfattandet helt till AI, såvida inte organisationen säger att de anställda ska göra det, utan snarare för att effektivisera processerna. Detta nyanserar Tomkins (2001) syn på relationen mellan tillit och kontroll, där lägre tillit kräver mer information och större behov av övervakning och transparens. Som resultat av analysen framkommer dock att behovet av kontroll existerar även när AI upplevs som användbar, där respondenterna beskriver att samtliga rekommendationer från AI kontrolleras, trots att AI som beslutsstöd i vissa situationer bedöms ha en god förmåga. Detta skulle möjligen kunna förstås som en konsekvens av att banker är reglerade och att flera av respondenterna uppger att deras beslutsfattande omges av riskmedvetenhet, men även deras upplevda osäkerheter kring vilken data AI använder vid rekommendationer och beslutsstöd.

Ett annat resultat från studien är att tillit till AI framstår som tydligt situationsbunden. I arbetsuppgifter med låg upplevd risk används AI i högre grad, medan mer komplexa eller känsliga beslut kräver ännu högre grad av transparens, kontroll och mänsklig bedömning. Detta ligger i linje med tidigare forskning som visar att tillit varierar beroende på situationens osäkerhet och de konsekvenser som ett beslut kan få, snarare än att vara generell eller konstant (Tomkins 2001; McKnight et al. 2011). De bankanställda kan uppleva att AI fungerar väl och samtidigt vara försiktiga i användningen, eftersom beslutsansvaret och kraven på motivering innebär att osäkerhet måste hanteras aktivt.

Studiens resultat pekar därför på att mänskligt ansvar fungerar som ett tydligt tak för hur långt tilliten till AI-baserade beslutsstöd kan sträcka sig. Detta utgör ett centralt bidrag för studien, då samtliga respondenter tydliggör en principiell skillnad mellan AI som beslutsstöd och AI som beslutsfattare, där AI inte används som självständig beslutsfattare. Även när AI upplevs som kapabel kvarstår behovet av mänskligt omdöme och ansvar, särskilt i den reglerade bankkontexten, där beslut måste kunna motiveras och granskas i efterhand.

I detta sammanhang framstår förklarbar AI som ett medel för att möjliggöra professionellt omdöme. Samtidigt stödjer studiens resultat Rai (2020) gällande att förklarbarhet inte bör maximeras. Respondenternas erfarenheter visar att förklaringar behöver vara funktionella och anpassade till arbetsuppgiften för att vara meningsfulla. Alltför omfattande eller tekniskt detaljerade förklaringar riskerar i stället att skapa kognitiv belastning och minska effektiviteten i beslutsfattandet, vilket även lyfts i tidigare forskning om transparensens dubbla effekter av Yu och Li (2022) samt Jiang, Li och Liu (2025). Detta kan ses som en form av paradox där

respondenterna vill ha insyn för att känna kontroll, men där för mycket eller fel typ av insyn riskerar att minska användbarheten och därmed också viljan att använda AI (Yu & Li 2022; Rai 2020). Detta öppnar för tolkningen att transparens bör utformas som en situations- eller individanpassad resurs snarare än ett konstant informationsflöde.

Sammantaget visar studien att transparens är avgörande för hur tillit till AI-baserade beslutsstöd formas, men att transparens inte enbart kan reduceras till en teknisk fråga om förklarbarhet. I stället framstår transparens som något som skapas i praktiken genom samspelet mellan algoritmisk insyn, de bankanställdas möjlighet till interaktion och organisationens sätt att rama in, styra och stödja användningen av AI. Tillit till AI i bankkontexten framstår därmed som villkorad, situationsbunden och nära knuten till mänskligt ansvar och organisationen, snarare än som ett stabilt tillstånd som kan uppnås genom teknisk design ensam.

7. Slutsats

Syftet med denna studie har varit att undersöka hur bankanställda uppfattar transparens i AI-baserade beslutsstöd och hur denna upplevda transparens påverkar deras tillit till AI i det dagliga arbetet. Studien visar att bankanställdas tillit till AI-baserade beslutsstöd påverkas av transparens i den mån transparens möjliggör förståelse, granskning och ansvarstagande i det dagliga beslutsarbetet. När transparens upplevs som otillräcklig begränsas både tilliten och användningen av AI, medan en funktionell och situationsanpassad transparens skapar förutsättningar för en villkorad tillit under fortsatt mänsklig kontroll.

