DIGITALA SYSTEM
SYSTEMS ENGINEERING

Transparenta algoritmer i försäkringsbranschen

Ulrik Franke

RISE Rapport 2021:04

Transparenta algoritmer i försäkringsbranschen

Ulrik Franke

Abstract

Transparent algorithms in insurance

This is the final report of the project Transparent algorithms in insurance, which has been conducted at RISE and KTH Royal Institute of Technology, funded by Länsförsäkringars forskningsfond, from 2018 to 2021. The report starts by discussing some of the difficulties that arise with automated decision-making and black box-like algorithms, as well as the prospects for alleviating these through appropriate transparency. Next, some research results are presented. From an analysis of national AI strategies from the Nordic countries, it is concluded that ethics play an important part in them: ethical AI is said to be a competitive advantage. However, the documents do not offer any convincing arguments for this position, and do not offer any clear ethical guidance. Based on two case studies, the prospects for using the Value Sensitive Design method to conduct design work while taking transparency and other values into account are discussed. The method is deemed promising for use in actual design work, e.g., within insurance. An empirical investigation of how the right to explanation under GDPR works in practice in Swedish insurance finds large discrepancies between companies. The times required to respond were long and the contents of the answers showed considerable variability. A follow-up study in several EU countries is planned and will in due time give a better picture of what the GDPR right to meaningful information actually means in practice. Based on interviews with some 30 senior managers and board members within banking and insurance, possibilities and challenges of implementing transparent and explainable AI in large organizations are discussed. Difficulties include coming to grips with different points of view and creating common frames of reference. The perspective of the Swedish insurance industry on AI and transparency was also investigated through interviews. There is a widespread belief that transparency could be a competitive advantage, but most informants are uncertain about how this advantage can actually be realized. It is also apparent that AI is not yet used in the insurance core business in Sweden, meaning that the questions of transparent and explainable AI are still somewhat theoretical to the industry. Still, they are expected to become important in due time. The report is concluded with some observations that are practically relevant to insurance, and a few directions for future research.

Keywords: Insurance, transparency, Value Sensitive Design (VSD), GDPR, national AI strategies, AI governance

RISE Research Institutes of Sweden AB

RISE Rapport 2021:04
ISBN:978-91-89167-87-2
Stockholm 2021

Innehåll

Förord

Projektet Transparenta algoritmer i försäkringsbranschen (TALFÖR) har med finansiering från Länsförsäkringsgruppens forsknings- och utvecklingsfond (P4/18) under perioden oktober 2018-januari 2021 studerat olika aspekter av transparens, AI och etik med försäkringsbranschen som tematiskt fokus. Projektet har syftat till att ta fram praktiskt relevant kunskap om hur vi kan dra nytta av fördelarna med automatiserat beslutsfattande utan att drabbas av dess nackdelar.

Projektets resultat redovisas kortfattat i denna populärvetenskapliga sammanfattning. I ett appendix finns en sammanställning av projektets vetenskapliga publicering. Ytterligare några artiklar genomgår i skrivande stund vetenskaplig granskning (*peer review*) och kommer att publiceras i sinom tid.

Projektet hade inte varit möjligt utan de informanter som ställt sin tid och kunskap till förfogande i delstudierna. Ett stort tack riktas till dessa, liksom till forskningsfondens kansli och till medlemmarna av referensgruppen som har bidragit med uppmuntran, idéer och goda råd.

Ulrik Franke, projektledare

Sammanfattning

Denna rapport utgör slutrapport för projektet Transparenta algoritmer i försäkringsbranschen (TALFÖR) som har genomförts vid RISE och KTH med finansiering från Länsförsäkringars forskningsfond från 2018 till 2021. Inledningsvis diskuterar rapporten några av de svårigheter som kan uppstå med automatiserat beslutsfattande och algoritmer som liknar svarta lådor, liksom möjligheterna att genom lämplig transparens komma till rätta med dessa. Därefter presenteras några forskningsresultat. Utifrån en analys av de nordiska ländernas nationella AI-strategier konstateras att etik spelar en framträdande roll i dessa: etisk AI sägs vara en konkurrensfördel. Däremot lägger dokumenten inte fram några övertygande argument för denna ståndpunkt och ger inte heller någon tydlig vägledning i etiska frågor. Baserat på två genomförda försök diskuteras utsikterna att använda metoden Value Sensitive Design för design med hänsyn till transparens och andra värden. Metoden bedöms ha god potential för användning i konkret designarbete, exempelvis i försäkringsbranschen. En empirisk undersökning av hur GDPR-rätten till förklaring fungerar hos svenska försäkringsbolag finner stora skillnader bolagen emellan. Svarstiderna var långa och själva innehållet i svaren skiljer sig betänkligt åt. En uppföljande studie i fler EU-länder planeras och kommer på sikt att ge en bättre bild av vad GDPR-rätten till meningsfull information faktiskt innebär i praktiken. Utifrån intervjuer med ett trettiotal högre chefer och styrelseledamöter inom bank och försäkring diskuteras möjligheter och utmaningar med att införa transparent och förklarbar AI i större organisationer. Svårigheterna handlar bland annat om att hantera olika utgångspunkter och lyckas skapa gemensamma referensramar. Den svenska försäkringsbranschens syn på AI och transparens undersöktes också genom intervjuer. Det finns en spridd uppfattning om att transparens kan vara en konkurrensfördel, men de flesta informanter är osäkra på hur denna faktiskt kan realiseras. Det är också uppenbart att AI ännu inte används i försäkringsbolagens kärnverksamhet i Sverige, vilket innebär att frågorna om transparens och förklarbarhet än så länge är lite teoretiska för branschen. Samtidigt väntas de bli viktiga på sikt. Rapporten avslutas med några iakttagelser av praktisk relevans för försäkringsbranschen och några uppslag till framtida forskning.

Nyckelord: Försäkring, transparens, Value Sensitive Design, GDPR, nationella AI-strategier, ledning och styrning av AI

1 Inledning

I vårt dagliga liv stöter vi på automatiserat beslutsfattande hela tiden. Vid varje kortköp görs en automatisk bedömning av sannolikheten att det rör sig om ett bedrägeriförsök. Den som ringer till eller chattar med en kundtjänst möter oftast först ett automatiskt system som ibland kan lösa ärendet på egen hand, ibland slussar en vidare till en människa. Alla som streamar en TV-serie har vant sig vid att få tips på ytterligare, liknande, serier. Det verkar också högst sannolikt att utvecklingen fortsätter med mer automatisering driven av AI-framsteg på allt fler områden, även om en del förutsägelser på senare år tycks har varit överdrivna [37, 18].

En sådan utveckling kan få många goda konsekvenser. Mänskligt beslutsfattande har många kända svagheter som ankring, bekräftelsebias, social anpassning etc. [49]. Ur det perspektivet är det ingen långsökt tanke att datorer i vissa sammanhang skulle kunna göra ett bättre jobb än människor, eller åtminstone komplettera människor på ett positivt sätt: Datorer blir exempelvis inte trötta eller irriterade och drivs inte heller av prestige eller självhävdelsebehov. Många länder, inklusive Sverige, tar därför fram nationella AI-strategier för att navigera rätt bland de möjligheter som den nya tekniken skapar (se [22] för en studie av nordiska AI-strategier och [39] för en global översikt).

Samtidigt medför mer automatiserat beslutsfattande nya risker. Maskininlärda system har visat sig kunna fatta beslut som diskriminerar vissa grupper, exempelvis minoriteter eller människor med knappa ekonomiska resurser [41], utan att detta egentligen var avsikten från början. Sådana oförutsedda konsekvenser har lett till omfattande diskussioner om algoritmer som liknar ”svarta lådor” [42, 16]. Hur denna sorts system kan göras mer förståeliga och transparenta är följaktligen också ett högaktuellt forskningsområde [34, 27].

Både potentialen och problemen med ökad AI-användning går att illustrera med svenska exempel. I januari 2021 gick Lunds kommun ut och berättade att deras automatiska system hade hanterat 66 840 ärenden och därmed sparat in 7 150 mantimmar – knappa resurser som därmed kunde användas på bättre sätt inom kommunen.[1] Men utvecklingen ifrågasätts också. I januari 2020 anmälde Akademikerförbundet SSR Trelleborgs kommun till Justitieombudsmannen för att kommunen låter en algoritm snarare än en socialsekreterare fatta beslut om försörjningsstöd.[2] Vilken väg utvecklingen ska ta är omtvistat.

1.1 Transparens och AI

Traditionella datorprogram är regelbaserade. Reglerna – algoritmerna – är som kokboksrecept och även om de kan vara komplicerade så kan den som har rätt utbildning läsa koden och förstå vad som sker. En mänsklig programmerare som skriver ett regelbaserat program för att känna igen bilder av katter skriver troligen kod som går att begripa: kanske letar den efter morrhår, fyra ben och svans? På senare tid har istället maskininlärning, baserad på statistiska samband, möjliggjort slående framsteg på områden som bildigenkänning [35] och språkteknologi [36]. Här går det inte att följa

[1] https://computersweden.idg.se/2.2683/1.745387/lund-kommunala-bottar, läst 2021-01-13
[2] https://www.svd.se/obehorig-algoritm-tar-beslut-i-socialtjansten, läst 2020-03-11

receptet på samma sätt. Principen är istället att systemen utvecklas kontinuerligt genom att matas med nya indata. För att fortsätta exemplet så skulle ett sådant system exponeras för miljoner kattbilder och därigenom bygga upp en modell som å ena sidan fungerar oerhört bra, men å andra sidan är helt oöverskådlig för den som försöker förstå den. Sådan maskininlärning är bara en delmängd av det större området AI, men det är en delmängd som just nu tilldrar sig stor uppmärksamhet och troligen också det område där frågorna om transparens och förklarbarhet blir mest akuta.

I den datavetenskapliga litteraturen finns det gott om exempel på olika tekniska lösningar för att ta fram förklaringar till varför en algoritm beter sig på ett visst sätt [34, 27]. Bildigenkänningssystem kan märka upp vilka delar av en bild som påverkar klassificeringen, system som arbetar med text kan på liknande vis markera hur olika ord har tolkats och mer abstrakta system för exempelvis kreditprövningar kan översättas till tumregler som är enklare att tolka för en människa.