I linje med problemformuleringen visar detta att tillit och transparens i AI-baserade beslutsstöd inte kan förstås isolerat från den organisatoriska kontext där tekniken används. Studien bidrar därför till förståelsen av transparens som ett sociotekniskt fenomen, där transparens formas i samspelet mellan människa, teknik och organisation. Med stöd i arbetssystemteorin enligt Alter (2013) kan detta förstås som att transparens förutsätter fungerande sociala och organisatoriska ramar, som tydlig ansvarsfördelning, etablerade arbetssätt och organisatoriskt stöd.

Samtidigt går det att konstatera att transparens utgör en grundläggande förutsättning för att tillit överhuvudtaget ska kunna utvecklas, men att transparens inte enbart kan förstås som en teknisk egenskap hos AI-systemen, utan där de olika dimensionerna i stället samverkar och påverkar varandra. Algoritmisk transparens, i form av insyn i AI-verktygets beslutslogik, är centralt för att bankanställda ska uppleva tillit till AI som beslutsstöd och för att skapa förståelse för hur rekommendationer har genererats. Samtidigt utvecklas tillit även genom interaktionell transparens, där förståelsen för beslutslogiken succesivt byggs upp genom dialog med AI-beslutsstödet.

Å andra sidan framgår det att dessa interaktionsmöjligheter i praktiken är beroende av social transparens, det vill säga organisatoriska riktlinjer, utbildning och ansvarsfördelning som gör det tydligt hur AI får och bör användas. Om sådana ramar saknas, leder det i stället till försiktig och selektiv användning av AI som beslutsstöd. Detta visar att de olika dimensionerna av transparens samverkar och påverkar varandra.

Ett annat centralt resultat är att bankanställda inte efterfrågar transparens i form av detaljerad teknisk insyn i algoritmernas konstruktion, utan snarare i form av spårbarhet och möjlighet att

förstå vilka grunder AI genererar rekommendationer och beslutsstöd på. Transparens fungerar därmed främst som ett praktiskt och professionellt stöd för att kunna granska, motivera och ta ansvar för beslut, snarare än som ett ideal om fullständig förklarbarhet. När denna form av transparens saknas begränsas både tilliten till AI och dess faktiska användning som beslutsstöd eftersom de anställda har svårt att förstå vilka grunder AI baserar rekommendationerna på.

Ett ytterligare bidrag är att tillit till AI framstår som tydligt situationsbunden. Studien visar att AI främst används i enklare och mer standardiserade arbetsuppgifter med låg upplevd risk, medan komplexa beslut fortsatt kräver mänsklig kontroll, transparens och professionellt omdöme. Denna försiktighet bör inte nödvändigtvis förstås som motstånd mot tekniken eftersom respondenterna i hög grad uttrycker en positivitet gällande användningen av AI, utan snarare som ett uttryck för professionell riskhantering i en reglerad verksamhet där beslut måste kunna motiveras och granskas i efterhand.

Följaktligen går det att konstatera att mänskligt ansvar utgör en tydlig gräns för hur långt tilliten till AI-baserade beslutsstöd kan sträcka sig. Även när AI upplevs som tekniskt kapabel förblir tilliten villkorad, eftersom det slutliga ansvaret inte kan delegeras till tekniken. Förklarbar AI framstår därför som ett medel för att möjliggöra professionellt omdöme snarare än som ett mål i sig. Samtidigt indikerar resultaten att förklarbarhet behöver vara individbaserad och kontextanpassad för att stödja beslutsfattande utan att skapa kognitiv belastning eller minska effektiviteten för de anställda.

Avslutningsvis visar studien att tillit till AI-baserade beslutsstöd utvecklas i ett samspel mellan AI:s egenskaper, som autonomi, lärande och upplevd oförklarbarhet, samt transparensens olika dimensioner det vill säga algoritmisk, interaktionell och social. Dessa faktorer påverkar hur bankanställda i praktiken kan bedöma AI-beslutsstödet förmåga, integritet och välvilja, vilket enligt tillitsteorin utgör centrala förutsättningar för att tillit ska kunna uppstå och som därefter påverkar hur de anställda använder verktyget. Studien visar att teknisk förklarbarhet inte är tillräckligt för att skapa tillit till AI, utan att tilliten formas som resultat av hur transparens realiseras i relation till arbetssystemets organisatoriska och professionella villkor och reglering.

8. Framtida forskning

Framtida forskning kan fördjupa förståelsen för hur transparens och tillit till AI-baserade beslutsstöd formas inom bankverksamhet genom att undersöka variationer mellan olika yrkesroller och arbetsuppgifter. En jämförande analys av exempelvis rådgivare, riskfunktioner, regel efterlevnadsfunktioner och ledning kan synliggöra hur behovet av transparens och graden av tillit till AI skiljer sig åt inom arbetssystemet, samt hur transparenskrav bör anpassas utifrån skilda professionella ansvar, beslutsmiljöer och risknivåer.

En ytterligare forskningslinje rör den paradoxala roll som transparens kan spela i relation till tillit till AI-baserade beslutsstöd, där ökad insyn inte nödvändigtvis leder till ökad tillit. Mot denna bakgrund går det därför att genomföra experimentella studier där både omfattningen och typen av AI:s förklaringar varieras, exempelvis genom att jämföra korta samt beslutsnära motiveringar med mer omfattande teknisk insyn. Genom att mäta effekter på tillit, kognitiv belastning och besluts kvalitet kan sådan forskning bidra till att identifiera vilka former av förklaringar som främjar tillit utan att leda till informationsöverbelastning eller försämrad effektivitet i beslutsarbetet.

Vidare går det även att genom longitudinella studier följa bankanställda över tid för att undersöka hur den eventuella erfarenhetsbaserade tilliten utvecklas i takt med att AI integreras i arbetsprocesser. Sådan forskning kan synliggöra hur tillit påverkas av införandeprocessen, organisatoriska förändringar och successiva erfarenheter av verktygets träffsäkerhet och begränsningar. Detta är särskilt relevant i organisationer där tillit inte enbart är en individuell fråga, utan också formas av rutiner, ansvar och styrning i arbetssystemet.

Slutligen kan framtida forskning undersöka hur mänskligt ansvar påverkar graden av tillit och automatisering, det vill säga varför AI som denna studie visat, förblir ett beslutsstöd snarare än att bli ett fullt beslutsunderlag. Med utgångspunkt i Lamb och Kling (2003) och deras sociotekniska aktörsperspektiv kan detta analyseras genom hur professionella roller, normer och institutionella miljöer formar ansvar och legitimitet i användningen av AI. Samtidigt kan Alter (2013) och arbetssystemteorin användas för att identifiera sociotekniska mekanismer som begränsar automatisering, som ansvarsfördelning, riskuppfattning och krav på spårbarhet. Sammantaget kan denna ansats belysa hur selektiv AI-användning påverkar organisationens nytta och riskhantering.

Källförteckning

- Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J. & Díaz-Rodríguez, N. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*, 99, 101805. doi: 10.1016/j.inffus.2023.101805
- Alter, S. (2006). *The work system method: Connecting people, processes, and IT for business results*. Work System Press.
- Alter, S. (2013). Work System Theory: Overview of Core Concepts, Extensions, and Challenges for the Future. *Journal of the Association for Information Systems*, 14(2), s. 72-121. doi:10.17705/1jais.00323
- Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing Artificial Intelligence. *MIS Quarterly*, 45(3), s. 1433-1450. doi: 10.25300/MISQ/2021/16274
- Bryman, A., Bell, E. & Harley, B. (2024). *Företagsekonomiska forskningsmetoder*. 4:e uppl. Stockholm: Liber.
- Crawford, J. (2017). Regulation's influence on risk management and management control systems in banks. *Företagsekonomiska institutionen Department of Business Studies*, Uppsala universitet. urn:nbn:se:uu:diva-332037
- Dwivedi, R., Dave, D., Naik, H., Singhal, S., Rana, O., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G. & Ranjan, R. (2023). Explainable AI (XAI): Core Ideas, Techniques and Solutions. *ACM Computing Surveys*, 55(9), s. 1-33. doi: 10.1145/3561048
- European Commission. (2025). *AI Act*. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> [Hämtad: 17-11-2025]
- European Union. (2024). *Regulation (EU) 2024/1689 - Artificial Intelligence Act (AI Act)*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689> [Hämtad: 24-11-2025]
- Finansinspektionen (2024a). *AI allt vanligare i finanssektorn men riskhanteringen släpar efter*. <https://www.fi.se/sv/publicerat/rapporter/rapporter/2024/ai-allt-vanligare-i-finanssektorn-men-riskhanteringen-slapar-efter/> [Hämtad: 07-11-2025]
- Finansinspektionen (2024b). *AI-användningen i den svenska finanssektorn*. <https://www.fi.se/contentassets/084ebc13d6364a28a87a37c9a557ec9c/rapport-ai-svenska-finanssektorn.pdf> [Hämtad: 07-11-2025]

Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B. & Spaulding, R. (2021) Attachment and trust in artificial intelligence. *Computers in Human Behavior*, 115, 106607. doi: 10.1016/j.chb.2020.106607

Haresamudram, K., Larsson, S. & Heintz, F. (2023). Three Levels of AI Transparency. *Computer*, 56(2), s. 93-100. doi: 10.1109/MC.2022.3213181

Hennink, M., Hutter, I. & Bailey A. (2020). *Qualitative Research Methods*. Utgåva 2, Sage.

Lazo, M. & Ebarido, R. (2023). Artificial Intelligence Adoption in the Banking Industry: Current State and Future Prospects. *Journal of Innovation Management*, 11(3), s. 54-74. doi:10.24840/2183-0606_011.003_0003

Islam, M., Mehadi, G., Sohag, H.J. & Uddin, K. (2025). The Role and Impact of-Management Information Systems in the Banking Sector: Enhancing Operational Efficiency & Customer Satisfaction. *International Journal of Economics and Management*, 10, s. 274-290. doi: 10.2139/ssrn.5471566

Jiang, Z., Li, Z. & Liu, C. (2025). Understanding dimensions of trust in AI through quantitative cognition: Implications for human-AI collaboration. *PLOS ONE*, 20(7), e0326558. doi: 10.1371/journal.pone.0326558

Kovari, A. (2024). AI for Decision Support: Balancing Accuracy, Transparency, and Trust Across Sectors. *Information*, 15, Artikel 725. doi: 10.3390/info15110725

McKnight, D.H., Carter, M., Thatcher, J.B. & Clay, P.F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, 2(2), Artikel 12, s. 1-25. doi: 10.1145/1985347.1985353

Maier, J.D., Menold, J. & McComb, C. (2022). The relationship between performance and trust in AI in e-finance. *Frontiers in Artificial Intelligence*, 5, 891529. doi: 10.3389/frai.2022.891529

Polinati, A.K., Singh, S., Akula, S., Pasala, R.R., Sharma, M., Korkanti, S.K. & Bose, B. (2025). Revolutionizing Information Management: AI-Driven Decision Support Systems for Dynamic Business Environments. *Journal of Information Systems Engineering and Management*, 10(35s), s. 322-335. doi: 10.52783/jisem.v10i35s.6010

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, 48, s. 137-141. doi: 10.1007/s11747-019-00710-5

Rienecker, L. & Stray Jørgensen, P. (2025). *Att skriva en bra uppsats*. Utgåva 5, Liber.

Stryker, C. & Kavlakoglu, E. (u.å.). *What is artificial intelligence (AI)? IBM Think*. <https://www.ibm.com/think/topics/artificial-intelligence> [Hämtad: 07-11-2025]

Söllner, M., Benbasat, I., Gefen, D., Leimeister, J. M. & Pavlou, P. A. (2016). Trust: MIS Quarterly Research Curation. *MIS Quarterly*, June 2016. <https://ssrn.com/abstract=2985418> [Hämtad: 07-11-2025]

Tomkins, C. (2001) Interdependencies, trust and information in relationships, alliances and networks, *Accounting, Organizations and Society*, 26(2), s. 161-191. Doi: 10.1016/S0361-3682(00)00018-0

Valacich, J. S., Schneider, C. & Hashim, N. H. (2022). *Information Systems Today: Managing the Digital World*. 9th ed. Pearson Education.

Wilson, A.D., Onwuegbuzie, A.J. & Manning, L.P. (2016) Using Paired Depth Interviews to Collect Qualitative Data. *The Qualitative Report*, 21(9), s. 1549-1573. doi: 10.46743/2160-3715/2016.2166

Yu, L. & Li, Y. (2022). Artificial Intelligence Decision-Making Transparency and Employees' Trust: The Parallel Multiple Mediating Effect of Effectiveness and Discomfort. *Behavioral Sciences*, 12(4), s. 127. doi: 10.3390/bs12050127

Bilagor

Bilaga 1: Intervjuguide

De intervjuer vi genomför ingår som en central del av vår kandidatuppsats i ämnet informatik. Syftet med studien är att undersöka hur transparens i AI-baserade beslutsstöd påverkar bankanställdas upplevelse av tillit i sitt beslutsfattande.