Trots detta är det tydligt att dessa lösningar också har viktiga begränsningar. Markeringar i bilder och text kan vara mångtydiga, missvisande eller överförenklade. Förenklade tumregler har per definition ett begränsat förklaringsvärde, eftersom de bara beskriver en förenkling av det verkliga systemet. Ett visst mått av oöverskådlighet återstår alltså – vi kan inte förvänta oss att bygga bort den genom rent tekniska lösningar.

System som är svåra eller omöjliga att förstå kan naturligtvis vara problematiska. Som nämnt ovan kan oöverskådliga algoritmer, utan avsikt, efterlikna diskriminerande beteenden som fanns i den data som de har tränats upp på [41]. Ett företag som köper in ett automatiserat rekryteringssystem riskerar att överraskas av att det exempelvis missgynnar kvinnor, helt enkelt för att det utgår från en datamängd där kvinnor var missgynnade. Det finns gott om exempel på hur detta sker i praktiken, på alla möjliga områden. Maskinerna kan föra vidare våra mänskliga fördomar och skevheter; inte bara när människor har märkt upp träningsexempel för så kallad *övervakad* inlärning, utan till och med vid *oövervakad* inlärning, om denna bygger på data som i någon mening är skev [14]. I Sverige initierade Jämställdhetsmyndigheten i december 2020 ett pilotprojekt för att upptäcka och mäta möjliga risker med AI avseende jämställdhet, utifrån just denna sorts överväganden.[3]

När ett AI-system byggs för att vara transparent finns det alltid en risk att transparensen uppstår på bekostnad av sämre prestationsförmåga. Då uppstår svåra avvägningar. Därför är det mycket lättare att peka ut problemen än att lösa dem – åtminstone om lösningarna inte ska kasta bort fördelarna med AI av bara farten.

1.2 Transparens och tillit

En viktig pusselbit i frågorna om transparens handlar om människors attityder och tillit. I den statsvetenskapliga litteraturen finns det starka belägg för att transparens i den offentliga sektorn minskar korruption och leder till bättre hushållning med skattemedel, men däremot inte nödvändigtvis till ökad tillit [19]. I Sverige undersöktes inom ramen för den nationella SOM-undersökningen 2018 medborgarnas syn på automatiserat beslutsfattande i offentlig sektor — när algoritmer ersätter, eller väljer ut beslutsunderlag

[3] https://www.jamstalldhetsmyndigheten.se/nyhet/unikt-pilotprojekt-om-ai-utveckling-och-jamstalldhet-pa-statliga-myndigheter, läst 2021-01-22.

åt, handläggare som beslutsfattare [21]. Bara en minoritet, 20%, kände ens till att detta sker. Även om de flesta visserligen tror att automatiserade beslut blir mer opartiska så tvivlar de flesta på att de blir mer tillförlitliga. En majoritet tror också att automatiska beslut tar mindre hänsyn till människornas situation och ger mindre insyn i beslutsfattandet. Det finns alltså all anledning att ta frågorna på största allvar för att säkerställa att medborgarna fortsatt har anledning att känna tillit till myndigheterna.

För att den offentliga sektorn ska kunna dra nytta av fördelarna med automatisering utan att drabbas av dess nackdelar fick Lantmäteriet och DIGG i december 2019 ett regeringsuppdrag om att testa ny teknik vid automatisering inom offentlig förvaltning.[4] I den slutrapport som presenterades ett år senare föreslår myndigheterna en förtroendemodell som syftar till att skapa tillit till myndighetsbeslut – även när dessa går emot ens förhoppningar.[5]

Förtroendefrågorna är naturligtvis inte begränsade till offentlig sektor. Transparens har visat sig öka tilliten även gentemot företag [38], och ur det perspektivet skulle transparens kunna vara en konkurrensfördel. För att undersöka attityder till transparens deltog TALFÖR i utformandet av undersökningen *Delade meningar 2019* med något tusental respondenter kring årsskiftet 2018/2019.[6] Respondenterna fick svara på ett tjugotal frågor relaterade till integritet och digitalisering, varav två handlade specifikt om transparens i digitala tjänster.

På frågan ”Skulle du kunna tänka dig att byta från en digital tjänst till en annan om den nya tjänsten är mer öppen och transparent med hur de använder din personliga information?” svarade hela 63% ja. Ingen större överraskning, eftersom transparens är ett positivt laddat begrepp.

För att borra djupare ställdes en följdfråga till de som svarat positivt: ”Skulle du kunna tänka dig att byta även om den nya tjänsten är dyrare?” Här svarade 47% ja. Det är många som faller ifrån jämfört med den tidigare frågan, men uppemot 30% av befolkningen (63% · 47%) uppger sig alltså vara beredda att betala för mer transparens. Det är ganska många möjliga kunder.

Iakttagelsen väcker många följdfrågor. Ekonomer brukar skilja på *stated preference* (vad respondenter säger sig vilja i en enkät) och *revealed preference* (hur de faktiskt agerar i verkligheten) och de skiljer sig ofta rejält. Men kanske finns det ändå ett segment som faktiskt har betalningsvilja? De marknadsaktörer som först förstår hur denna potential kan konverteras till försäljning kan ha mycket att vinna.

1.3 Försäkring, teknik och transparens

Försäkringsbranschen står inför ett stort förändringstryck till följd av digitalisering och skärpt konkurrens [28]. Branchorganisationen Svensk Försäkring menar att digitaliseringen ändrar hela värdekedjan: produktdesign, underwriting, distribution och skadereglering [47]. Algoritmer kan skräddarsy produkter, anpassa villkor och prissättning, träffsäkert marknadsföra försäkringar och riskhantera

[4] https://www.regeringen.se/regeringsuppdrag/2019/12/uppdrag-om-att-testa-ny-teknik-vid-automatisering-inom-offentlig-forvaltning/, läst 2019-12-12.
[5] https://www.digg.se/om-oss/nyheter/2020/en-fortroendemodell-skapar-tydlighet-och-tillit-i-myndigheters-automatisering-med-ny-teknik, läst 2020-12-17.
[6] https://www.insightintelligence.se/delade-meningar/delade-meningar-2019/

försäkringsbedrägeri. På konferensen *Forsikring 2018* i Köpenhamn marknadsförde en IBM-partner AI-systemet Watson, som sades snabba upp skadeärendehanteringen med 25%. Men Watson är precis den svarta låda som många är oroade för. Kunderna kan snabbt försvinna om Watson visar sig diskriminera vissa grupper.

På den amerikanska marknaden försöker bolaget Lemonade använda transparens som en konkurrensfördel. Öppenheten gäller framförallt vart de inbetalde premierna går: skador, återförsäkring, vinst – och donationer till välgörenhet som kunderna själva tillåts välja. Det sistnämnda har även det nystartade försäkringsbolaget Hedvig anammat på den svenska marknaden. Det är tydligt att det händer saker på marknaden, men exakt hur AI-användningen får genomslag för svenska försäkringstagare och hur lång tid det tar återstår att se. De flesta svenska försäkringsbolag är ännu mycket långt ifrån att nyttja den fulla potentialen i den nya tekniken [24].

1.4 Forskningsansats och disposition

Resonemangen ovan pekar på att fördelarna med AI och automatiserat beslutsfattande riskerar att gå om intet om vi inte förmår tydliggöra hur maskinerna fattar beslut, åtgärda eventuella fel och bädda in maskinellt beslutsfattande i en juridisk och social kontext som gör att det går att lita på algoritmerna. Problemen med automatiserat beslutsfattande är inte bara tekniska och går knappast heller att lösa med rent tekniska lösningar. Litteraturen pekar snarare på behovet av en bredare ansats, där olika metoder och många intressenter måste ge sina bidrag för att hitta hållbara lösningar [44, 52].

I en sådan bredare ansats måste olika verksamheter lära sig att göra genomtänkta designval både på den tekniska nivån (algoritmer, inlärningsdata, etc.) och på den organisatoriska (styrning, mätetal, kontrakt etc.) [45]. För att lyckas uppstår här ett behov av att kombinera kunskap om hur metoder från olika fält kan användas för att designa rätt sorts transparens i rätt sammanhang. Det är sådan kunskap som projektet TALFÖR har byggt upp, naturligtvis utan att göra anspråk på att fullständigt lösa hela problemkomplexet.

För att lyckas med detta har TALFÖR studerat frågorna om transparens ur ett brett perspektiv. Projektet har bemannats av och samarbetat med forskare med olika bakgrunder; utöver datavetenskapliga ämnen även företagsekonomi, statsvetenskap, filosofi och juridik. Forskningen uppvisar därför en ganska stor metodologisk bredd, men förenas i sin *möjliggörande ansats*: hur kan vi dra nytta av fördelarna med AI och automatiserat beslutsfattande utan att drabbas av dess nackdelar?

De fem följande kapitlen är baserade på vetenskapliga artiklar som har producerats inom ramen för TALFÖR. Kapitel 2 bygger på tidskriftsartikeln [22], kapitel 3 på konferensartikeln [23], kapitel 4 på konferensartikeln [25], kapitel 5 på två manuskript som i skrivande stund författas respektive genomgår vetenskaplig granskning och kapitel 6 på tidskriftsartikeln [24]. Den intresserade läsaren finner mer fullständiga redogörelser av såväl metod som resultat och slutsatser i originalpublikationerna. Tillsammans ger dessa bakåtblickande kapitel en god översikt över vad projektet har åstadkommit i vetenskapligt avseende. Rapportens avslutande, sjunde, kapitel är istället framåtblickande. Där ges en utblick i form av dels några iakttagelser av praktisk relevans för försäkringsbranschen, dels några uppslag till framtida forskning.

2 Nationella AI-strategier och etik

Som nämndes i inledningskapitlet är det många länder och internationella samarbetsorgan som har tagit fram strategier och verksamhetsplaner för AI-området. Teknikutvecklingen erbjuder både möjligheter och svårigheter och det gäller att navigera dem så klokt som möjligt. En översikt över policy-dilemman ges i [15].