Intervjun kommer att spelas in och inspelningen används endast för att kunna skriva ut och analysera materialet. Transkriberingen delas inte offentligt i den färdiga studien, men kommer att delas med vår handledare och examinator.

Alla som deltar kommer att vara helt anonyma i uppsatsen och inga personliga uppgifter kommer att kunna kopplas till dig. Du kan också avbryta ditt deltagande när som helst utan att ange någon anledning.

Inledning och rollförståelse (socioteknisk systemteori/arbetsystemteori)

1. Kan du berätta om din roll och vilka typer av beslut du fattar i ditt arbete?

- Hur ofta fattar du sådana beslut?
- Vad är viktigt för dig i beslutsprocessen?

2. Vilken typ av AI använder ni?

3. På vilket sätt ingår AI i din arbetsprocess i nuläget?

- Beskriv hur AI används, eller planerar att användas.

4. Hur har AI-baserat beslutsstöd förändrat arbetsprocesser, roller eller ansvar i er organisation jämfört med hur det såg ut innan?

- Kan du ge exempel på hur det märks i praktiken?

AI:s egenskaper och upplevda utmaningar (AI:s framkant)

5. Hur mycket frihet upplever du att AI har när den ger rekommendationer eller beslutsunderlag?

6. Om systemet förändrar sina rekommendationer över tid baserat på exempelvis historik, hur uppfattar du den utvecklingen?

- Upplever du att systemet blir bättre eller sämre?
- Hur märker du det?

7. Vilka situationer har fått dig att uppleva AI:s beslutslogik som svår att förstå?

- Hur hanterar du sådana situationer?
- Hur påverkar det din vilja att använda systemet?

Transparens

Algoritmisk transparens

8. Hur skulle du beskriva din förståelse av hur AI behandlar information och kommer fram till sina förslag?

- Vilken information saknar du?
- Hur påverkar det dig?

9. I vilka situationer upplever du att insynen i systemets beslutsprocess är tillräcklig, och i vilka situationer uppstår brister?

Interaktionell transparens

10. Hur upplever du samspelet med systemet när du använder det i dina beslut?

11. På vilket sätt hjälper systemet dig att förstå dess förslag under själva användningen?

- Hur kan du följa logiken bakom förslagen?

Social transparens

12. Hur tydligt är det vilka regler, rutiner eller ansvar som gäller kring AI i din organisation?

- Vet du vem som ansvarar för fel?
- Hur påverkar det din trygghet?

13. Hur hanteras frågor om risker, begränsningar eller osäkerheter kring AI i din organisation?

- Vad saknas?

Förklarbar AI (XAI)

14. Hur presenterar systemet sina förklaringar när det ger ett beslutsunderlag?

- Hur användbara är de i praktiken?

15. Hur skulle mer förklarbarhet påverka din användning av AI?

- Vilken typ av förklaring skulle du vilja ha?

16. Om du har saknat en förklaring från AI i en specifik situation, hur påverkade det ditt fortsatta beslut?

- Hur påverkade den situationen ditt senare arbete med AI?

Tillit till AI

Förmåga

17. Vilka erfarenheter har format din uppfattning om AI-systemets kompetens i dess beslutsstöd?

- Vad gör att du litar/litar inte på dess kompetens?

18. När du väger din egen expertis mot AI:s rekommendationer, hur brukar du resonera?

- Vad är det som gör att du agerar på detta sätt?

Välvilja

19. Hur uppfattar du samspelet mellan dina mål i arbetet och vad systemet verkar vilja stödja?

Pålitlighet/etik

20. Hur gör du för att avgöra om AI:s rekommendationer känns etiskt hållbara, rättvisa eller välgrundade?

- Hur bedömer du det i praktiken?

21. Vilka händelser eller observationer har påverkat din uppfattning om systemets pålitlighet?

Transparens & tillit

22. Vilka typer av information eller insyn från AI ger dig en känsla av förtroende i ditt arbete?

- Kan du beskriva en situation där transparens ökade din tillit?
- Och en där lägre transparens minskade den?

23. Hur navigerar du i situationer där AI:s bedömning skiljer sig från din egen?

- Exempel på en sådan situation?

Konsekvenser för beslutsfattande

24. Hur brukar du avgöra i vilken utsträckning du ska använda AI som stöd i ett beslut?

25. Hur skulle du beskriva hur AI-systemet uppfattas i organisationen, både formellt och informellt?

26. Hur har relationen mellan ditt arbete och tekniken förändrats sedan AI infördes?

- Hur påverkar det ditt sätt att arbeta?

Risk, ansvar och organisatoriska faktorer (Stöds av sociotekniska teorin)

27. Hur hanterar du situationer där AI:s rekommendationer kan innebära risker för kund eller organisation?

28. Vilken typ av stöd, utbildning eller resurser skulle hjälpa dig att förstå eller värdera AI-systemet?

- Vad saknas?

Avslutande reflektioner

29. Vilken aspekt av AI-systemet skulle göra störst skillnad för dig och eventuellt öka tillit om den förbättrades?

30. Finns det något du vill lägga till, nyansera eller korrigera innan vi avslutar?

Bilaga 2: Deduktiva koder

Deduktiva koder:	Teori	Förkortning
Autonomi: Hur AI uppfattas agera självständigt, "Tar egna beslut " och ändrar rekommendationer.	AI:s framkant (Berente et al. 2021)	#A
Lärande: Utsagor om att systemet "lärsig", blir bättre över tid, anpassar sig till nyadata.	AI:s framkant (Berente et al. 2021)	#L
Oförklarbarhet: Beskrivningar av att AI är svårt att förstå, svårt att se logiken bakom besluten.	AI:s framkant (Berente et al. 2021)	#O
Algoritmisk Transparens: Hur de uppfattar insyn i hur modellen fungerar, vilka faktorer som vägs in, varför ett visst beslut ges.	Olika dimensioner av transparens (Haresamudram, Larsson & Heintz 2023)	#AT
Interaktionell Transparens: Hur transparens uppstår i mötet mellan de bankanställda och system.	Olika dimensioner av transparens (Haresamudram, Larsson & Heintz 2023)	#IT
Social Transparens: Hur banken informerar om AI, policys, riktlinjer, vem som ansvarar. hur besluten kan ifrågasättas.	Olika dimensioner av transparens (Haresamudram, Larsson & Heintz 2023)	#ST
Förklarbarhet XAI: Uttalanden om att systemet "förklarar" eller "motiverar" eller gör beslut mer begripliga/spårbara	Förklarbar AI (Dwivedi et al. 2023)	#XAI

(Uppfattningar om bra/dåliga förklaringar).		
Förmåga: Upplevelser av AI:s kompetens.	Tillitsteori (Söllner et al. 2016)	#F
Välvilja: Om systemet upplevs vilja "hjälpa" respondenterna och stödja jobbet.	Tillitsteori (Söllner et al. 2016)	#V
Integritet: Etik, bias, om AI följer regler och normer.	Tillitsteori (Söllner et al. 2016)	#I
Erfarenhetsbaserad tillit: Hur tilliten har ändrats genom faktisk erfarenhet.	Utvecklad tillit till AI (McKnight et al. 2011)	#ET
Arbetsroll och ansvar: Hur deltagarna beskriver sin roll, ansvar för beslut, vem som äger beslutet när AI är inblandat.	Arbetsystemteorin (Alter 2013)	#AA
Samspel (Människa och AI): Hur AI används i praktiken: stöd, dubbelkontroll, ersätter vissa moment, konflikter mellan eget omdöme och AI:s förslag.	Socioteknisk systemteori (Alter 2013; Lamb & Kling 2003)	#S
Organisatoriska villkor: Kultur, ledningssyn, utbildning, resurser och styrning kring AI-användning.	Arbetsystemteorin (Alter 2013)	#OV

Bilaga 3: Induktiva koder

Induktiva koder	Förkortning
Personlig kompetens: Avser anställdas yrkesmässiga kompetens, inklusive utbildning, erfarenhet, och professionellt omdöme.	#PK
Effektivisering av arbete: Avser upplevelser av hur AI bidrar till snabbare, smidigare eller mer strukturerade arbetsprocesser.	#EA
Riskmedvetenhet: Avser den medvetenhet som anställda uttrycker kring potentiella risker med AI-användning.	#RM
AI-dominans: Avser föreställningar och uppfattningar om att AI i allt högre grad dominerar eller förväntas komma att dominera beslutsfattande och arbetsprocesser.	#AD
AI-integrationsansvar: Avser hur ansvar för AI:s integration uppfattas.	#IA