Ett område som är både teoretiskt och praktiskt komplicerat handlar om AI och etik. Enkelt uttryckt: Hur kan vi säkerställa att de nya tekniska möjligheterna används för att göra gott snarare än för att göra ont? Till del relaterar dessa svårigheter till transparens: Det kan vara svårt att handla rätt när för den som använder teknik som är svår eller rentav omöjlig att förstå.

2018 gjorde de nordiska länderna ett intressant gemensamt uttalade inom ramen för Nordiska ministerrådet:[7]

> ”Countries that are successful in utilising and realising the benefits of AI, while managing risks responsibly, will have advantages in international competition and in developing more efficient and relevant public sector activities.”

Med andra ord uttrycker de nordiska länderna här en tydlig hypotes: det finns en konkurrensfördel i att bli först och bäst på etisk och hållbar AI. Däremot *argumenterar* de inte för hypotesen och säger inte eller någonting om *hur* de nordiska länderna ska lyckas ta vara på konkurrensfördelen.

Eftersom hypotesen är mycket intressant genomförde TALFÖR en dokumentstudie av nordiska AI-strategier i syfte att noggrannare granska tänkbara argument och mekanismer.

2.1 Analys av strategidokument från fyra länder

Följande strategidokument lästes och analyserades:

- **Danmark:** *National strategi for kunstig intelligens* [7] och *Data i menneskets Tjeneste: Anbefalinger fra Ekspertgruppen om dataetik* [3] ingick i studien. Det första dokumentet är den officiella AI-strategin, gemensamt framtagen av finans- och näringsdepartementen. Den andra rapporten var ett uppenbart komplement för att på ett bra sätt täcka frågorna om transparens, etik och hållbarhet.
- **Finland:** *Finland’s Age of Artificial Intelligence* [1] och *Leading the way into the era of artificial intelligence* [6] ingick i studien. Det första rapporten skrevs 2017 av en arbetsgrupp tillsatt av finansministern. Den andra rapporten utgör slutrapport för programmet och presenterades 2019.
- **Norge:** *Nasjonal Strategi for kunstig intelligens* [8] är den officiella norska AI-strategin, framtagen av kommun- och moderniseringsdepartementet. Rapporten *Kunstig intelligens – muligheter, utfordringar og en plan for Norge* [4] är framtagen av Teknologirådet, som ger råd till Stortinget och regeringen om ny teknik, men inte har något officiellt mandat att utforma politiken.

[7] https://www.norden.org/sv/node/5059, läst 2018-11-09.

- **Sverige:** *Nationell inriktning för artificiell intelligens* [5] och *Artificiell intelligens i svenskt näringsliv och samhälle: Analys av utveckling och potential* [2] ingick i studien. Även om det förstnämnda dokumentet är mycket kort och inte har ordet "strategi" i titeln så utgör det i praktiken den officiella svenska AI-strategin. Den andra rapporten, framtagen av Vinnova på uppdrag av regeringen, är mycket längre om mer omfattande.

Sammantaget speglar urvalet av dokument på ett jämförelsevis heltäckande sätt läget i respektive land. Genom att varje land inte bara representeras av en enda strategi i analysen utan även av ytterligare ett relevant dokument erhålls en mer nyanserad bild.

För att närmare analysera vilka etiska principer som förekommer i dokumenten användes ett ramverk som tagits fram av Floridi et al. [29] som ett led i att EUs *High Level Expert Group (EU HLEG) on Trustworthy AI* skulle ta fram sina rekommendationer. Ramverket är en syntes av flera andra förslag till etiska ramverk, och består av fem principer: *göra gott*, *inte göra skada*, *rättvisa*, *autonomi* och *förklarbarhet* (eng. *explicability*).

Utifrån detta ramverk lästes strategierna och etiskt relevanta stycken kodades utefter de fem principerna. Om det exempelvis hävdades att AI-användning kommer att leda till ekonomisk tillväxt placerades uttalandet i kategorin *göra gott*.

2.2 Slutsatser

De analyserade strategidokumenten innehåller många påståenden och resonemang som avspeglar de fem etiska principerna. Avseende *göra gott* råder påtaglig samsyn kring goda ekonomiska effekter såsom tillväxt, innovation och effektivitetsförbättringar, både i privat och offentlig sektor. Avseende att *inte göra skada* råder samsyn kring frågorna om cybersäkerhet och missbruk av AI samt nära nog samsyn om privatliv och skydd av persondata. Avseende *rättvisa* råder samsyn om effekter på arbetsmarknaden samt kring att undvika (maskiners) fördomar och diskriminering. Avseende *autonomi* råder det påtaglig samsyn mellan de danska, finska och norska strategierna som alla talar om kontroll över och äganderätt till data, kunskap om tekniken och dess följder, vikten av att låta människor utforma den tekniska utvecklingen, människocentrerad AI samt informerade val och samtycke. Slutligen råder det påtaglig samsyn kring *förklarbarhet* i form av ansvarstagande datahantering och öppna data i offentlig sektor. Generellt har de svenska dokumenten en viss övervikt mot att avspegla kategorin *göra gott* medan *autonomi* nästan helt saknas. De norska dokumenten avspeglar främst *autonomi* och *rättvisa*, medan de finska och danska dokumenten är tämligen välbalanserade och avspeglar samtliga kategorier i jämförbar utsträckning.

Angående hur länderna kan realisera konkurrensfördelen med etisk AI var den mest framträdande tanken att det finns en fördel i att vara först (eng. *first mover advantage*). En annan återkommande tanke handlar om att ställa krav på etik i offentlig upphandling för att därigenom skapa en mer mogen AI-etisk marknad med leverantörer som förmår konkurrera internationellt. Andra konkreta åtgärder som föreslås är tydliga standarder och regleringar, framförallt med inriktning på att påverka internationella överenskommelser. Även framtagandet av nationella etiska riktlinjer ses som ett sätt att skapa en konkurrensfördel.

Sist och slutligen är det dock tydligt att strategierna inte är tänkta att vara vägledande i etiska spörsmål. De är oftast framtagna av närings- eller finansdepartement och prioriterar därmed att beskriva de fördelar som tekniken kan ge för marknad och myndigheter, snarare än att prioritera vilka etiska konsekvenser AI kan få och hur dessa bäst ska hanteras.

Att etisk AI är en konkurrensfördel är en attraktiv hypotes. De flesta skulle önska att den är sann. Tyvärr leder strategierna inte detta i bevis på något rigoröst sätt. Den etiska vägledning som ges är implicit och vag, och den empiriska frågan om hur etisk AI faktiskt kan bli en konkurrensfördel – och i så fall under vilka omständigheter – är fortsatt öppen för vidare forskning.

3 Design med hänsyn till transparens

I takt med framstegen på AI-området har många andra akademiska områden fått upp ögonen för frågor om etik och transparens i AI och liknande system. Till det som studeras hör hur begreppen transparens och förklarbarhet bör förstås i olika sammanhang, hur befintliga system kan göras mer transparenta, vilka effekter transparens kan medföra och hur system kan konstrueras för ökad transparens från de allra första designfaserna.

Det är det sistnämnda problemet – att designa med hänsyn till transparens – som behandlas här. TALFÖR har i två små försök studerat möjligheterna att använda ramverket *Value Sensitive Design* (VSD), och speciellt metoden *Envisioning Cards*,[8] för att säkerställa transparens. VSD introducerades av Batya Friedman 1996 [30] som ett ramverk för att systematiskt och proaktivt ta hänsyn till mänskliga värden under en designprocess. Även om ursprungslitteraturen inte definierar exakt vad som avses med mänskliga värden så skriver Friedman i senare arbete att ett värde är ”sådant som är viktigt för människor i deras liv, med fokus på etik och moral” [31]. Hon ger också exempel såsom välmående, privatliv, (maskiners) fördomsfrihet, tillit, autonomi, informerat samtycke och identitet. VSD har fått stort genomslag i den akademiska litteraturen och har bland annat använts i design av informationssystem i allmänhet [32] och AI i synnerhet [26, 50].

Envisioning Cards är en kortlek som används för att förmå workshopdeltagare att tänka bortom sina förutfattade meningar och bryta invanda mönster. Korten är indelade i fyra tematiska kategorier; (i) intressenter, (ii) tidens gång, (iii) värden och (iv) närvaro. Korten är avsedda att vara självförklarande och försedda med frågor som användarna ska reflektera över.

Som antyddes i inledningskapitlet är design av transparens komplicerat. Det är inte så enkelt som att mer transparens alltid är bättre. Turilli & Floridi [48] försöker reda ut svårigheterna genom att tala om transparens som en *pro-etisk omständighet*: transparens *möjliggör* etiska beslut, utan att nödvändigtvis vara etiskt värdefull i sig. Konkret betyder det att transparens ibland kan krävas för att kunna fatta etiska beslut, exempelvis kring när och hur en viss AI-baserad tjänst ska användas eller inte användas. Ibland är det i och för sig etiskt rimligt att *inte* vara transparent, exempelvis med exakt hur system för att leta efter brott eller försäkringsbedrägerier är konstruerade, men reglerna som reglerar denna praxis bör i sin tur vara transparenta.

De försök som genomfördes inom TALFÖR syftade därför till att se om *Envisioning Cards* fungerar som en bra katalysator i diskussioner om transparens som pro-etisk omständighet.

3.1 Två försök

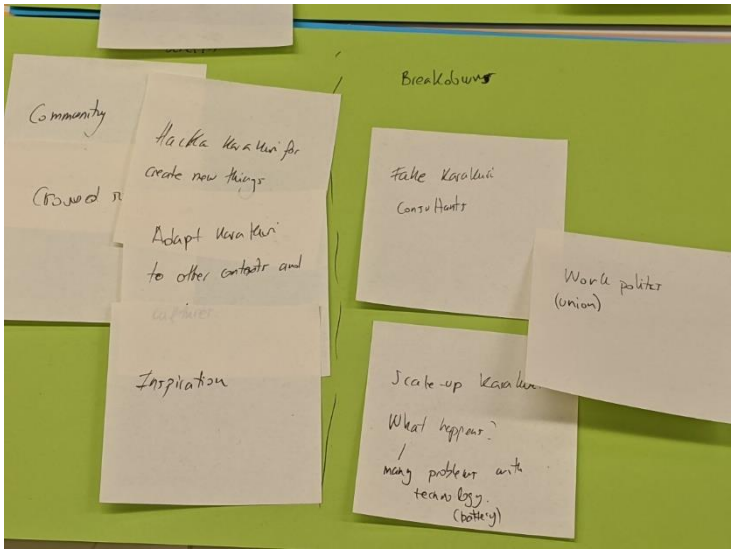
I det första försöket användes *Envisioning Cards* som samtalskatalysator i ett sammanhang med många deltagare som bara var ytligt bekanta med varandra. Försöket genomfördes vid eventet *Financial Markets Transparency*, medarrangerat av TALFÖR, som hölls på Handelshögskolan i Stockholm. Efter ett inledningsanförande om

[8] https://www.envisioningcards.com

finansmarknader och transparens som hölls av Paolo Sironi från IBM Watson Financial Services användes *Envisioning Cards* som medel för att moderera deltagarnas reflektioner. Fyra frågor ställdes och svaren samlades in elektroniskt via tjänsten menti.com. En av frågorna illustreras i Figur 1.

Figur 1: En av frågorna som ställdes och besvarades i det första försöket.

I det andra försöket användes *Envisioning Cards* som samtalskatalysator i ett sammanhang med få deltagare som alla ingick i ett befintligt projektteam. Försöket genomfördes vid ett projektmöte inom ett forskningsprojekt inriktat på sakernas internet som genomfördes i samarbete mellan RISE, en högskola och fem privata bolag i tillverkningsindustrin. En tvåtimmarsworkshop genomfördes med fem projektdeltagare och *Envisioning Cards* användes på mer traditionellt sätt för att stimulera till diskussion. Ett exempel på resultat av brainstorming illustreras i Figur 2.



Figur 2: Ett resultat av brainstorming i det andra försöket.

3.2 Slutsatser

En första slutats som kan dras av de genomförda försöken är att även om *Envisioning Cards* är ett effektivt verktyg för att reflektera över värden och etiska principer så är det

inte säkert att dessa värden och principer för den sakens skull omsätts i praktiken genom transparens.

Den första slutsatsen leder omedelbart vidare till den andra: Om ett transparenskort tillfördes kortleken är det rimligt att *Envisioning Cards* på ett än mer effektivt sätt skulle kunna bidra till att värden förverkligas genom transparens. Här är det viktigt att notera att de två försök som beskrivs ovan är begränsade. Mer empirisk forskning behövs för att utvärdera dels hypotesen om vad ett transparenskort skulle kunna tillföra, dels mer generellt belysa vad som kan och inte rimligen kan åstadkommas med VSD och dess *Envisioning Cards*.

En tredje slutsats är att det finns potential för att använda *Envisioning Cards* som verktyg för etisk reflektion även i stora grupper. I det första försöket lyckades en lättviktsversion av metoden engagera publiken och stimulera till interaktion. Samtidigt är det tydligt att mer forskning behövs för att bättre förstå möjligheterna och begränsningarna med VSD i storgrupp.

4 Hur fungerar GDPR-rätten till förklaring i praktiken?

Den allmänna dataskyddsförordningen (GDPR) har på senare år kommit att bli en viktig referenspunkt i snart sagt alla diskussioner om hur personuppgifter används vid automatiserat beslutsfattande. Det beror inte minst på att EU:s lagstiftning har fått ett starkt genomslag världen över. Stora multinationella företag tycker ofta att det är lättare att globalt anpassa sina tjänster till EU-lagstiftning än att krångla med olika villkor i olika jurisdiktioner. Därmed kan EU genom att gå före med relativt tuffa regler sätta de facto-, om än inte de jure-standarder för hela världen [20, 13].

En av de rättigheter som GDPR inbegriper är en juridisk rätt för individer att begära att få ut ”meningsfull information om logiken bakom” automatiserat beslutsfattande som bygger på deras egna personuppgifter (artikel 15 punkt 1h). Eftersom det finns (legitima) begränsningar för vad företag kan avslöja är det emellertid inte uppenbart vad denna rätt innebär i praktiken. För det första pågår diskussionen vilken tolkning av GDPR-kraven som är juridiskt korrekt. Detta kommer i sinom tid troligen att klargöras genom domstolsbeslut. För det andra finns det en mer praktisk och empirisk fråga: Hur ser praxis ut just nu? Vilken typ av svar kan förväntas på GDPR-förfrågningar med hänvisning till rätten till ”meningsfull information”? Det är det senare som undersöktes av TALFÖR i ett litet experiment.

4.1 Förfrågningar till sju försäkringsbolag

Frågan undersöktes genom förfrågningar till svenska försäkringsbolag. GDPR gäller alla branscher, men försäkringsbranschen är särskilt intressant. För det första hävdar yrkesverksamma inom försäkringsbranschen ofta att de är i en förtroendebransch [51]: produkten går inte att ta på och levereras först i framtiden, så den är omöjlig att sälja om kunden inte litar på försäkringsgivaren. För det andra känner endast 40% av svenskarna förtroende för hur försäkringsföretag hanterar digital personlig information.[9] Även om det är mindre än hälften så ledde det till en tredjeplacering i enkäten, efter banker (68%) och offentlig sektor (46%). Kanske kan försäkringsbranschen därmed vara en förebild för mindre betrodda branscher.

Eftersom rätten till förklaring endast gäller den individ vars information behandlas ställde TALFÖR-forskarna och ytterligare några kolleger förfrågningar till sina försäkringsbolag i egenskap av konsumenter, specifikt av hemföräkringar. De frivilliga personerna hade valts ut för att tillsammans ge en så heltäckande bild av branschen som möjligt och frågan ställdes vintern 2018–2019 till sju bolag som tillsammans står för cirka 90–95% av marknaden för hemförsäkring i Sverige. Den exakta frågeformuleringen återges i Figur 3.

[9] https://www.insightintelligence.se/delade-meningar/delade-meningar-2018/

> Hej!
>
> I enlighet med artikel 15 punkt 1h i EU:s dataskyddsförordning 2016/679 skulle jag vilja få information om hur premien på min hemförsäkring sätts. Artikeln i förordningen torde vara tillämplig om prissättningen (i) är automatiserad och (ii) baseras på personuppgifter (både sådana jag själv har uppgivit och sådana som inhämtats på annat sätt).
>
> Jag tar tacksamt emot informationen i lämplig form (t.ex. matematiska formler eller beskrivande text) som uppfyller förordningens krav på meningsfull information om logiken bakom automatiserat beslutsfattande. Stort tack för hjälpen!
>
> Vänliga hälsningar etc.

Figur 3: Förfrågan till försäkringsbolag.

Svaren skilde sig en hel del åt, både avseende svarstid och svarslängd. Det snabbaste bolaget svarade på ungefär två timmar, medan det långsammaste tog ungefär två månader på sig. Det kortaste svaret var på ungefär 50 ord, medan det längsta var på ungefär 600. Varken svarstid eller svarslängd verkar hänga ihop med kvaliteten eller innehållet i svaret på något uppenbart sätt.

Även vad gäller innehållet fanns det slående skillnader. Tabell 1 ger en översikt över de inkomna svaren. Det är intressant att notera att det inte finns *någon enda* kategori av information som omnämns av samtliga bolag. Det är också slående att uppgifterna om självrisk och försäkringsbelopp, som måste ingå i försäkringsprissättning värd namnet bara omnämns av ett enda bolag.

Tabell 1 Översikt över svarsinnehåll. X betyder att denna sorts uppgift nämndes i svaret; L att både uppgiften och logiken bakom dess användande nämndes. Fastighetsinformation avser villaförsäkringar.

Försäkringsbolag	1	2	3	4	5	6	7
Boyta (m^2)	X	X				X	
Familjeförhållanden (t.ex. antalet boende)	X	X		X			X
Adress	X	L	L	X	X		X
Fastighetsinformation		X		X		X	
Ålder	X	L	L		X		X
Självrisk		X					
Försäkringsbelopp		X					
Inkomster och andra taxeringsuppgifter		X		X			
Säkerhetsåtgärder (t.ex. lås)		X		X			
Försäkringens ålder (lojalitetsmått)		L					X
Historik över ersättningsärenden		X		X			X

4.2 Slutsatser

Det är tydligt att den rätt till förklaring som GDPR ger enskilda individer i praktiken fungerar olika beroende på omständigheterna. I undersökningen studerades försäkringsbranschen, en bransch vars datahantering åtnjuter relativt högt förtroende bland allmänheten och som även får sägas vara van vid att hantera relativt komplex reglering. Ändå skiljer sig svaren betänkligt åt bolagen emellan.

Vad avser svarstiderna kan konstateras att inget företag överskred den övre tremånadersgräns som GDPR stipulerar. Samtidigt är förfrågningarna liksom svaren relativt enkla, så det finns egentligen inget skäl att tro att svarstiden skulle kunna förlängas enligt artikel 12.3 i GDPR, med hänvisning till att begäran skulle vara komplicerad. Hursomhelst kräver en sådan förlängning att den registrerade inom en månad underrättas om förseningen, vilket försäkringsbolaget som dröjde 2,5 månader inte gjorde. Hanteringen av denna begäran uppfyller uppenbarligen inte lagen. Även om alla andra förfrågningar hanterades inom en månad är det osannolikt att de två svar som dröjde upp till en månad uppfyller kravet att svara ”utan onödigt dröjsmål”. Även om dess betydelse ännu inte har tolkats av EU-domstolen, antyder dess vanliga innebörd i lag att frågan bör behandlas snarast och att varje försening som överstiger några arbetsdagar måste motiveras, vilket inte skedde.

Att avgöra huruvida innehållet i svaren faktiskt uppfyller lagens krav på att vara meningsfullt är naturligtvis en svårare fråga som skulle kunna utforskas genom en djupare juridisk analys. Ett alternativt tillvägagångssätt, mer empiriskt orienterat, vore att genomföra användarstudier som försöker svara på frågan om mottagarna av denna sorts svar faktiskt blir klokare.

För att ytterligare belysa frågan och sätta den i europeisk kontext initierades mot slutet av TALFÖR en fortsättningsstudie där forskargrupper i andra EU-länder kontaktades för att göra liknande experiment på sina respektive hemförsäkringsmarknader. Denna studie, som i skrivande stund pågår, kommer i sinom tid att ge en mognare och mer fullständig bild av vad GDPR-rätten till meningsfull information faktiskt innebär i praktiken.

5 Strategisk ledning och styrning av AI

Som beskrivits i tidigare kapitel pågår det flera utvecklingar på både politisk och teknisk nivå som dels främjar utvecklingen av AI, dels lyfter transparens eller förklarbarhet som nycklar till att dra full nytta av tekniken. Men hur avspeglas denna samhällsutveckling på insidan av enskilda organisationer? Den frågeställningen var drivkraften bakom de två studier som beskrivs i det följande. Studierna kom att arbeta utifrån delprojektnamnet *governance*, som i någon form motsvaras av styrning på svenska.

Organisationer kommer troligen att behöva ta fram och underhålla både strukturer och processer för att möjliggöra att verksamheterna å ena sidan kan dra nytta av AI-trenden och å ena sidan tillfredsställa kraven på transparens. Men det brukar finnas interna drivkrafter både inom branscher och organisationer som dels stödjer, dels hindrar sådan förändring. Många gånger blir det en uppgift för högsta ledningen i ett företag att balansera dessa drivkrafter. Frågorna kan därmed hamna på styrelsernas bord. Här finns det dock en lucka i den befintliga forskningen, där den högsta ledningen ännu inte har studerats empiriskt för att förstå vad design för transparens innebär. Tidigare forskning har snarare nöjt sig med normativa uttalanden om hur ledningarna *borde* agera.

Datainsamlingen genomfördes genom att i olika omgångar träffa och intervjua ett trettiotal högre chefer och styrelseledamöter, dels inom Länsförsäkringar, dels hos ett antal andra storbanker och försäkringsbolag. Det som kom ut från dessa möten var främst beskrivningar av avsikter och ståndpunkter som senare genom tolkningar har blivit de två studier som beskrivs i det följande.

5.1 Studie 1, en infrastrukturmodell

Den första studien kom att handla om förekomsten av en infrastruktur för transparens. Idén med infrastruktur är ofta tillämplig för att förstå större designproblem på samhällsnivå, men här har idén alltså använts för att förstå designproblem på organisationsnivå. För mindre organisationer är detta kanske ett olämpligt perspektiv. Men många gånger är större försäkringsbolag eller storbanker relativt komplexa organisationer, inte enbart för att många olika verksamheter, strukturer och processer ständigt finns och pågår, utan också utifrån de olikartade idéer som samexisterar i styrningen av företaget. Denna grad av komplexitet går att jämföra med den komplexitet som förekommer i ett samhälle. Genom ett antal intervjuer och uppföljande verifieringar målades en bild av en möjlig infrastruktur upp. Genom ett analytiskt grepp i tolkningen av intervjudata skapades tre huvudkategorier: redovisning, organisering och mening samt sjutton underkategorier.

5.2 Slutsatser utifrån studie 1

De tre huvudkategorierna i den tänkta infrastrukturen framgår av Tabell 2 nedan.

Tabell 2 En modell av infrastruktur för transparens.

Redovisning	Organisering	Mening
Redogörelser för branschkollegor etc. Redovisning utifrån expertis Preferens för informationstyp Redovisning för medlemmar av organisationer Regelefterlevnadsbehov	Driv för AI Push & pullstrategier Managementstil Resurstillgång Kompetensprofiler på styrelsenivå Kompetensprofil inom AI-organisationer Reglering	Hopp om AI Kontrollbenägenhet av aktivister Inställning till ny teknik Människors läggning Attityder till förändring

Av utrymmesskäl beskrivs huvudkategorierna nedan. De sjutton underkategorierna är förhoppningsvis relativt självförklarande.

- **Redovisning** kan i sin enklaste form beskrivas som muntliga och skriftliga redogörelser för händelser. Forskningen har visat att redovisning använder sig av teknik som framställer befintliga redogörelser som objektiva och särkopplade från de händelser som de handlar om, till exempel bokföring, men redovisning har även visat sig kunna inkludera avsikter och redogörelserna kan vara subjektiva till sin natur. Vikten av en sådan del i infrastrukturen är att den överensstämmer med förslaget att transparens kan separeras från skeenden och framställas vid behov. Detta förhållningssätt förekommer många gånger i de krav som lagstiftare och myndigheter ställer på finansföretagen att verka för omfattande redovisning (eng. *disclosure*).
- **Organisering** är alla försök att skapa överbyggnader och processer som på ett eller annat sätt ger insyn i verksamheter och möjliggör övervakning (eng. *monitoring*) av pågående processer. Exempelvis förekommer det många gånger att revisorer kontrollerar affärsprocesser på avstånd och producerar mått, såsom produktivitetsmått. Sådana mått kan användas för att fokusera ledningens uppmärksamhet på eventuella problem och reglera verksamheters resursanvändning. Vikten av denna del i infrastrukturen relaterar till tanken att transparens leder till skapandet av nya informationslager som möjliggör skapandet av sanningar (såsom fakta) som ger nytta i form av lärande.
- **Mening** legitimerar handlingar och beslut. Den tredje komponenten i infrastrukturen bygger på idén att transparens bygger på positiva förväntningar om ökad insyn och genomskinlighet. Transparens baseras på att särskilda enheter (till exempel högt kvalificerade specialister) och teknik (till exempel avancerade system) tillsammans erbjuder specialverktyg, som har tillräckligt förtroendekapital för att motivera beslut. Exempelvis bygger AI på idén att experter och informationssystem ska kunna genomföra avancerade beräkningar som ger inblick i risk. På så sätt kan AI fatta beslut såsom att sätta pris på en försäkring. Vikten av denna del i infrastrukturen är sammankopplad med förväntningar om att ökad insyn, till exempel *big data*, ska ha positiv effekt på utbyten och relationer, till exempel i form av lägre transaktionskostnader och fler bättre beslut.

Upptäckten att design för transparens kan förstås utifrån en infrastrukturmodell kan vara viktig att ha med sig för att ledningen i ett företag ska kunna fånga alla de aktiviteter som pågår för att möjliggöra AI och skapa transparens. Genom denna vetskap kan

ledningen lättare säkerställa att aktiviteter sker på ett samordnat sätt för att bibehålla kontrollen under utvecklingens gång.

Denna tolkning av studiens resultat bygger på ett antal tester som genomfördes med både styrelsepersoner och personer som ansvarar för att lyckas med AI-projekt. Intervjupersonerna fick utifrån idén att utvecklingen av AI kan genomgå tre skilda faser, (i) reaktiv, (ii) proaktiv och (iii) integration, svara på hur de förhåller sig till de tre delarna i infrastrukturen. Testresultaten visade på att det kan finnas risk att aktiviteterna i ett företag inte sker på ett sammanhållet sätt. Detta testverktyg kan alltså användas för att undersöka hur personer i ett företag tänker om infrastrukturen och dess beståndsdelar. Därigenom blir det alltså möjligt att jämföra personers tankesätt och diskutera gemensamma ståndpunkter.

5.3 Studie 2, ett synsätt på riskhantering

Det andra studien kom att handla om behovet av att förstå hur transparens kan användas för riskhantering. Många gånger framställs AI som ett projekt som medför både för- och nackdelar. Inom forskningen om risk finns idén att riskhantering handlar om att arbeta utifrån både möjligheter och faror och att det är genom att ändamålsenligt hantera båda två samtidigt som framgång nås.

Även den andra studien tog avstamp i en empiriskt grundad problemställning och observationen att det har utvecklats standarder och ramverk som uppmanar till att hantera AI-relaterade risker (se även kapitel 2 ovan). Gemensamt för dessa storskaliga försök är hänvisningar till förtroende som den självklara mekanismen och att vägen till detta förtroende kan gå via transparens. Däremot verkar dessa standarder och ramverk sakna förståelse för processer som pågår inom organisationer sett utifrån den normativa hållning som ofta finns i formuleringar och anvisningar. Samtidigt behöver organisationer leva upp till sådana krav på AI-relaterad riskhantering. Det är väl känt att finansbranschen står inför båda strategiska och operationella risker relaterade till AI-utvecklingen. Tabell 3 nedan ger några exempel.

Tabell 3 Exempel på risker för finansföretag.

Riskkategori	Faror	Möjligheter
Operationella risker	Diskriminering av kunder	Eliminera ineffektivitet
Strategiska risker	Företaget hamnar på efterkälken	Bättre marginaler och nya marknader

Modellen i tabellen har med framgång testats på fyra styrelseordföranden i olika finansbolag. Ytterligare tester är planerade under våren 2021.

Vidare finns det i finansbranschen flera grundläggande omständigheter som starkt talar för transparens, exempelvis teorin om effektiva marknader och regleringar som ställer krav på omfattande redovisningar. Den teoretiska ingången till studien var en iakttagelse i tidigare försök till teoretisering: transparens kan ses som ett flytande objekt, det vill säga den tenderar att inte ha en fast form utan förklaras bäst av en mångfald ståndpunkter. Utifrån sådana empiriska och teoretiska ledtrådar fick studien upp ögonen för ett antal frågeställningar som på sistone dykt upp inte bara i

försäkringsföretag utan även i banker. I Tabell 4 nedan presenteras ett axplock av frågeställningar utifrån formuleringar i intervjuer med olika befattningshavare.

Tabell 4 Exempel på aktuella frågor i finansföretag

AI experter	Affärsfolket	Toppchefer	Styrelsepersoner
Problemet är hierarkin och att vi inte talar samma språk	Vad kan AI-experter bevisa när det handlar om risk?	Hur vet jag att AI fattar rätt beslut?	Magkänslan, då?

Även den andra studien krävde intervjuer och verifieringar av de resultat som kortfattat presenteras nedan.

5.4 Slutsatser utifrån studie 2

Den första upptäckten var att transparens kan ses som något som möjliggör argumentation kring för- och nackdelar med AI. Transparens kan efter behov användas i utvecklingsprojekt för att ta fram och förfina AI och hantera uppkomna risker. Det hela kan liknas vid hur en diskussion förs. Samtidigt medför inslaget av löpande diskussion att det är orimligt att förvänta sig någon slutgiltig sanning om AI-riskerna, som en gång för alla slås fast efter avslutad diskussion.

Flera observationer från intervjuerna gav bilden av att hantering av AI-risker kräver en kollektiv bedömning av information som uppnås genom diskussion på olika organisatoriska nivåer. Det förekom även uppfattningar om att beslut kring AI-risker inte bör fattas endast genom att alla underkastar sig ett chefsbeslut. I en organisation kan makt utövas av många; från specialister såsom AI experter till formell ledning såsom styrelseledamöter. Specialister kan påverka beslut genom sin kunskap, även om denna möjlighet kan vara begränsad i förhållande till den formella beslutsmakt som en ledningsgrupp besitter. För den som är satt att hantera AI-risker i en organisation kan det vara av stort praktiskt värde att förstå hur och vad som uppfattas som sanningen om dessa risker är kontextberoende och formas av en löpande diskussion. Det är en mer nyanserad bild än vad gängse standarder och ramverk för AI ger.

Den andra upptäckten var att hantering av strategiska risker med AI ställer krav på gemensamma referensramar, vilket i sig utgör ett slags transparens. Men i verkliga styrelser finns det ofta olika uppsättningar referensramar, exempelvis avseende frågor om risk, ersättning etc. Denna upptäckt gjordes genom en kvalitativ genomgång av intervjupersonernas beskrivningar av strukturer, ansvarsområden och kompetens. En slutsats är att gemensamma referensramar skapar stabilitet avseende transparens och därmed reglerar vilken information som görs tillgänglig. Omvänt gäller att brist på gemensamma referensramar hindrar transparens från att uppstå, samtidigt som tillgången kan omförhandlas om de strategiska riskerna tilltar.

Medan den första upptäckten handlar om frihet att tänka och diskussion över gränser handlar den andra upptäckten om en begränsning av diskussionen av hur strategiska risker ska hanteras. Det är väl känt att etablerade finansbolag behöver hantera fler strategiska risker i och med utvecklingen av AI. I praktiken kan upptäckten bidra till ändamålsenlig design av gemensamma referensramar i större eller mindre grupperingar.

Behövs det exempelvis ett särskilt utskott för att diskutera AI på styrelsenivå? Eller kan redan befintliga utskott dela upp AI-frågorna sinsemellan?

Den tredje upptäckten är att för hanteringen av operationella risker verkar transparens inte gå att reducera till enskilda beståndsdelar. Denna upptäckt gjordes genom att studera hur intervjupersonerna beskrev riskhanteringen kopplad till olika AI-projekt i respektive organisation. Enligt beskrivningarna krävs det att flera delar, såsom strukturer, processer, kompetens, roller, möten, information, rapporter etc. *alla* ordnas på ett sådant sätt att transparens uppstår. Bakgrunden var både den högsta ledningens behov av uppföljning och projektmedlemmarnas behov av insyn i vad som sker – även bortom sådana gränser som strukturer och processer definierar. I praktiken kan upptäckten bidra till ändamålsenlig design för transparens i samtliga delar av en organisation och undvika ”suboptimeringar”; att vissa delar lyfts fram på ett sätt som missgynnar helheten. Denna designuppgift behöver också förhålla sig till en teknisk kärnproblematik: AI med begränsad transparens förväntas kunna fatta beslut på ett självständigt sätt, exempelvis när en chatbot interagerar med en kund. Brist på transparens i själva tekniken medför såklart en fara för felaktiga beslut; en risk som är svår att bilda sig en rättvisande uppfattning om.

6 Den svenska försäkringsbranschens syn på AI och transparens

Som diskuterades i inledningskapitlet medför teknikutvecklingen nya utmaningar. Å ena sidan finns det mycket att vinna på ökad digitalisering och mer automatisering, inte minst i försäkringsbranschen [47]. Å andra sidan är det riskabelt att ta svåröverskådlig teknik i bruk. Vems är ansvaret om ett tekniskt system visar sig hantera persondata oförsiktigt eller fatta fördomsfulla beslut baserat på skev inlärningsdata? Det behöver inte ha funnits ont uppsåt för att någon ska drabbas av konsumenternas eller lagstiftarens vrede om misstag uppdagas.

Dessa utmaningar gäller generellt, men försäkringsbranschen är speciell just eftersom den är en utpräglad förtroendebransch som bygger på att såväl kunder [51] som exempelvis försäkringsmäklare [53] känner tillit. Om användningen av kundernas data för mer automatiserat beslutsfattande leder till minskad tillit så är det med andra ord ett stort problem; inte bara för försäkringsbranschen, utan också för samhället i stort. Försäkring är nämligen en väldigt bra mekanism för att kollektivt hantera och sprida sådana risker som är för stora att hantera individuellt.

En till synes uppenbar lösning på problemet är att införa AI och automatiserat beslutsfattande på ett öppet och transparent sätt och därigenom bevara tilliten. Precis den mekanismen har föreslagits i andra försäkringssammanhang, exempelvis vikten av att ge kunder tid att förstå sina avtal [12]. Samtidigt finns det en generell misstro bland konsumenter gentemot företag i största allmänhet [10] och konsumenter är, med viss rätta, oroliga för hur deras persondata används [40].

För att ytterligare belysa dessa problemställningar genomförde TALFÖR en intervjustudie med representanter från den svenska försäkringsbranschen.

6.1 Intervjuer med åtta representanter för fyra försäkringsbolag

Åtta informanter intervjuades under hösten 2019 och våren 2020. Samtliga hade chefspositioner på svenska försäkringsbolag. Fyra bolag var representerade, vilket svarar mot knappt hälften av den svenska konsumentmarknaden för skadeförsäkring. Det var en medveten strategi att tala med flera representanter per bolag i syfte att få en mer heltäckande bild utifrån olika roller och olika försäkringsprodukter. Intervjuerna tog en knapp timme.

Intervjuerna var semistrukturerade, vilket betyder att det fanns en uppsättning förberedda frågor som täcktes, men att informanterna också gavs frihet att lägga till ytterligare kommentarer och reflektioner. Intervjuerna kretsade kring huruvida transparens och öppenhet kan vara en konkurrensfördel, vad som är rimligt att vara transparent kring, hur transparenta konkurrenterna på den svenska försäkringsmarknaden är och om det finns någon information som kunderna absolut inte bör ha tillgång till.

6.2 Slutsatser

Bland informanterna fanns det en utbredd uppfattning om att transparens kan vara en konkurrensfördel. Däremot hade de svårare att peka på exakt under vilka omständigheter detta gäller, eller vilka mekanismer som ligger bakom. (Detta är inte helt olikt de nationella strategierna som diskuteras i kapitel 2.) Därför är det ingen överraskning att de flesta av de fyra bolagen *inte* använde transparens på något strategiskt sätt för att uppnå konkurrensfördelar.

I avsaknad av strategi är det heller inte förvånande att informanter från samtliga bolag ser förbättringspotential: områden där de inte är transparenta idag, men borde vara det. Det finns också slående likheter i hur informanterna vill använda ökad transparens för att ge kunderna rätt förväntningar på försäkringsprodukten. Därigenom minskar risken för framtida besvikelser.

Vad gäller begränsningar i transparens – sådant som bolagen inte bör eller vill vara öppna med – identifierade informanterna tre kategorier: (i) begränsningar som följer av rättsliga krav, (ii) begränsningar som följer av att det är för svårt att göra vissa typer av information förståelig samt (iii) begränsningar som beror på risken att avslöja affärskänslig information för konkurrenter och kunder. Det är också värt att notera att informanterna *inte* var överens om exakt vilken information som är svår att göra förståelig, eller hur detta kan bli bättre.

Vad gäller AI-användningen är det också uppenbart att AI ännu inte används i försäkringsbolagens kärnverksamhet i Sverige (däremot för mer avgränsade uppgifter som chattbottar). Flera av informanterna såg därför frågan om transparens rörande AI-användning som ett problem som än så länge är teoretiskt. I takt med ökad AI-användning väntas frågorna dock bli viktigare.

Sammanfattningsvis tycks det finnas en tro på transparens som en potentiell konkurrensfördel i försäkringsbranschen, men det råder oenighet om hur stor fördelen är och om hur den kan förverkligas. Ingenting tyder på att transparens och öppenhet ännu används strategiskt i branschen, även om intervjuerna pekade ut områden där bolagen skulle kunna förbättra sin transparens. Informanterna såg flera begränsningar i hur mycket transparens som är lämplig. Det råder också stor osäkerhet kring framtida krav, eftersom branschen ännu inte använder AI i själva kärnverksamheten.

7 Utblick och avslutning

Det är tydligt att frågorna om transparens och förklarbarhet väcker stort intresse, såväl i forskarvärlden som i det offentliga samtalet. Potentialen för digitalisering och nya AI-baserade tjänster att göra gott är stor, både generellt och inom försäkring [47]. Samtidigt är det tydligt att samhällets syn på smarta maskininlärningsbaserade tjänster kan skifta snabbt. I ena stunden kan sådana tjänster vara oerhört populära och framgångsrika, för att nästa stund se sitt rykte snabbt solkas om det framkommer tveksamheter kring etik och hantering av data. Det kan i sin tur leda till att de datamängder som tjänsterna bygger på sinar, när konsumenter väljer att sluta använda tjänsten eller nyttjar GDPR-möjligheten att få data raderad. Detta kan naturligtvis vara helt förödande för en datadriven verksamhet. Försäkringsbranschen är inget undantag. Forskning visar att den negativa effekten av en dålig erfarenhet av ett försäkringsbolag är starkare än den positiva effekten av en god erfarenhet [17].

De föregående kapitlen beskriver den forskning som har genomförts inom projektet TALFÖR. Perspektivet har varit möjliggörande: Går det att använda transparens för att kunna dra nytta av teknikens fördelar utan att drabbas av dess nackdelar? Vikten av den frågan kan knappast överdrivas: Det vore en ofantlig förlust om mänskligheten skulle bli utlämnad till obegripliga och fördomsfulla algoritmers godtycke. Det vore också en ofantlig förlust om mänskligheten skulle avstå från algoritmer som med gott resultat kan hjälpa oss att analysera stora mängder data och fatta klokare beslut. Transparens och förklarbarhet utgör otvivelaktigt nycklar till att undvika detta dilemma och skapa ett tredje, bättre, alternativ.

7.1 Praktisk relevans för försäkringsbranschen

Försäkringsbranschen behöver å ena sidan anamma den nya teknikens möjligheter, men å andra sidan göra det på ett genomtänkt och förtroendeingivande sätt, i samklang både med gällande regelverk och med det omgivande samhällets uppfattning om rättvisa. Några praktiskt relevanta observationer för att lyckas med ett sådant ansvarsfullt AI-införande är följande:

- **Osäkert tempo:** Som framgår av kapitel 6 används AI ännu inte på bredden i försäkringsbolagens kärnverksamhet i Sverige. Men potentialen är stor, så detta kan snabbt ändras. Om någon aktör plötsligt drar ifrån så kan det skapa ett plötsligt tryck hos andra att komma ifatt. Det är kanske i denna situation som riskerna utifrån transparens- och förklarbarhetsperspektiv är som störst: framskyndade lösningar löper stor risk att innehålla svagheter som straffar sig senare.
- **Konsumentreaktioner:** Som framgår av kapitel 1 finns det åtminstone en potentiell betalningsvilja för digitala tjänster som är mer transparenta med hur de använder data. Hur stor den potentialen faktiskt är återstår att se, liksom hur den fördelas mellan olika kundkategorier och produkter. Kanske finns det nischer för företag som möter dessa kunder? Internetleverantören Bahnhof är ett exempel på ett företag i en annan bransch som appellerar till ett särskilt kundsegment med en viss syn på frågor om integritet och dataskydd. Det går att tänka sig liknande nischer inom försäkring också, men antalet kunder i en sådan nisch kan vara begränsat.

- **AI-risker och försäkring:** Försäkringsbranschen delar många utmaningar med andra branscher: hur AI bäst ska infogas i verksamheten, hur den ska upphandlas och kravställas, vad som bör vara transparent och på vilket sätt. Men försäkringsbranschen har också en unik utmaning: hur ska AI-relaterade risker hos försäkringstagare bedömas, prissättas och försäkras? Det är en frågeställning som tarvar en hel uppsats i sig, men den är en kärnfråga för försäkringsbranschen. Lyckas branschen så kan det möjliggöra teknikskiften till gagn för hela samhället. Ett misslyckande kan å andra sidan förhindra önskvärda förändringar eftersom nya tjänster och produkter inte går att försäkra.
- **Tänka utanför lådan:** Det är oerhört svårt att föreställa sig konsekvenserna av ny teknik, inte minst på längre sikt. Historien är full av spektakulära felbedömningar som Nokias underskattning av marknaden för smarta telefoner, eller för den delen chefsingenjören på det brittiska postverket sir William Preeces uttalande 1882 att telefonen kanske kunde vara användbar i Amerika, men knappast för britterna som ”hade gott om springpojkar” [33]. Men även om det är svårt att förutsäga framtiden så är det rimligt att försöka undvika de värsta misstagen genom ett ansvarstagande designarbete. Som diskuteras i kapitel 3 är *Value Sensitive Design* en metod som låter olika intressenter tillsammans tänka bortom invanda tankemönster och därigenom skapa mer genomtänkta produkter. Systematiskt användande av sådana metoder kan vara ett sätt att kvalitetssäkra AI-användning och skapa genomtänkt ändamålsenlig transparens.

7.2 Framtida forskning

Ur ett vetenskapligt perspektiv går det, utan anspråk på fullständighet, att identifiera några områden där framtida forskning skulle kunna leda till värdefull kunskap:

- **Människa-datorinteraktion:** Mycket av forskningen inom förklarbar AI utgår från tekniska lösningar. Exempelvis tas mått på förklarbarhet fram baserat på önskvärda matematiska egenskaper och verktyg för ökad transparens utvecklas ofta för att passa den som programmerar sådana system. Mer sällan syns i litteraturen att sådana mått eller verktyg valideras genom systematiska och välplanerade användarförsök. Detta område – forskning inom människa-datorinteraktion med inriktning på förklararhet [9] – framstår som mycket fruktbart för framtiden.
- **Upphandling och kravställning:** De allra flesta företag, myndigheter och organisationer bygger inte själva sina maskininlärda system från grunden. Sådana tjänster *upphandlas* istället från andra – underleverantörer som är experter på mjukvara sådant som bildigenkänning, texthantering eller riskbedömning. Därför finns det behov av mer forskning kring upphandling och kravställning av transparens och förklarbarhet. Sådan kunskap är nödvändig för att kunna dra nytta av arbetsdelning och specialisering utan att tappa greppet om hur tekniken används i den egna verksamheten.
- **Ekonomi och transparens**: Avslutningsvis vore det intressant med mer transparensforskning i skärningen mellan datavetenskap och nationalekonomi. Hur tillgång till information påverkar ekonomiskt beslutsfattande är ett stort och etablerat område i ekonomisk forskning [46, 11, 43]. 2020 års pris i ekonomisk vetenskap till Alfred Nobels minne är ett bra och aktuellt exempel. Pristagarna Paul Milgrom och Robert Wilson belönades för sitt arbete med auktioner – ett praktiskt verktyg som används hela tiden i samhället och där det spelar stor roll exempelvis om budgivningen är öppen eller sluten. Kombinationen av ett sådant

ekonomiskt perspektiv på information med de datavetenskapliga perspektiven på maskininlärning och förklarbarhet framstår som ett område med framtiden för sig. Det finns all anledning att tro att sådan kunskap inte bara kan öka vår förståelse, utan även vår förmåga att designa system på ett önskvärt sätt.

Referenser

[1] *Finland's Age of Artificial Intelligence: Turning Finland into a leading country in the application of artificial intelligence Objective and recommendations for measures*. Finnish Ministry of Economic Affairs and Employment, 2017. http://urn.fi/URN:ISBN:978-952-327-290-3.

[2] *Artificiell intelligens i svenskt näringsliv och samhälle: Analys av utveckling och potential*. Vinnova, 2018. https://www.vinnova.se/publikationer/artificiell-intelligens-i-svenskt-naringsliv-och-samhalle/.

[3] *Data i menneskets tjeneste: Anbefalinger fra Ekspertgruppen om dataetik*. Erhvervsministeriet, 2018. https://em.dk/publikationer/2018/ekspertgruppen-om-dataetiks-rapport.

[4] *Kunstig intelligens – muligheter, utfordringar og en plan for Norge*. Teknologirådet, 2018. https://teknologiradet.no/wp-content/uploads/sites/105/2018/09/Rapport-Kunstig-intelligens-og-maskinlaering-til-nett.pdf.

[5] *Nationell inriktning för artificiell intelligens*. Näringsdepartementet, 2018. https://www.regeringen.se/49a828/contentassets/844d30fb0d594d1b9d96e2f5d57ed14b/2018ai_webb.pdf.

[6] *Leading the way into the era of artificial intelligence: Final report of Finland's Artificial Intelligence Programme 2019*. Finnish Ministry of Economic Affairs and Employment, 2019. http://urn.fi/URN:ISBN:978-952-327-437-2.

[7] *National strategi for kunstig intelligens*. Finansministeriet og Erhvervsministeriet, 2019. https://em.dk/publikationer/2019/national-strategi-for-kunstig-intelligens/.

[8] *Nasjonal Strategi for kunstig intelligens*. Kommunal- og moderniseringsdepartementet, 2020. https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/no/pdfs/ki-strategi.pdf.

[9] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim och M. Kankanhalli. Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. I: *Proceedings of the 2018 CHI conference on human factors in computing systems*, ss 1–18, 2018.

[10] J. E. Adams, S. Highhouse och M. J. Zickar. Understanding general distrust of corporations. *Corporate Reputation Review*, 13(1):38, 2010.

[11] G. A. Akerlof. The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*, 84(3):488–500, 1970.

[12] B. K. Atchinson. Walking the talk: Ethics as corporate culture. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 29(1):40–44, 2004.

[13] A. Bradford. *The Brussels Effect: How the European Union Rules the World*. Oxford University Press, 2020.

[14] A. Caliskan, J. J. Bryson och A. Narayanan. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186, 2017.

[15] R. Calo. Artificial Intelligence Policy: A Primer and Roadmap. *University of Bologna Law Review*, 3(2):180–218, 2018.

[16] D. Castelvecchi. Can we open the black box of ai? *Nature News*, 538(7623):20, 2016.

[17] C. Courbage och C. Nicolas. Trust in insurance: The importance of experiences. *Journal of Risk and Insurance*, 2020. In press.

[18] T. Cross. Artificial intelligence and its limits: Steeper than expected. *The Economist*, 2020. Technology Quarterly, 13 juni.

[19] M. Cucciniello, G. A. Porumbescu och S. Grimmelikhuijsen. 25 years of transparency research: Evidence and future directions. *Public Administration Review*, 77(1):32–44, 2017.

[20] C. Damro. Market power Europe. *Journal of European Public Policy: EUSA 2011 Boston Conference papers*, 19(5):682–699, 2012.

[21] T. Denk, K. Hedström och F. Karlsson. Medborgarna och automatiserat beslutsfattande. I: U. Andersson, B. Rönnerstrand, P. Öhberg och A. Bergström, red., *Storm och stiltje*. SOM Institute, University of Gothenburg, 2019.

[22] J. Dexe och U. Franke. Nordic lights? National AI policies for doing well by doing good. *Journal of Cyber Policy*, 5:332–349, 2020.

[23] J. Dexe, U. Franke, A. Avatare Nöu och A. Rad. Towards Increased Transparency with Value Sensitive Design. I: *Artificial Intelligence in HCI. HCI International 2020.*, ss 3–15. Springer, juli 2020.

[24] J. Dexe, U. Franke och A. Rad. Transparency and insurance professionals – A study of Swedish insurance practice attitudes and future development. *The Geneva Papers on Risk and Insurance – Issues and Practice*, 2021. Under tryckning.

[25] J. Dexe, J. Ledendal och U. Franke. An empirical investigation of the right to explanation under GDPR in insurance. I: *Trust, Privacy and Security in Digital Business. The 17th International Conference on Trust, Privacy and Security in Digital Business – TrustBus 2020*. Springer, sept. 2020.

[26] V. Dignum. Responsible artificial intelligence: designing AI for human values. *ITU Journal: ICT Discoveries*, (1), 2017.

[27] M. Du, N. Liu och X. Hu. Techniques for interpretable machine learning. *Communications of the ACM*, 63(1):68–77, 2019.

[28] The future of insurance: Counsel of protection. *The Economist*, (March 11):67–68, 2019.

[29] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi m.fl. AI4People – An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4):689–707, 2018.

[30] B. Friedman. Value-sensitive Design. *interactions*, 3(6):16–23, dec. 1996.

[31] B. Friedman och D. G. Hendry. *Value Sensitive Design: Shaping Technology with Moral Imagination*. The MIT Press, Cambridge, MA, 2019.

[32] B. Friedman, P. H. Kahn, A. Borning och A. Huldtgren. Value sensitive design and information systems. I: N. Doorn, D. Schuurbiers, I. van de Poel och M. E. Gorman, red., *Early engagement and new technologies: Opening up the laboratory*, ss 55–95. Springer Netherlands, Dordrecht, 2013.

[33] J. Grafström. *Moderna tider 4.0: Från kugge i maskineriet till vinnare bland algoritmerna*. Volante, 2020. s. 44.

[34] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti och D. Pedreschi. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5):1–42, 2018.

[35] K. He, X. Zhang, S. Ren och J. Sun. Deep residual learning for image recognition. I: *Proceedings of the IEEE conference on computer vision and pattern recognition*, ss 770–778, 2016.

[36] J. Hirschberg och C. D. Manning. Advances in natural language processing. *Science*, 349(6245):261–266, 2015.

[37] M. Hutson. Eye-catching advances in some AI fields are not real. *Science*, 2020.

[38] J. Kang och G. Hustvedt. Building trust between consumers and corporations: The role of consumer perceptions of transparency and social responsibility. *Journal of Business Ethics*, 125(2):253–265, 2014.

[39] J. Kung. *Building an AI world: Report on National and Regional AI Strategies*. Toronto, ON: Canadian Institute for Advanced Research CIFAR, 2020. https://www.cifar.ca/docs/default-source/ai-reports/building-an-ai-world-second-edition-f.pdf.

[40] T. Morey, T. Forbath och A. Schoop. Customer data: Designing for transparency and trust. *Harvard Business Review*, 93(5):96–105, 2015.

[41] More accountability for big-data algorithms. *Nature*, 537(7621):449, 2016.

[42] C. O'Neil. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books, 2016.

[43] A. Prat. The wrong kind of transparency. *American economic review*, 95(3):862–877, 2005.

[44] F. Rossi. Building trust in artificial intelligence. *Journal of International Affairs*, 72(1):127–134, 2018.

[45] H. Schildt. Big data and organizational design–the brave new world of algorithmic management and computer augmented transparency. *Innovation*, 19(1):23–30, 2017.

[46] G. J. Stigler. The economics of information. *Journal of political economy*, 69(3):213–225, 1961.

[47] Omvärldstrender 2017 – utmaningar och möjligheter för försäkringsbranschen. Teknisk rapport, Svensk försäkring, 2016.

[48] M. Turilli och L. Floridi. The ethics of information transparency. *Ethics and Information Technology*, 11(2):105–112, 2009.

[49] A. Tversky och D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.

[50] S. Umbrello. Beneficial artificial intelligence coordination by means of a value sensitive design approach. *Big Data and Cognitive Computing*, 3(1):5, 2019.

[51] A. van Rossum. Ethics, governance, trust and customer relations. *The Geneva Papers on Risk and Insurance. Issues and Practice*, 29(1):52–55, 2004.

[52] A. F. Winfield och M. Jirotka. Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133):20180085, 2018.

[53] M. Zboron. *Insurance underwriting and broking in the London insurance market: the role of reputation and trust in the insurance decision making process*. Doktorsavhandling, University of Southampton, 2015.

Appendix: Projektets resultatspridning

Vetenskaplig publicering

- J. Dexe och U. Franke. Nordic lights? National AI policies for doing well by doing good. *Journal of Cyber Policy*, 5:332–349, 2020. https://doi.org/10.1080/23738871.2020.1856160
- J. Dexe, U. Franke, A. Avatare Nöu och A. Rad. Towards Increased Transparency with Value Sensitive Design. I: *Artificial Intelligence in HCI. HCI International 2020.*, ss 3–15. Springer, juli 2020. https://doi.org/10.1007/978-3-030-50334-5_1
- J. Dexe, U. Franke och A. Rad. Transparency and insurance professionals – A study of Swedish insurance practice attitudes and future development. *The Geneva Papers on Risk and Insurance – Issues and Practice*, 2021. Under tryckning. https://doi.org/10.1057/s41288-021-00207-9
- J. Dexe, J. Ledendal och U. Franke. An empirical investigation of the right to explanation under GDPR in insurance. I: *Trust, Privacy and Security in Digital Business. The 17th International Conference on Trust, Privacy and Security in Digital Business – TrustBus 2020*. Springer, sept. 2020. https://doi.org/10.1007/978-3-030-58986-8_9

Ännu ej publicerade manuskript, under slutförfattande, granskning och revidering:

- A. Rad. Transparency for risk management: A field study of AI in the financial services.
- A. Rad. Infrastructure for Transparency.

Populärvetenskaplig publicering

- U. Franke. Towards Increased Transparency in Digital Insurance. *ERCIM NEWS*, (116), ss. 23–24, 2019. https://ercim-news.ercim.eu/en116/special/towards-increased-transparency-in-digital-insurance

Konferensdeltagande, föredrag och liknande

- Den 4–8 februari 2019 deltog Jacob Dexe i EU JRC:s *HUMAINT Winter school on AI and its ethical, legal, social and economic impact* i Sevilla.
- Den 29 mars 2019 deltog Jacob Dexe i en workshop om riktlinjer för *big data* i försäkringsbranschen på Zürichs universitet
- Den 3–4 juni deltog Ulrik Franke i *International Workshop on Cyber Insurance and Risk Controls* (CIRC 2019) på Oxfords universitet.
- Den 24 september 2019 deltog Jacob Dexe i två expertpaneler rubricerade *Algorithmic decision making* respektive *AI and insurance companies* på *Nordic Privacy Arena* i Stockholm.
- I november 2019–februari 2020 genomfördes den första omgången av doktorandkursen *Transparens i tekniska och sociala system* (7,5 hp) på KTH under ledning av Ulrik Franke.

- Den 16–17 januari 2020 deltog Alexander Rad på *Nordisk workshop i ekonomistyrning* i Stockholm och presenterade en första version av en kommande tidskriftsartikel.
- Den 29 januari 2020 höll Alexander Rad ett föredrag om transparens och AI för Governos AI-nätverk.
- Den 19 februari 2020 genomförde Jacob Dexe inom ramen för sina doktorandstudier 30%-seminarium på KTH.
- I april 2020 togs en PM till vd:ar och ordföranden i Länsförsäkringsbolagen fram i samarbete med forskningsfonden.
- Den 12 maj 2020 höll Jacob Dexe och Ulrik Franke föredrag om transparens för personal från Lantmäteriet.
- Den 1 juni 2020 publicerades, i samarbete med forskningsfonden, en populärvetenskaplig video.
- Den 15 juni 2020 höll Ulrik Franke, i samarbete med forskningsfonden, föredrag om projektets resultat för intern publik på Länsförsäkringar.
- Den 23 juli 2020 presenterade Jacob Dexe ”Towards Increased Transparency with Value Sensitive Design” vid konferensen *HCI International 2020*.
- Den 15 september 2020 presenterade Jacob Dexe ”An empirical investigation of the right to explanation under GDPR in insurance” vid konferensen *TrustBus 2020*.
- Den 17 november 2020 höll Jacob Dexe och Ulrik Franke föredrag om transparens för personal från Regeringskansliet.
- Den 30 november 2020 genomförde Jacob Dexe inom ramen för sina doktorandstudier 50%-seminarium på KTH.
- I januari 2021 publicerades en kort beskrivning av artikeln om nordiska nationella AI-strategier på den populärvetenskapliga webbplatsen *Svensk filosofi*.

Referensgruppsmöten

Projektets referensgrupp har bestått av docent Ester Appelgren (Södertörns högskola), professor Till Grüne-Yanoff (KTH), docent Stefan Larsson (Lunds universitet/FORES), Mareike Häggkvist (Länsförsäkringar) samt Peter Griepenkerl Lööf respektive Mari Sparr (representanter för forskningsfonden). Referensgruppen har sammanträtt den 19 februari 2019, den 5 november 2019 och den 23 september 2020.

Through our international collaboration programmes with academia, industry, and the public sector, we ensure the competitiveness of the Swedish business community on an international level and contribute to a sustainable society. Our 2,800 employees support and promote all manner of innovative processes, and our roughly 100 testbeds and demonstration facilities are instrumental in developing the future-proofing of products, technologies, and services. RISE Research Institutes of Sweden is fully owned by the Swedish state.

I internationell samverkan med akademi, näringsliv och offentlig sektor bidrar vi till ett konkurrenskraftigt näringsliv och ett hållbart samhälle. RISE 2 800 medarbetare driver och stöder alla typer av innovationsprocesser. Vi erbjuder ett 100-tal test- och demonstrationsmiljöer för framtidssäkra produkter, tekniker och tjänster. RISE Research Institutes of Sweden ägs av svenska staten.

RISE Research Institutes of Sweden AB
Box 1263, 164 29 KISTA
Telefon: 010-516 50 00
E-post: info@ri.se, Internet: www.ri.se

Systems engineering
RISE Rapport 2021:04
ISBN: 978-91-89167-87-2