Bilaga 4: Kategorier och teman

Teman	Kategori	Deduktiva koder	Induktiva koder
Transparensens betydelse för tillit	Transparens genom interaktion för att förstå och värdera AI:s rekommendationer.	#AT, #IT, #ST, #XAI, #O	#RM
	Organisatoriska strukturer och begränsningar som formar tillit och transparens.	#OV, #AA, #ST, #S	#IA, #EA
	Förklarbar AI som praktisk strategi, inte tekniskt krav.	#IT, #XAI, #AT	#RM
Tillit som situationsbunden, erfarenhetsbaserad och begränsad	Tillit till AI som växer genom erfarenhet och upplevd förmåga.	#F, #V, #I, #ET	#PK
	Tillit som situationsbunden och begränsad.	#ET, #F, #S	#RM
Mänskligt ansvar och professionellt omdöme i AI-baserade beslutsstöd	AI som stöd, men inte ersättning för mänskligt beslutsfattande.	#AA, #S, #F, #I, #A	#RM, #PK
	Individuellt ansvar gällande användning av AI.	#AA, #I, #S, #A	#RM, #PK

Bilaga 5: Missivbrev



Informationsbrev till deltagare

Titel på studien

Bankanställdas tillit till artificiell intelligens som beslutsstöd: Ett organisationsperspektiv

Forskargrupp

Masah Alkurdi och Saga Adolphsson, studenter vid Södertörns högskola, kursen *Informatik C - inriktning digital affärsutveckling* (HT25).

Handledare

Pär-Ola Zander, Södertörns högskola

Syfte med studien

Syftet med denna studie är att undersöka hur transparens i AI-baserade beslutsstöd påverkar bankanställdas upplevelse av tillit i sitt beslutsfattande. Genom att analysera hur transparens upplevs och tolkas i mötet mellan människa och system syftar studien till att belysa hur graden av insyn i AI:s beslutsstöd processer påverkar tillit, trygghet och ansvar i beslutsfattande.

Om intervjun

Intervjun beräknas ta cirka **60 minuter** och kommer att genomföras på plats, platsen är beroende på vad som passar bäst för deltagaren. Med deltagarens tillstånd kommer intervjun att spelas in för att underlätta transkribering (skriva ut intervjun i textform). Inspelningarna och utskrifterna kommer endast att användas av oss som genomför studien.

Vid intervjutillfället kan även **ett fotografi tas av intervjumiljön**. Bilden kan komma att visa deltagaren, men endast om denne samtycker till det. Syftet med bilden är att illustrera intervjusituationen i uppsatsen. Bilden kommer **inte att publiceras eller spridas utanför uppsatsen** och används endast i forskningssyfte. Deltagaren kan välja att avstå från att vara med på bilden utan att det påverkar deltagandet i studien.

Frivilligt deltagande och konfidentialitet

Deltagandet är **frivilligt**, och du kan när som helst avbryta intervjun eller dra tillbaka ditt deltagande utan att ange något skäl. All information som framkommer kommer att behandlas konfidentiellt och anonymiseras i den färdiga uppsatsen. Det betyder att inga namn, roller eller uppgifter som kan kopplas till en specifik person kommer att publiceras.

Materialet kommer endast att användas för denna kandidatuppsats och raderas efter att studien har examinerats.

Kontaktinformation

Om du har frågor eller vill veta mer om studien är du välkommen att kontakta oss:

Masah Alkurdi - “*mailadress*”

Saga Adolphsson - “*mailadress*”

Handledare: “*mailadress*”

Samtyckesformulär

Jag har tagit del av informationen ovan och samtycker till att delta i studien “*Bankanställdas tillit till artificiell intelligens som beslutsstöd: Ett organisationsperspektiv*”.

Jag är medveten om att mitt deltagande är frivilligt och att jag när som helst kan avbryta utan att behöva ange något skäl. Jag samtycker till att intervjun får spelas in och att mitt svar får användas i anonymiserad form i uppsatsen.

Namn:
Organisation:

Datum:

Underskrift: