AI i samhällets tjänst – Balansgång mellan innovation och risk inom samhällsviktiga verksamheter

Evelin Ericsson och Erika Hagberg | Avdelningen för Riskhantering och samhällssäkerhet | LTH | Lunds universitet

AI i samhällets tjänst: Balansgång mellan innovation och risk inom samhällsviktiga verksamheter

Evelin Ericsson och Erika Hagberg

Lund 2024

AI i samhällets tjänst: Balansgång mellan innovation och risk inom samhällsviktiga verksamheter

Evelin Ericsson och Erika Hagberg

Number of pages: 76
Illustrations: 11

Keywords
AI, artificial intelligence, risk, organizations, critical service, risk exposure, private sector, public sector

Abstract
This study examines various risks associated with the use of artificial intelligence (AI), which have become increasingly relevant as the technology rapidly develops and integrates into society. The risks have been categorized into six main areas: technological, data- and analytical, informational and communicational, economic, social, ethical and legal and regulatory. These areas include challenges such as technological complexity, the spread of false information, economic losses, social stress, discrimination, and difficulties navigating current legislation. What is also highlighted is the upcoming EU AI Act, which aims to regulate AI systems by categorizing them based on their level of risk.

The results from literature review, surveys, and interviews with 19 respondents knowledgeable in AI from critical societal operations show that the risks are currently low but are expected to increase. This is partly due to the increased use of advanced, generative AI technologies that are more opaque and complex.

The study also highlights similarities and differences in how the private and public sectors manage these challenges. Particularly in public operations, and to some extent in some private organizations, data often subject to stricter legal requirements is handled. This sensitive data management imposes high demands and can risk limiting how AI can be utilized, potentially creating obstacles for operations where the need for efficiency is great.

Finally, the challenge that organizations must face is emphasized as they need to balance the utilization of AI for its many benefits while also managing and minimizing the risks that the technology entails. This balancing act is a complex and crucial challenge in the development and implementation of AI technology.

Riskhantering och samhällssäkerhet
Lunds tekniska högskola
Lunds universitet
Box 118
221 00 Lund

http://www.risk.lth.se

Telefon: 046 - 222 73 60

Division of Risk Management and Societal Safety
Faculty of Engineering
Lund University
P.O. Box 118
SE-221 00 Lund
Sweden

http://www.risk.lth.se

Telephone: +46 46 222 73 60

Sammanfattning

Detta examensarbete utforskar de risker som associeras med artificiell intelligens (AI), vilka har fått stor uppmärksamhet i media det senaste året. Särskilt sedan lanseringen av ChatGPT, har mediebilden av AI både breddats och fördjupats, vilket skapat en omfattande debatt om de potentiella risker som AI kan medföra. Diskussionerna idag är huvudsakligen centrerad kring etiska risker, med särskilt fokus på hur individuella fördomar och samhällsnormer kan bli en del av AI-systemets programmering. Detta medför potentiellt en risk för diskriminering och orättvis behandling på en skala som kan överstiga tidigare teknologiska tillämpningar. Utöver dessa etiska frågeställningar har även andra riskdimensioner uppmärksammats och det är sannolikt att ytterligare risker kommer att identifieras allteftersom AI-tekniken fortsätter att avancera och integreras djupare i vårt samhälle.

Denna studie har identifierat sex huvudområden för AI-relaterade risker:

- Teknologiska, data- och analytiska risker: Dessa risker omfattar komplexiteten i AI-systemets design och funktion. Dessa risker involverar utmaningar i att förstå och hantera system som kan göra att man missbedömer och fel hanterar AI, vilket kan få negativa konsekvenser såsom spridning av data.
- Informativa och kommunikativa risker: Här ingår riskerna för hur AI kan manipulera bilder och media för att sprida falsk eller vilseledande information, vilket kan ha stora och allvarliga effekter för både enskilda individer och samhällen i stort.
- Ekonomiska risker: Dessa inkluderar de ekonomiska förluster som kan uppstå på grund av driftstopp, ineffektiv implementering av AI, samt kostnaden för förlorade möjligheter om AI inte utnyttjas på rätt sätt.
- Sociala risker: Här rör det sig om hur AI påverkar människors psykologiska och sociala välbefinnande, exempelvis genom ökad stress och osäkerhet på arbetsmarknaden samt hur teknologin kan bidra till social isolering när människor interagerar mer med maskiner än med varandra.
- Etiska risker: Dessa inkluderar frågor kring diskriminering, rättvisa, och etisk behandling av individerna som AI-systemen berör. Denna kategori är kanske den mest diskuterade givet AI-systemens kapacitet att replikera och förstärka existerande sociala fördomar.
- Juridiska och regelmässiga risker: Dessa spänner över flera av de andra riskområdena och berör de svårigheter som finns i att anpassa sig till befintlig lagstiftning, samt de framtida utmaningar som exempelvis EU:s AI Act, som ska träda i kraft om cirka två år, för med sig.

Grunden till denna studie är en litteraturgenomgång som täcker de mest diskuterade och aktuella AI-riskerna, kompletterat med en enkätstudie och intervjuer med 19 personer från samhällsviktiga verksamheter. Dessa respondenter, som antingen redan har implementerat AI eller befinner sig i början av en implementeringsprocess, ger värdefulla insikter eftersom deras verksamheter ständigt hanterar säkerhetsrelaterade frågor. Detta ger dem ett unikt perspektiv på de risker som AI kan innebära.

Resultaten från studien tyder på att även om riskerna för närvarande anses vara låga, förväntas de öka över tid, delvis på grund av att dagens AI-system fortfarande erbjuder användarna

relativt god insyn i deras funktionssätt. Framöver förväntas användningen av mer avancerade teknologier som generativ AI öka, vilket medför ökad komplexitet och minskad transparens. Dessa mer sofistikerade system innebär en minskad transparens, vilket gör dem svårare att förstå och övervaka. Detta påverkar framför allt juridiska och etiska risker, där man, på grund av låg insyn, kanske inte kommer att veta om AI:n själv sätter ihop diskriminerande attribut eller att man oavsiktligt bryter mot lagar. Studien belyser också skillnader mellan privat och offentlig sektor, där respondenter inom den offentliga sektorn bland annat uttrycker oro för att bli begränsade i användningen av AI-system på grund av lagkrav. Detta kan förhindra att dessa sektorer drar nytta av AI:s fördelar, trots att de ofta är de som mest behöver stöd, exempelvis inom hälso- och sjukvården där utmaningar med underskott av personal råder. Vidare anses även informativa och kommunikativa AI-risker leda till stor samhällspåverkan i framtiden då organisationerna måste förhålla sig till en tid när allmänheten kanske tappat tillförlitligheten till information på grund av deep fakes och liknande tekniker, vilket AI bidrar till att förstärka. Ekonomiska och sociala risker anses inte vara speciellt närvarande idag och förväntas öka måttligt i framtiden, varpå dessa risker kan anses vara lägre än de andra studerade riskerna.

Avslutningsvis är risker kopplade till AI ett komplext och mångfacetterat område som fortsätter att expandera och förändras. AI är fortfarande en relativt ny teknologi som förväntas ha en stor påverkan i samhället, vilket gör den till ett väldiskuterat ämne. Denna studie bidrar till en översikt av hur de idag mest debatterade riskerna kan finnas bland samhällsviktiga verksamheter, och understryker behovet av fortsatt forskning, samarbete och dialog för att förstå och hantera dessa risker på ett effektivt sätt.

Summary

This thesis explores the risks associated with artificial intelligence (AI), which have garnered significant attention in the media over the past year. Particularly since the launch of ChatGPT, the media portrayal of AI has both broadened and deepened, sparking extensive debate about the potential risks AI may pose. Present discussions primarily center around ethical risks, with a specific focus on how individual biases and societal norms may become embedded in AI system programming. This potentially entails a risk of discrimination and unfair treatment on a scale that may surpass previous technological applications. Beyond these ethical considerations, other dimensions of risk have also been highlighted, with additional risks likely to be identified as AI technology continues to advance and become more deeply integrated into our society.

This study has identified six main areas of AI-related risks:

- Technological, data, and analytical risks: These risks encompass the complexity of AI system design and functionality. Challenges in understanding and managing systems may lead to misjudgment and mishandling of AI, resulting in negative consequences such as data breaches.
- Informational and communicative risks: These involve the risks of how AI can manipulate images and media to spread false or misleading information, potentially having significant and serious effects on both individuals and societies at large.
- Economic risks: These include economic losses due to downtime, ineffective AI implementation, and the cost of missed opportunities if AI is not utilized properly.
- Social risks: This concerns how AI affects people's psychological and social well-being, such as increased stress and uncertainty in the job market, as well as how the technology may contribute to social isolation as people interact more with machines than with each other.
- Ethical risks: These encompass issues of discrimination, fairness, and ethical treatment of individuals affected by AI systems. This category is perhaps the most discussed given AI systems' capacity to replicate and reinforce existing social biases.
- Legal and regulatory risks: These span several of the other risk areas and address the difficulties in adapting to existing legislation, as well as the future challenges such as the EU's AI Act, which is expected to come into force in about two years.

The foundation of this study is a literature review covering the most discussed and current AI risks, supplemented by a survey and interviews with 19 individuals from essential sectors. These respondents, who either have already implemented AI or are in the early stages of implementation, provide valuable insights as their organizations consistently deal with security-related issues. This gives them a unique perspective on the risks AI may entail.

The study results suggest that while the risks are currently considered low, they are expected to increase over time, partly because today's AI systems still offer users relatively good insight into their functioning. Going forward, the use of more advanced technologies such as generative AI is expected to increase, bringing increased complexity and reduced transparency. These more sophisticated systems imply reduced transparency, making them

harder to understand and monitor. This particularly affects legal and ethical risks, where, due to low transparency, one may not know if AI itself is assembling discriminatory attributes or if unintentionally breaking laws. The study also highlights differences between the private and public sectors, where respondents in the public sector express concerns about being restricted in the use of AI systems due to legal requirements. This may prevent these sectors from benefiting from the advantages of AI, despite often being the ones most in need of support, such as in healthcare where challenges with staff shortages exist. Furthermore, informational and communicative AI risks are also expected to have significant societal impact in the future as organizations must adapt to a time when the public may have lost trust in information due to deep fakes and similar techniques, which AI contributes to reinforcing. Economic and social risks are not particularly prevalent today and are expected to increase moderately in the future, making these risks lower compared to the other studied risks.

In conclusion, AI-related risks are a complex and multifaceted area that continues to expand and evolve. AI is still a relatively new technology expected to have a significant impact on society, making it a widely debated topic. This study contributes to an overview of how the most debated risks today may exist among essential sectors and underscores the need for continued research, collaboration, and dialogue to understand and effectively manage these risks.

Förord

Detta examensarbete är den avslutande delen av civilingenjörsutbildningen i riskhantering och har utförts under vårterminen 2024 mellan januari och maj i Lund respektive Malmö. Studien har genomförts i samarbete med WSP Sverige AB för att kartlägga närvaron och exponeringen för de risker som för närvarande får mest uppmärksamhet inom området för artificiell intelligens (AI), samt andra eventuella risker som organisationerna själva upplever. Detta med syfte att fördjupa kunskapen om AI i allmänhet men i synnerhet för att lyfta hur tjänstemän inom samhällsviktiga verksamheter upplever riskerna med AI idag och i framtiden.

Ett stort tack önskas riktas till Emelie Laurin och Henrik Selin för god handledning å WSPs vägnar. Tack även till Christina Österlin och resterande Malmö-team för att vi fått vara en del av ert arbetslag.

Ett särskilt stort tack önskas riktas till Misse Wester för handledning, värdefulla insikter och vägledning.

Innehållsförteckning

1 Inledning

Artificiell intelligens (AI) har fått mycket uppmärksamhet på senare tid, och i takt med det blir tekniken även vanligare inom organisationer. Det har estimerats att mer än hälften av alla företag implementerat någon sorts AI-teknik under 2020, och det fortsätter att implementeras i hög takt (Berente et al., 2021).

Enligt Myndigheten för digital förvaltning (DIGG, 2023) har man estimerat att AI förväntas leda till en besparing om 140 miljarder kronor varje år inom den offentliga sektorn i Sverige. I december 2023 meddelades det även att regeringen skulle tillsätta en AI-kommission, i form av Finansdepartementets Dir. 2023:164, i syfte att stärka Sveriges konkurrenskraft (Regeringskansliet, 2023). Det betonas här att Sverige måste dra nytta av de betydande möjligheterna som AI erbjuder samtidigt som teknikens risker hanteras. För att vara en aktiv del i den fortsatta utvecklingen och användningen av AI krävs det att Sverige ökar sina insatser utöver nuvarande nivå (Regeringskansliet, 2023).

AI har haft positiva effekter inom den digitala arbetsmiljön genom att effektivisera och öka produktiviteten och kan frigöra tid hos anställda genom att snabbt kunna analysera stora mängder data, automatisera repetitiva uppgifter och bidra till att generera värdefulla insikter (Cipu, 2023; Alhosani & Alhashmi, 2024). Några områden där AI idag har fått genomslag i vardagen är bland annat digitala personliga assistenter och automatiska översättningar, städer och infrastruktur där AI är en del i utvecklingen av s.k. “smarta städer”, cybersäkerhet där AI bland annat bekämpar cyberangrepp, medicin och hälsa där AI kan förbättra diagnostik, tillverkning genom användning av AI-styrda robotar samt offentlig administration och tjänster där AI exempelvis kan varna för naturkatastrofer (Europaparlamentet, 2023a).

Som ovanstående redogörelse poängterar visar AI mycket potential och kommer med all sannolikhet att forma vår framtid, men i och med denna innovativa teknik har också frågor väckts kring potentiella risker för individ och samhälle. Wirtz et al. (2022) summerar de AI-risker som är de mest framträdande inom litteraturen fram till 2022 och organisaerar dessa utefter sex olika riskkategorier, 1) teknologiska, data- och analytiska, som handlar om bland annat cybersäkerhet och svårigheter i användning av AI, 2) informativa och kommunikativa, som handlar om AI:ns bidrag till spridning av desinformation, exempelvis “deep fake”-bilder, 3) ekonomiska som handlar om risken för ekonomiskt förlust, 4) sociala, där AI kan komma ersätta vissa yrken, 5) etiska, där AI riskerar diskriminera grupper eller individer samt 6) juridiska och regelmässiga AI-risker, som handlar om svårigheterna i att möta rådande och kommande AI-lagstiftning. Sedan artikeln publicerades har AI-användningen ökat globalt efter att OpenAI lanserade sin språkmodell ChatGPT i november 2022 (Zhou et al., 2024). Trots att denna teknik lanserades efter kartläggningen som gjordes av Wirtz et al. (2022) tyder senare publicerad litteratur på att dessa risker fortsätter vara de mest debatterade (Hammond, 2023; Raynovich, 2023; Marr, 2023; Cipu, 2023; Srivastava, 2023; Europaparlamentet, 2023b; Europeiska kommissionen, 2024; Campbell & Jovanovic, 2024; Nigmatov & Pradeep, 2023; Kaminski, 2024).

Verksamheter som potentiellt behöver hantera dessa risker mer än andra är de som klassats som *samhällsviktig verksamhet*. Enligt Myndigheten för skydd och beredskap (MSB, 2023) definieras en *viktig samhällsfunktion* som "en sådan samhällsfunktion som är nödvändig för samhällets grundläggande behov, värden eller säkerhet", och de verksamheter som upprätthåller och säkerställer de viktiga samhällsfunktionerna kallas för *samhällsviktiga.* Detta föranleder ett intresse att utforska hur dessa organisationer, klassade som samhällsviktiga, förhåller sig till AI-risker. Dessa verksamheter står inför striktare krav på riskhantering, kontinuitetshantering och hantering av händelser jämfört med andra organisationer. Det är även av vikt att undersöka eventuella skillnader mellan den offentliga och privata sektorn, delvis då det kan finnas vissa skillnader i hur pass lång AI-implementeringen kommit bland de två sektorerna, men framförallt då litteraturstudien indikerade att denna uppdelning är vanlig.

1.1 Syfte

Denna studie syftar till att undersöka hur exponerade samhällsviktiga verksamheterna i Sverige anser sig vara för de AI-risker som är mest framträdande inom litteraturen. Vidare utforskas andra potentiella risker kopplade till AI som dessa organisationer anser sig vara exponerade för. Ett delsyfte är också att belysa om det föreligger eventuella skillnader i riskexponering mellan privata och offentliga organisationer.

1.2 Frågeställningar

För att uppfylla syftet med studien formuleras följande forskningsfrågor:

- Hur exponerade anser samhällsviktiga verksamheter sig vara för de sex mest debatterade riskerna med AI inom litteraturen inom sin organisation idag och i framtiden?
- Föreligger skillnader mellan privat och offentlig sektor?

2 Bakgrund

Nedan behandlas en beskrivning av vad samhällsviktig verksamhet är samt en kortfattad beskrivning av vad AI är, hur det fungerar, samt hur användningen med AI ser ut i världen och Sverige idag. Utöver detta presenteras även de identifierade riskområdena med AI, för att slutligen presentera den kommande EU-lagstiftningen EU AI Act.

2.1 Samhällsviktig verksamhet

Samhällsviktiga verksamheter bedrivs av både privata och offentliga aktörer, och definieras som "verksamhet, tjänst eller infrastruktur som upprätthåller eller säkerställer samhällsfunktioner som är nödvändiga för samhällets grundläggande behov, värden eller säkerhet" (MSB, 2023). Enligt MSB (2013) anses en verksamhet vara samhällsviktig om den uppfyller minst ett av följande villkor:

- Ett bortfall av, eller en svår störning i verksamheten som ensamt eller tillsammans med motsvarande händelser i andra verksamheter på kort tid kan leda till att en allvarlig kris inträffar i samhället.
- Verksamheten är nödvändig eller mycket väsentlig för att en redan inträffad kris i samhället ska kunna hanteras så att skadeverkningarna blir så små som möjligt.

De samhällsviktiga verksamheterna måste bedriva ett systematiskt säkerhetsarbete som bygger på ansvar och samverkan mellan aktörer på olika nivåer och inom olika ansvarsområden i samhället och omfattar samtliga aktörer som äger eller bedriver samhällsviktig verksamhet, det vill säga kommuner, landsting, länsstyrelser, centrala myndigheter och privata aktörer (MSB, 2013). Vidare kräver ägande och drift av samhällsviktig verksamhet rapportering till ansvarig aktör samt förutsätter stark privat-offentlig samverkan och delaktighet i utvecklingsnätverk för att garantera verksamhetens och dess underleverantörers funktionalitet (MSB, 2013). Skyddet av samhällsviktig verksamhet kräver ett kontinuerligt och uppdaterat systematiskt säkerhetsarbete som anpassas efter samhällets föränderliga risker och utmaningar, där arbete med *riskhantering*, *kontinuitetshantering* och *hantering av händelser* för att säkerställa verksamhetens skydd ingår (MSB, 2013).

Det finns ingen myndighet som ansvarar för att utse eller sammanställa de samhällsviktiga verksamheterna i Sverige (MSB, 2023). Organisationer ansvarar därmed själva för att identifiera den samhällsviktiga verksamhet som de förser samhället med, oavsett privata eller offentliga. De offentliga aktörerna behöver identifiera samhällsviktig verksamhet inom sina respektive ansvarsområden, inklusive geografiskt område, något som inte står utskrivet för de privata aktörerna (MSB, 2023).

2.1.1 Sektorer inom samhällsviktig verksamhet

Riksdagen och regeringen tog i sin proposition *En politik för det civila samhället* (Prop. 2009/10:55) fram fem viktiga värden som ska skyddas. Bland dessa är det främst inom *samhällets funktionalitet* som samhällsviktig verksamhet lyfts (MSB, 2023).

MSB har pekat ut 10 sektorer som ska upprätthållas och som är av prioritet gällande civilt försvar, och av dessa har följande sju samhällssektorer valts ut: 1) energiförsörjning, 2) livsmedel och dricksvatten, 3) information och kommunikation, 4) finansiella tjänster, 5) skydd och säkerhet, 6) transporter och 7) hälso- och sjukvård (MSB, 2020). Mer ingående förklaring och exempel inom dessa områden återfinns i Bilaga A.

2.2 Vad är AI?

Själva termen "artificiell intelligens" myntades av John McCarthy år 1956 (McGuire et al., 2006; Council of Europe, u.å). Mellan 50- och 70-talet såg framtiden ljus ut för artificiell intelligens och många trodde att teknologin inom kort skulle kunna utföra uppgifter lika bra som människor. Bristande finansiella tillgångar och framförallt begränsningar av dåtidens datorer, där kostnad att tillverka och bristande förmåga att lagra data och processa data tillräckligt snabbt, hindrade däremot den utvecklingen (Council of Europe, u.å; Rockwell, 2017; Brooks et al., 2016). Under 1990- och 2000-talet väcktes det allmänna intresset igen, framförallt eftersom datorer blev billigare, bättre hårdvara utvecklades och datorerna klarade av att lagra större mängder data (Brooks et al., 2016). I samband med internets utveckling som möjliggjort enorm tillgänglighet till information, tillsammans med datorers ytterligare ökade förmåga för datalagring och bearbetning, har AI-teknologier blivit allt mer avancerade och används i allt större utsträckning världen över (Rockwell, 2017).

AI medför även utmaningar, som ofta uppstår i diskussionen kring dess definition (Nolan et al., 2024; Goodson, 2021; O'Shaughnessy, 2022; Casey & Lemley, 2020). Det har konstaterats att beslutsfattare tenderar att jämföra AI med mänskliga beteenden, medan forskare inom AI fokuserar på mer tekniska definitioner som betonar dess funktionella aspekter (Krafft et al., 2019). Enligt Krafft et al. (2019) speglar de tekniska definitionerna mer korrekt AI:s nuvarande användning och kapaciteter, medan en syn på AI som om den hade mänskliga drag mer överensstämmer med framtida förväntningar på tekniken. Författarna argumenterar även för att skapa en bättre överensstämmelse mellan forskningens och politikens perspektiv, vilket är avgörande för att framgångsrikt navigera i AI:s utmaningar (Kraft et al., 2019). Casey och Lemley (2020) menar att det troligen inte finns en "rätt" definition och att man vid försök att skapa sådan kan riskera göra denna för bred eller för snäv, och därmed exkluderar viktiga aspekter eller att den efter kort tid blir irrelevant på grund av AI:ns snabba tekniska utveckling.

Enligt ovan redogörelse är en gemensam definition alltså önskvärd, och EU har påbörjat detta arbete genom ett lagförslag som lades fram 2021, kallat EU AI Act, som nu är i slutstadiet då de 27 medlemsstaterna enhälligt godkänt AI-akten (Europeiska kommissionen, 2024a & 2024b). Detta arbete utgår från definitionen av ett AI-system ur EU Artificial Intelligence Act (EU AI Act, 2024a) som efter översättning utifrån bästa förmåga lyder:

"AI-system är ett maskinbaserat system som är utformat för att fungera med varierande grad av autonomi och som kan uppvisa anpassningsförmåga efter driftsättning och som, för uttryckliga eller underförstådda mål, från den indata det tar emot drar slutsatser om hur man ska generera utdata såsom prognoser, innehåll, rekommendationer eller beslut som kan påverka fysiska eller virtuella miljöer."

2.3 Hur är AI uppbyggt?

För att beskriva hur AI är uppbyggt bör utvecklingen inom programmering och datorteknik studeras. I traditionella datorprogram är det en programmerares ansvar att definiera programmets funktioner genom att skriva koder, också kallat algoritmer, som specificerar exakt hur en uppgift ska utföras, inklusive vilka steg som ska tas och i vilken sekvens de bör genomföras (Bäck, 2023). *Maskininlärning* (eng. machine learning) (ML) är sådana avancerade algoritmer och används ofta synonymt med AI, dock finns det vissa skillnader (Coursera, 2024). AI syftar huvudsakligen på det övergripande konceptet att skapa människoliknande tankeprocesser med hjälp av datorprogram, medan maskininlärning bara är en metod för att uppnå detta och således utgör ett delområde inom AI (Coursera, 2024; Holzinger et al., 2018).

Deep learning (DL), eller *djupa neurala nätverk*, är ett delområde inom ML som bygger på neurala nätverk (NN) och som är en essentiell del vad gäller skapande av AI-system (Brynjolfsson & McAfee, 2019; V. Joshi, 2024; Holzinger et al., 2018). De neurala nätverken består av lager av små enheter (neuroner) som bearbetar information där varje neuron tar emot information, bearbetar denna, skickar vidare till nästa neuron som gör samma sak tills dess att alla neuroner i nätverket bearbetat informationen (DIGG, 2024). Att skilja på DL och ML kan vara svårt eftersom båda metoderna/systemen kan använda sig av NN, men den största skillnaden ligger i typen av NN och graden av mänsklig inverkan (IBM, u.å.). Klassisk ML är ofta *övervakad*, där modellen tränas på data som blivit annoterad och därmed känner man till vad datan visar (DIGG, 2024; V. Joshi, 2024). DL är oftare *oövervakad*, där träningsdata inte är annoterad och modellen skapar struktur och mönster på egen hand, och använder flera djupare lager av neuroner (IBM, u.å.; Lunds universitet, 2020). Många populära AI-system som finns idag, bland annat ChatGPT, är en DL-modell som ofta benämns generativ AI som är en typ av AI-teknik som kan lära sig från omfattande datamängder. Det intressanta är att denna teknik kan skapa helt *nya* objekt baserat på det den har lärt sig, och detta samtidigt som den behåller vissa likheter med ursprungsdatan (Sasaki, 2023).

2.4 Användningen av AI

Nedan presenteras statistisk information om användningen och förekomsten av AI, med en kort översikt av globala trender och en djupare översikt över användning i Sverige.

2.4.1 AI globalt

Organisationen för ekonomiskt samarbete och utveckling (OECD, 2023) beskriver att eftersom AI kan tillämpas på en mängd olika sätt och är under ständig utveckling så finns det vissa utmaningar vid försök att mäta dess användning. Detta gör bedömningen och mätningen av vilka länder som ligger i framkant komplicerad och beroende av flera faktorer. OECD (2024) har valt att mäta AI-användning bland länder baserat på faktorer som utbildning, forskning och investeringar i AI, och baserat på dessa kriterier är framförallt USA, EU och Kina ledande aktörer.

Användningen av AI bland företag i EU beskrivs av Eurostat (2021), som uppger att ungefär 8 % av organisationerna integrerat AI i sina verksamheter. Danmark utmärker sig som

ledande, där 24 % av företagen har antagit AI-teknologier, och Sverige presterar marginellt bättre än genomsnittet (8 %) med en införandenivå på 9 % (Eurostat, 2021).

2.4.2 AI i Sverige

Statistiska centralbyrån (SCB, 2023) har i en rapport fokuserat på upptaget av AI-teknik i den offentliga och privata sektorn från 2019-2021, där en skillnad mellan den offentliga och privata sektorn redovisas, varav 29,7 % av de större bolagen inom privata sektorn angav att de använde AI i sin verksamhet, medan motsvarande siffra inom offentlig sektor var 26,6 %. Det syns även en ökning av användningen framförallt hos mindre och medelstora organisationer inom företagssektorn, men högst andel som använder AI fortsätter vara de större organisationerna (SCB, 2023).

SCB (2023) rapporterar att företag inom privat och offentlig sektor svarade på deras AI-användning utefter syfte och från dessa mätningar påvisas att AI främst används till att utveckla eller förbättra interna processer samt förbättra en existerande produkt eller tjänst.

Utmaningar med implementering av AI

Statskontoret (2024) rapporterar om användningen av AI bland statliga myndigheter där nästan alla respondenter lyfte kompetensbrist, både teknisk och juridisk, men även kunskap om AI-utveckling tillsammans med kunskap om verksamheten, som ett hinder för användning av AI. Den offentliga sektorn rapporterar att faktorer som kunskap om teknologin och tillämpningar samt anställdas kompetens, utbildning eller erfarenhet utgjorde de största hindren för AI-implementering (SCB, 2023). Inom privata sektorn var det främst avsaknad av relevant expertis på företaget som utgjorde det största hindret. Data för privata företag som använder AI från 2021 återges ej i rapporten, men år 2019 där företag som använt AI fick svara angav majoriteten att anställdas kompetens, utbildning eller erfarenhet var det största hindret (SCB, 2023).

2.4.3 Incidenter med AI

Det har rapporterats i media om AI-orsakande incidenter. Stora företag som Samsung upptäckte att personal råkat läcka känslig data till ChatGPT, varpå de förbjöd användandet av det inom verksamheten, och OpenAI har själva poängterat att integriteten inte är säkrad när sådant sker (Marr, 2023). Ett AI-system som implementerats på Amazon med syfte att ge arbetssökande kandidater betyg genom att granska jobbansökningar sållade bort kvinnliga sökande då träningsdatan som lämnats in av sökande under en 10-årsperiod bestod av män (BBC, 2018). Andra AI-program såsom Googles Bard AI fabricerade själv ihop fakta, ett fenomen kallat “hallucinationer” där AI skapar eller presenterar felaktig eller påhittad information som om den vore sann (Raynovich, 2023).

2.5 Risker med AI

Det finns flera faktorer som gör det utmanande att helt förstå riskerna med AI. Systemen består av olika delar som måste samverka såsom algoritmer, data, hård- och mjukvara vilket gör dem komplexa samt svårt att förutse alla potentiella konsekvenser, särskilt om det är en ny tillämpning som ännu inte har testats i verkligheten (Tartaro et al., 2024).

Wirtz et al. (2022) publicerade ett risk- och riktlinjebaserat ramverk i februari 2022 med en omfattande litteraturöversikt av de AI-risker som var mest diskuterade fram till studiens publiceringsdatum. Studiens syfte var att utveckla ett ramverk baserat på risker och riktlinjer för styrning av AI inom den offentliga sektorn. Arbetet resulterade i att sex riskområden identifierades: (1) *teknologiska, data- och analytiska*, (2) *informativa och kommunikativa*, (3) *ekonomiska*, (4) *sociala*, (5) *etiska*, samt (6) *juridiska och regelmässiga* AI-risker (Wirtz et al., 2022). Författarna beskriver hur etiska risker är väl representerade inom forskningen och dominerar områden vad gäller risker med AI samt att risker inom områdena informations- och kommunikationsrelaterade risker samt ekonomiska AI-risker inte är lika väl studerade och att det finns kunskapsbrist.

Följande avsnitt bygger på riskkategoriseringen föreslagen av Wirtz et al. (2022), men har berikats med ytterligare perspektiv och infallsvinklar. Särskild uppmärksamhet har ägnats åt litteratur publicerad efter artikeln av Wirtz et al. (2022) med ett fokus på studier som framträder i samband med den betydande utvecklingen inom AI-forskningen, speciellt efter lanseringen av OpenAI:s ChatGPT i november 2022 (Zhou et al., 2024). Många av dessa risker är sammanflätade och påverkar varandra både direkt och indirekt. Trots deras inbördes samband har en uppdelning gjorts för att ge en strukturerad överblick över detta komplexa område. Nedan presenteras de enligt hur omskrivna de är enligt Wirtz et al. (2022).

2.5.1 Teknologiska, data- och analytiska

Den första kategorin berör främst två områden enligt Wirtz et al (2022), där det första gäller den potentiella förlusten av kontroll över teknologin, med särskilt fokus på automatiserade beslutsprocesser där mänskligt inflytande inte existerar. Den andra gäller brist på teknologisk (expert-)kunskap alternativt den ökade komplexiteten och/eller black-box-problematiken, som innebär att vi inte kan se eller förstå hur maskinen tänker eller fattar beslut, som kan leda till oönskade konsekvenser.

AI får alltmer mandat för beslutsfattande inom organisationer och institutioner, trots att människor inte har full insikt i vad AI baserar sina beslut på (Berente et al., 2021). Detta fenomen har exempelvis dykt upp inom sjukvården, där AI kan användas för att diagnostisera patienter eller rekommendera vård i form av läkemedel, och AI har visat sig rekommendera något som faktiskt funkar för patienter, trots att man inte förstår hur AI:n kommit fram till det (Chan, 2023). Detta illustrerar hur AI blir som en black box där även AI-experter får svårt att veta hur AI:n kommit fram till beslut, då systemen blivit allt för komplexa och saknar transparens (Raynovich, 2023; Europeiska kommissionen, 2020; Du & Yuan, 2022; Statskontoret, 2024; Brynjolfsson & McAfee, 2019). Detta kan göra det svårt att verifiera överensstämmelse med och hindra effektivt genomförande av regler inom befintlig EU-lagstiftning avsedda att skydda grundläggande rättigheter (Europeiska kommissionen, 2020). Denna brist på transparens kan leda till misstro bland användare och intressenter, och organisationer bör prioritera transparens genom att utforma AI-modeller och algoritmer som ger insikt i deras beslutsprocesser. Transparent AI ökar det övergripande förtroendet mellan parterna och deras beslut och underlättar efterlevnad av regelverk (Srivastava, 2024).

AI kräver ofta insamling och analys av stora mängder (ibland personlig) data, vilket väcker integritets- och säkerhetsfrågor (Srivastava, 2024). Dessa risker kan finnas både internt och externt, där internt exempelvis kan vara anställdas felaktiga användning av AI-tjänster där de av misstag läcker företagskänslig data (Marr, 2023). Ett exempel på detta är implementeringen av generativ AI på arbetsplatsen som introducerar flera möjligheter. Att inte använda tekniken kan leda till att man hamnar efter i förhållande till konkurrenter men företag måste se över så att anställda inte trillar ner i fallgropar med dess tillämpning där företagen måste ta itu med upphovsrätt, dataläckage och säkerhetsrisker (Campbell & Jovanovic, 2024). Termen "Shadow AI" innebär oauktoriserad användning av generativ AI utanför IT-styrningen inom en organisation, och många företag är idag omedvetna om vilka AI-verktyg som anställda använder och vilka risker detta innefattar (Campbell & Jovanovic, 2024).

Vid användning av AI finns det även risk att man förlitar sig för mycket på, och därmed övervärderar, resultaten som AI tar fram då det finns generella attityder att AI är smartare eller bättre än människan inom vissa områden, eftersom det baseras på objektiva och analytiska uppgifter (Keding & Meissner, 2021). I detta finns även risken att AI "förmänskligas", och att systemet därmed får högre tillit än vad det egentligen bör ha. Prest (2023) uttrycker detta som att "Risken är inte att maskinerna börjar tänka, utan att vi slutar göra det". En liknande risk är beroendet av AI där människans överdrivna tilltro till systemen kan lämna människor utan intuition, kreativitet och kritiskt tänkande (Raynovich, 2023).

Externa AI-risker är då angripare potentiellt kan skada organisationer genom användning av AI (Marr, 2023). Dessa AI-system kan utgöra cybersäkerhetsrisker och göra företag mer sårbara för cyberattacker och dataläckage (Nigmatov & Pradeep, 2023). AI är en allmän, dual-use[1] teknologi, som används med allt mer intuitiva gränssnitt (Stockholm universitet, 2023). Den användarvänlighet som dessa teknologier erbjuder gör dem tillgängliga för en bred publik, inklusive de som har onda avsikter (World Economic Forum, 2023). I takt med att teknologin utvecklas och intrång blir allt enklare, ökar även behovet för företag att hålla sig uppdaterade med den senaste tekniken för att kunna hantera säkerhetsriskerna (Srivastava, 2024; Campbell & Jovanovic, 2024).

2.5.2 Informativa och kommunikativa

Den andra riskkategorin fokuserar på risken för AI-manipulerad information, och Wirtz et al. (2022) delar in denna risk efter två huvudområden. Den ena gäller risker med den information vi får till oss som filtrerats genom AI-system och den andra risken handlar om att AI, genom att anpassad och personifierad information, manipulerar och provocerar/uppmanar individer genom AI-genererad målinriktad informationsförsörjning, även kallat "riktad censur" (Wirtz et al., 2022). AI har utvecklats till den grad att det nu kan skapa medieinnehåll som är så realistiskt att det blir mycket svårt att avgöra om det är äkta, så kallade "deep fakes" (Hammond 2023; Srivastava, 2024). Denna risk är särskilt påtaglig med tanke på den utbredda tillgängligheten av AI idag, vilket möjliggör för externa aktörer, som avsiktligt vill

[1] Dual-use produkter är produkter som kan användas både av civila och militära (Stockholm universitet, 2023)

skada eller gynna sig själva, att relativt enkelt lura och sprida falsk information (World Economic Forum, 2023).

2.5.3 Ekonomiska

Forskning inom denna riskkategorin syftar på risker där AI orsakar störningar i ekonomiska system, i synnerhet arbetsmarknaden, i form av att skatteintäkter från arbetare som ersatts av AI uteblir (Wirtz et al., 2022). Här inkluderas även organisatoriska risker, såsom brist på AI-talang/-strategi, eller ersättning av mänsklig arbetskraft (Wirtz et al., 2022) som författarna för denna studie istället valt att lägga under sociala risker. Inom litteraturen beskrivs de ekonomiska riskerna med AI framförallt som risker att hamna efter i utvecklingen, eller att underanvända AI, och därmed gå miste om ekonomiska fördelar som AI:n kommer att bistå med (Europaparlamentet, 2023b; Cipu, 2023). Vidare lyfts också hur organisationer i iver att dra nytta av teknikens många fördelar övernyttjar AI vilken kan leda till ekonomiska förluster till följd av kostnader av implementering och utbildning där avkastningen misslyckats eftersom de kostade mer än de producerade (Raynovich, 2023; Europaparlamentet, 2023b). Att automatisera uppgifter genom generativ AI kan också göra företag sårbara där användning av stora språkmodeller kan medföra operativa risker om systemen skulle felfungera eller fallera vilket leder till ekonomisk förlust (Campbell & Jovanovic, 2024).

2.5.4 Sociala

Den fjärde riskkategorin fokuserar på arbetslöshet till följd av AI-automatiserade processer och motståndet som uppstår mot AI bland anställda till följd av detta (Wirtz et al., 2022). Utöver detta är även associerade sociala konsekvenser, såväl som på integritets- och säkerhetshot eller oro i samhället till följd av integrering av AI-system (Wirtz et al., 2022; Hammond, 2023). Yrkesgrupper som jurister, textförfattare och kodare har yttrat oro för sin framtid (Hammond, 2023), och oron är även större bland yrken inom branscher mottagliga för automation, såsom tillverkning och transport (Nigmatov & Pradeep, 2023).

Danielsen (2023) menar att den emotionella och existentiella risk som uppstår i samband med att människors arbetsuppgifter blir ersatta av AI-system inte uppmärksammas tillräckligt mycket. Om AI-implementering leder till att vissa yrken blir ersatta kan känslor av hopplöshet, osäkerhet, likgiltighet och meningslöshet uppstå (Danielsen, 2023). En bank i Tyskland valde att ersätta personal med ett AI-system för rådgivning av kunder samt godkännande eller nekande av låneansökningar, som var mer effektivt än vad människor är kapabla till, och ansågs även minska risken för mänskliga fel (Mayer et al., 2020). Vid utvärderingen visades att AI-systemet uppfyllde sitt avsedda syfte, men också skapade en organisatorisk bieffekt, då en känsla av meningslöshet bland de anställda uppstått (Mayer et al., 2020). I och med att den förväntade ökningen av AI-implementeringar inom diverse branscher och organisationer, så lär emotionell och existentiell risk öka, något som chefer i framtiden bör ta hänsyn till (Danielsen, 2023).

Vidare kan denna tekniska utveckling leda till mindre interaktion människor emellan och ökad isolering bland anställda, som i sin tur leder till minskad empati och sociala färdigheter (Srivastava, 2024). Även om AI förväntas skapa många nya och förbättrade arbetstillfällen,

kommer utbildning och praktik att vara viktiga för att förhindra att människor förlorar sina jobb och hamnar i långvarig arbetslöshet (Europaparlamentet, 2023b).

2.5.5 Etiska

Inom den femte riskkategorin fokuserar Wirtz et al. (2022) främst på två faktorer, där den första är risken för bristen på eller felaktig integration av mänskliga värderingar i AI:s beslutsfattande och handlingar, då AI-system kan sakna en etisk grund. Den andra är de negativa konsekvenser som följer av det första fokusområdet, nämligen AI-baserad diskriminering som reproduceras av att människor, som har förutfattade meningar, har programmerats in AI-systemet. Europaparlamentet (2023b) lyfter också dessa risker med AI då dess algoritmer kan vara programmerade med fördomar som leder till diskriminering vid exempelvis rektyteringar och låneansökningar. De lyfter även hur AI kan användas för bildigenkänning och därmed inkräkta på integritet och dataskydd och hur organisations- och demonstrationsfrihet genom detta kan påverkas då den kan profilera och spåra individer kopplade till specifika övertygelser eller aktioner. Med generativ AI finns även risken att applikationer skapas med inbäddade fördomar och begränsad insyn i AI:ns beslutsfattandeprocess (Campbell & Jovanović, 2024).

Mänskliga beslut är inte immuna mot misstag samt fördomar och diskriminering utgör inneboende risker med alla samhälls- och ekonomiska aktiviteter. Dock kan samma fördomar få en mycket större effekt i AI-system och påverka många människor utan de sociala kontrollmekanismer som styr mänskligt beteende (Europeiska kommissionen, 2020). Detta kan också inträffa när AI-systemet lär sig under drift. I sådana fall, där utfallet inte kunde ha förutsetts eller förhindrats vid designfasen, kommer riskerna inte att härledas från ett fel i det ursprungliga systemets design utan snarare från de praktiska effekterna av de korrelationer eller mönster som systemet identifierar i ett stort dataset (Europeiska kommissionen, 2020).

Vidare kvarstår problemen att AI:n blir inte bättre än den data som den tränats på, och innehåller datan fördomar kommer AI:n lära sig detta. Detta kan orsaka problem om AI används för beslutsfattande om anställning, befordringar, beviljning av lån etc. som kan leda till skadat rykte och negativa konsekvenser för organisationer (Nigmatov & Pradeep, 2023). AI:n kan bli orättvis då det finns risk att de fördomar vi människor har som finns i trändningsdatan matas in i modellen som skadar särskilda samhällsklasser och därmed utsätter företaget för orättvisa risker och ansvar (Buehler et al., 2021). Inom den offentliga sektorn har potentiella risker som fördomar i träningsdatan, replikation av mänskliga fel, systematisk diskriminering lyfts och att utmaningar kommer finnas kopplat till att identifiera eventuella inbäddade fördomar som finns (Mellouli et al., 2024).

Du och Yuan (2024) har identifierat två etiska AI-risker i form av medveten eller omedveten diskriminering, där algoritmer kan undvika avsiktlig diskriminering, men fortfarande skapa omedveten diskriminering baserat på de datan som används, och därmed kan AI-algoritmer förvärra befintliga ojämlikheter. Artikeln diskuterar hur forskare hävdat att det är möjligt att tekniskt eliminera kulturella fördomar från ML, men att det finns potentiella risker och svårighet kopplat till komplexitet med att integrera AI i samhället, samt hur det är svårt att lagstiftningsmässigt reglera detta (Du & Yuan, 2024).

Enligt Statskontoret (2024) och Srivastava (2024) kommer AI:n kunna motverka fördomar samt att företag som implementerat sådana system menar att detta leder till ökad objektivitet och mindre fördomar än vad människor har.

2.5.6 Juridiska och regelmässiga

Den sjätte riskkategorin innefattar risken för oklar ansvarsskyldighet och redovisning vid AI-orsakade fel eller negativa konsekvenser, samt bristen på explicita regleringar och styrning kopplat till detta (Wirtz et al., 2022). Implementering av AI på arbetsplatser ställer nya krav och utmaningar för arbetsgivare att följa lagliga ramar för dataskydd, immateriell egendom och anställningsrätt för att säkerställa ansvarsfull och etisk användning av AI (Cipu, 2023). Centralt för användning av AI är data, och utöver att följa lagstiftningar kring detta måste organisationer också leva upp till konsumenters förväntningar kring hur personuppgifter ska hanteras, där försummelse av detta kan leda till brustet förtroende (Buehler et al., 2021). Medborgare kan ha svårt att förstå automatiserade processer och kan uppleva brist på insyn, särskilt när AI skyddas av patent och rättigheter, vilket kan skada förtroendet ytterligare (Statskontoret, 2024)

Frågor gällande ansvar och rättigheter blir mer komplexa med AI, bland annat frågan hur misstag orsakade av AI:n ska hanteras, och detta blir framförallt svårt då data är svårt att spåra (exempel vid hallucinationer) (Raynovich, 2023). Vidare uppstår svårigheter i hur man ska hantera “immateriell egendom” där AI exempelvis genererar ny media baserat på konst, musik och verk som skapats av andra (Raynovich, 2023).

Att veta om man följer rådande lagstiftning är också en utmaning då det gäller AI, och som tidigare nämnt kan AI:n på grund av dess komplexitet göra det svårt att verifiera att den följer lagstiftningen (Europeiska kommissionen, 2020). Tillsynsmyndigheter kan även bli osäkra på om de ska ingripa då de saknar teknisk förmåga att granska systemen eller är osäker på deras befogenhet i situationen (Europeiska kommissionen, 2020).

Vidare kan personer som lidit skada på grund av AI-system ha svårt att säkerställa bevisning och att bygga fall i domstol (Europeiska kommissionen, 2020). Ett problem vid användandet av AI är att det inte riktigt finns tydliga riktlinjer eller lagar som effektivt kan hantera då något går fel, något som märktes i Göteborgs stad där AI-algoritmer användes för att placera barn i skolan (Loudiyi, 2021). Algoritmen placerade barnen baserat på närmsta fågelvägen och missade därmed att ta hänsyn till stadens infrastruktur, där stora områden gränsas av kanaler, vilket ledde till att vissa barn fick lång pendling till skolan (Loudiyi, 2021). En doktorand inom digitalisering och juridik, valde att stämma kommunen för detta men förlorade (Sisask, 2023). Forskning gällande detta ämnet pågår fortsatt och det finns idag en brist på kunskap och förståelse hur det juridiska rättsväsendet ska bemöta tvister där AI ligger bakom avvikelser och/eller fel (Jarvenpaa et al., 2023; Berggren et al., 2023).

Implementering av AI inför således flera tekniska, organisatoriska och regleringsmässiga utmaningar. Befintliga lagar som styr dataskydd och säkerhet, såsom GDPR i EU och CCPA, begränsar hantering, användning och lagring av data, och med AI introduseras ytterligare en

dimension där företag måste beakta de etiska och sociala aspekterna av AI, inklusive risken för fördomar och diskriminering (Nigmatov & Pradeep, 2023).

2.5.7 Övriga risker kopplade till AI

Utöver de sex riskkategorierna har några risker varit framträdande inom litteraturen. En "modernare" risk är det miljömässiga avtryck som AI innebär, och Tartaro et al. (2024) menar att en viktig lärdom denna kategorisering tyder på är att organisationer inte bara kan tänka på organisatoriska risker (exempelvis kopplat till finans eller rykte), utan också måste ha miljömässiga, samhälleliga och etiska risker i åtanke. Sætra och Danaher (2023) lyfter en dystopisk risk där man menar att AI kommer att utrota mänskligheten, som är mycket i linje med en varning som Center for AI Safety (CAIS, u.å.) publicerat, nämligen:

> *"Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war"*

Denna oro för utrotning grundar sig delvis i idén att AI, om denna blir mer intelligent och mer kraftfull än oss människor, bestämmer sig för att avsluta våra liv (Raynovich, 2023; Du & Yuan, 2022; Srivastava, 2024). Enligt Bales et al. (2024) kan denna dystopiska risk med AI delas in i två kategorier, maktbegär och singularitetshypotesen. Den förstnämnda innebär att AI:n kommer vilja ha makt och därigenom hamna i konflikt med mänskligheten över resurser och inflytande. Den andra kategorin är en hypotes där AI:n blir så pass intelligent att den, om den inte är i symbios med mänskliga värderingar, kan utgöra ett hot mot mänskligheten (Bales et al., 2024).

En annan risk som också diskuteras är ojämn maktbalans där risken finns att några få stora företag med fler tillgångar får kontroll över AI-utvecklingen och därmed potentiellt förvärrar ekonomisk ojämlikhet till att gynna rika personer och stora företag (Raynovich, 2023). Författaren menar vidare att en "kapplöpning" om AI kan göra att tekniken utvecklas för snabbt. Detta kan också göra att större aktörer som sitter på mer data och information får ett övertag och slår ut konkurrenter (Europaparlamentet, 2023b). Företag kan också bli sårbara då de kan behöva externa aktörer för att bygga AI-modeller, vilket innebär outsourcing av exempelvis datainsamling och behöver känna till och förstå de riskhanterings- och styrningsstandarder som tillämpas av varje tredje part (Buehler et al., 2021).

2.6 EU AI Act

EU-kommissionen har föreslagit ett första rättsligt ramverk, med syfte att adressera de risker som finns med AI samt att se till att Europa blir ledande globalt inom AI-utveckling (Europeiska kommissionen, 2024). Efter att ha nått överenskommelse mellan Europeiska rådet och parlamentet i december 2023, förväntas den slutgiltiga texten träda i kraft år 2026 (Europeiska rådet, 2024). AI-lagen inför en klassificering av AI-system baserat på risknivå och delar in dessa i fyra kategorier: oacceptabel risk, hög risk, begränsad risk och minimal risk (EU AI Act, 2024b).

Oacceptabel risk beskrivs som AI-system som utgör hot mot människor och dessa kommer att förbjudas (Europaparlamentet, 2023c). Detta innefattar AI som används för biometrisk identifiering, manipulering av människor exempelvis via leksaker som är röststyrda samt användning av AI för social poängsättning. Undantag finns från några dessa typer av AI-användning, exempelvis biometrisk identifikation som endast får användas för brottsbekämpning (Europaparlamentet, 2023c).

Högrisksystem beskrivs som sådan AI som kan ha negativa effekter på säkerhet eller mänskliga rättigheter (Europaparlamentet, 2023c). Denna fördelas på två kategorier:
1) AI-system som används i produkter som omfattas av EU:s produktsäkerhetslagstiftning och
2) AI-system som faller under specifika områden som kräver registrering i EU-databas, såsom:

- Hantering och drift av kritisk infrastruktur
- Utbildning och yrkesutbildning
- Anställning, arbetarhantering och tillgång till egenföretagande
- Tillgång till och nyttjande av väsentliga privata tjänster samt offentliga tjänster och förmåner
- Brottsbekämpning
- Migrations-, asyl- och gränskontrollhantering
- Assistans vid juridisk tolkning och tillämpning av lagar

Den största delen av lagstiftningen rör klassificeringen och regleringen av dessa högrisksystem och aktörers vars AI faller under denna kategori kommer bli tvungna att följa vissa krav på riskhantering, dataskyddshantering, dokumentation, instruktioner för användning av AI-system, säkerställa mänsklig översyn och lämpliga nivå av noggrannhet, robusthet och cybersäkerhet samt inrätta ett kvalitetsledningssystem (EU AI Act, 2024b).

Vissa AI-system som faller under kategorin begränsad risk kommer ha krav om transparens och AI-system i kategorin minimal risk är oreglerade, vilket är majoriteten av de system som för närvarande finns tillgängliga på EU:s inre marknad, exempelvis AI i videospel och skräppostfilter (EU AI Act, 2024b).

De Cooman (2022) kritiserar valen av riskkategorisering som föreslagits, där framförallt högrisksystemets definition kritiseras för att vara alldeles för komplex och att AI-system som inte klassas som högrisk ändå riskerar vara det. Wörsdörfer (2023) analyserar kritiskt lagstiftningen ur ett affärs- och datoretiskt perspektiv och utvärderar dess styrkor och svagheter. Styrkor som lyfts är bland annat lagstiftningens förmåga att hantera risker med datakvalitet och diskriminering, samt institutionella innovationer. Brister som lyfts är bland annat en ineffektiv mekanism för att säkerställa efterlevnad av AI-förordningen samtidigt som skydd mot potentiella negativa konsekvenser säkerställs, bristen på demokratisk och rättslig tillsyn, bristen på tillräckligt arbetarskydd samt bristen på adekvat finansiering och bemanning (Wörsdörfer, 2023).

3 Metod

För bättre förståelse av kontexten genomfördes en litteraturstudie som sedan låg till grund för utformning av relevanta enkät- och intervjufrågor. För att svara på frågeställningarna gjordes en enkät- och intervjustudie där respondenterna först fick svara på enkätfrågor som sedan låg till grund för intervjuerna, där respondenterna ombads utveckla sin gradering gällande exponeringen av riskerna i enkäten. Enkäter med frågor som bedöms numeriskt faller under kategorin kvantitativ forskningsstrategi (Ponto, 2015), och intervjuer med syfte att utveckla de besvarade frågorna från enkäten faller under kategorin kvalitativ forskningsstrategi (Jamshed, 2014).

Intervjuerna utformades som en kombination av öppna, allmänna och mer riktade, snäva frågor, även kallat en halvstrukturerad intervju (Höst et al., 2006). Intervjuerna utfördes i fyra faser: sammanhang, inledande frågor, huvudfrågor och sammanfattning. Sammanhang inkluderar beskrivning av studiens syfte, varför personen blivit utvald, hantering och bearbetning av insamlad data, samt samtycke angående inspelning och deltagande (Höst et al., 2006). Därefter följde inledande frågorna med mer grundliga och neutrala om exempelvis arbetstitel samt nuvarande och planerad AI-användning, följt av huvudfrågorna där "tyngden" av intervjun vägs in. I sammanfattningen summerades intervjun kortfattat samt öppnade upp för frågor eller noteringar från respondenten, samt att formaliteter som intervjuns förutsättningar samt hur personen eventuellt kan dra tillbaka sitt deltagande avhandlades (Höst et al., 2006)

Användandet av både enkäter och intervjuer för datainsamling faller under forskningsstrategin som kallas för en mixad metod (eng. mixed methods), där element från både kvalitativa- och kvantitativa forskningsstrategier kombineras för att bredda och fördjupa förståelsen av ämnet (Schoonenboom & Johnson, 2017).

Då det noterades i ett tidigt skede att många respondenter upplever AI som ett känsligt ämne beslutades att samtliga respondenter skulle vara anonyma. För att tillgodose detta har en utformning av koder, dokumentation och hantering av data skett för att säkerställa respondenternas anonymitet. Vidare har citat valts med omsorg för att inte kunna spåras tillbaka till specifika respondenter.

3.1 Litteraturstudie

Litteraturstudien syftade till att identifiera de mest diskuterade AI-riskerna för att kunna formulera relevanta frågor för enkäter och intervjuer. Sökningen genomfördes huvudsakligen via universitetets bibliotekskatalog samt Google Scholar. Ytterligare källor som myndighetsrapporter och populärvetenskapliga artiklar användes också.

Riskkategoriseringen (beskriven i avsnitt 2.5) är till stor del baserad på Wirtz et al. (2022). För en fördjupad förståelse inom dessa kategorier har kompletterande information hämtats från andra källor. Fokus har främst riktats mot artiklar publicerade efter januari 2022 för att säkerställa att de senaste riskerna relaterat till AI-teknologin. Ett antagande som gjorts är att

det har publicerats mer litteratur om risker med AI under 2023 och 2024, delvis på grund av lanseringen av ChatGPT. Baserat på denna litteraturstudie kompletterades de 6 riskområdena ursprungligen från Wirtz et al. (2022) med ytterligare källor och den slutliga klassificeringen och samtliga källor till riskkategorierna återfinns i Bilaga B.

3.2 Urval

Då studiens datainsamling bland annat utgjordes av kvalitativa intervjuer har urvalet gjorts systematiskt utifrån teoretiskt och strategiskt definierade kriterier (Holme & Solvang, 1997). En utmaning var att ta reda på vilka verksamheter som klassas som samhällsviktiga, eftersom denna information inte finns tillgänglig, utan måste inhämtas via kontakt med organisationerna. Därmed inleddes urvalsprocessen med att först identifiera vilka organisationer som kommit längst i sin AI-implementering.

För att kartlägga detta studerades initialt hur AI-användningen ligger till i Sverige där det framkom att det är vanligare att AI implementerats hos större organisationer som har över 250 anställda (SCB 2019; SCB 2023). Storlek verkade således vara en faktor, varpå ett mål blev att få med så stora organisationer som möjligt och ett exklusionskriterie var organisationer som hade under 250 anställda. Ytterligare ett val var att exkludera företag som exklusivt arbetar med att utveckla AI-system, då dessa på grund av kommersiella incitament potentiellt kan undvika att framhäva risker med AI.

Genom onlinesökning skapades en lista som uppfyller ovan nämnda kriterier, från vilken en kartläggning gjordes över vilka organisationer som troligen faller inom någon av de sju samhällssektorerna för samhällsviktig verksamhet. Dessa verksamheter kontaktades sedan via mail där ytterligare två frågor ställdes: huruvida organisationen är klassad som samhällsviktig samt om de hade folk kunniga inom AI som kunde tänka sig att svara på enkäter och ställa upp på intervju. Av de 98 organisationer som kontaktades bokades 19 intervjuer, varav två stycken representerade samhällsperspektivet, åtta var från offentlig sektor och nio från privat sektor. För studien hittades respondenter inom varje sektor av de sju som omfattade samhällsviktiga verksamheter. För att säkerställa anonymitet i denna studie kommer detta inte redovisas ytterligare.

12 respondenter uppgav att de hade en ledande roll, varav vissa var chefer och andra del av ledande grupper. En majoritet av de som intervjuades hade bakgrund inom IT och digitalisering, men exakta titlar skilde sig en del, samt de olika avdelningarna titlarna härstammade från. Bland annat förekom titlar som CDO och CIO, men även ledande titlar där AI ingick. En del av respondenterna arbetade med AI på strategisk nivå, varav vissa mer ledande och andra som en del av utvecklingsgrupper. På grund av anonymitet publiceras inte samtliga titlar.

Organisationerna och respondenterna representerade en bred variation av verksamhetstyper, yrkesroller samt grad av AI-implementering, men trots det gav de relativt lika svar gällande riskkategorierna i stort, vilket talar för viss mättnad. Denna mättnad tyder på allmän konsensus eller enhetlighet i synsättet på riskerna, oavsett organisationstyp eller grad av AI-implementering.

3.2.1 Begränsningar

På grund av ämnets omfattning var det inte genomförbart att utforska varje risk med AI på djupet, vilket innebar att analysen i viss mån endast berörde ytan av dessa komplexa frågeställningar. Denna studie ger en mer generell bild och kan inte anses representera åsikter för samtliga samhällsviktiga verksamheter i Sverige. Vidare ger kan smalare urval erbjuda djupare insikter och perspektiv på de sex identifierade riskkategorierna. Om studien hade fokuserat på en specifik sektor inom samhällsviktig verksamhet och alla respondenter tillhört samma yrkeskategori, hade resultaten kunnat blir mer representativa för en bestämd population. Denna studie bör ses som en inledande bas för framtida forskning inom området, eftersom den erbjuder en översiktlig bild av riskernas omfattning inom samhällsviktiga verksamheter. Den ger en grundläggande förståelse, även om den endast skrapar på ytan av problematiken.

3.3 Datainsamling

Nedan beskrivs hur datainsamlingen gick till. Datainsamlingen inklusive transkribering skedde mellan februari och april 2024.

3.2.2 Enkät

Baserat på de slutliga sex riskkategorierna identifierade i litteraturstudien utformades 11 enkätfrågor där respondenterna fick ranka hur pass exponerade de ansåg sig vara i dagsläget för respektive exemplifierad risk inom respektive riskkategori på en skala 1-5, där 1 = liten risk och 5 = stor risk.

Ifall respondenten kände att någon riskkategori, alternativt de exemplifierade riskerna under riskkategorin, inte var relevant(a), eller att respondenten inte visste vad hen skulle svara, kunde hen välja att avstå att ranka risken i fråga. Utöver detta hade vardera riskkategori ett fritextsvar där respondenten kunde lägga till en risk som hen tyckte saknades. Med frisvarsfrågorna resulterade enkäten i totalt 17 frågor som presenteras i Bilaga C.

Begränsningar

En utmaning i arbetet var att skilja på respondenternas uppfattningar om nuvarande risker och framtida risker med AI eftersom några tolkade enkätfrågorna som att de främst berörde framtida förhållanden. Denna feltolkning uppdagades under intervjuerna. Trots att en liten grupp respondenter erkände att de eventuellt missförstått enkäten, ansågs dessa fall och avvikelsen vara för få och små för att motivera en korrigering av enkätresultaten.

3.2.3 Intervjuer

Syftet med intervjuerna var att uppnå en djupare förståelse för AI-riskerna än vad som är möjligt genom enkätsvar. Det ansågs också nödvändigt då ämnet är mångfacetterat.

Inför intervjuerna skapades en intervjuguide som återfinns i Bilaga D. Denna skickades till samtliga respondenter innan intervjun för att ge dem möjlighet till förberedelse.

Det område som tillägnades största tid under intervjun var riskkategorierna som respondenter tidigare fått svara på via enkäten. Dessa svar låg till grund för dialog kring riskerna med AI där respondenter fick utveckla varför de valt en viss rankning.

Fyra av intervjuerna genomfördes fysiskt och resterande digitalt. Vidare hölls fyra intervjuer där två respondenter deltog och resterande hölls med en. Alla intervjuer hölls på svenska utom en, och till denna översattes både intervjuguide och enkät till engelska.

En av författarna höll i intervjun, medan den andra förde anteckningar och ansvarade för teknik, med möjlighet att fylla i eventuella luckor i slutet av intervjun som krävde förtydligande. För att underlätta diskussion kring de sex riskkategorierna skapades ett visuellt verktyg som återfinns i Bilaga E. Ljudinspelning genomfördes efter respondenternas muntliga godkännande och för att skapa ett första ljud-till-text-utkast av transkriberingen användes AI-verktyget Klang.ai, med tillåtelse från respondenterna i syfte att informera om hur data hanteras. Därefter överfördes utkastet till egna dokument och materialet raderades från plattformen. Avslutningsvis kontrollerades utkastet mot ljudinspelning för korrigering av eventuella fel.

En av intervjuerna eliminerades från denna studie då respondenten inte hade arbetat med AI eller hade tillräckliga kunskaper inom AI för att kunna uttala sig kring de sex riskkategorierna.

Begränsningar

Under intervjuerna genomfördes vissa justeringar i intervjuguiden för att bättre anpassa den till studiens syfte och frågeställningar. Detta resulterade i att de tidigare intervjuerna inte hade samma struktur som de senare, och i vissa fall omformulerades frågor under processens gång. Dessa ändringar kan ha introducerat konsekvenser i den insamlade datan, vilket utgör en potentiell felkälla i studien.

Det var också en utmaning att upprätthålla objektivitet och undvika att leda respondenterna mot specifika svar. Svårigheter uppstod särskilt när respondenterna efterfrågade förtydliganden eller kategoriseringar, vilket ibland krävde exemplifiering av riskerna. Trots försök att vara objektiva kan vissa subjektiva inslag från författarna inte uteslutas.

Dessutom observerades skillnader i intervjuernas dynamik beroende på om de genomfördes med en eller två respondenter. Intervjuer med två deltagare tenderade att överskrida avsatt tid, som ibland nödvändiggjorde ett påskyndande av genomgångna svar, vilket påverkade svarens djup och bredd.

3.4 Analysmetod

Som analysram användes frågorna i intervjuguiden där den transkriberade texten kodades via analysverktyget NVivo. Initialt genomfördes kodningen gemensamt för att säkerställa samstämmighet mellan författarna angående kodningsprocessen. Därefter genomfördes kodningen individuellt för att öka effektiviteten. Vid eventuella svårigheter med kodningen

granskades texten av båda författarna. Från denna text utfördes analys där specifika teman identifierades och kategoriserades för att strukturera och fördjupa förståelsen av innehållet.

3.5 Reliabilitet och validitet

Oberoende av metodval för studier är det av stor vikt att den granskas kritiskt för att kunna bedöma om information som presenterats är pålitlig (Bell, 2006). Graziano och Raulin (1989) beskriver reliabilitet som att metoden eller mätningen ska ge liknande resultat vid ett annat mättillfälle, och är därmed ett mått på studiens pålitlighet. Reliabilitet knyts ofta samman med validitet, men kan uppnås utan validitet, medan validitet inte på motsvarande sätt kan uppnås utan reliabilitet (Graziano & Raulin, 1989). Validitet handlar istället om att studien mäter det den har för syfte att mäta (Bell, 2006).

För denna studie bedöms denna beskrivningen av reliabilitet och validitet vara enklare att tillämpa på enkäten, eftersom den mäter kvantitativ data. Genom att upprepa undersökningen kan det tydligare framgå om samma resultat uppnås vid ett senare tillfälle, vilket troligen är svårare att mäta i kvalitativa studier. Med tanke på AI:s snabba tekniska utveckling och frågornas subjektiva karaktär är det osannolikt att samma svar skulle erhållas vid ett annat mättillfälle, vilket innebär att reliabiliteten kan anses vara låg. Dock är studien huvudsakligen utformad kvalitativt och reliabilitet kan inom kvalitativ forskning beskrivas på flera sätt, bland annat genom analys av data mellan olika kodare (Creswell, 2013). För detta arbete tematiserade författarna enskilt transkriberingarna enligt intervjufrågorna, för att sedan kontrollera och analysera de svar som getts av respondenterna med varandra. Studien kan därmed anses ha god reliabilitet eftersom viss subjektivitet elimineras då författarna var två och överensstämmelse och eventuella skillnader systematiskt stämdes av.

Vid kvalitativa studier går även bedömning av validitet att göra på flera olika sätt (Creswell, 2013). Validitet avser ett sätt att bedöma “korrektheten” i resultaten, och författaren lyfter olika validitetsstrategier, och det föreslås att forskare försäkrar sig om att minst två appliceras. Bland dessa lyft bland annat *triangulering* där forskarna använder sig av flera källor, metoder och teorier för att stödja sitt bevis (Creswell, 2013). Detta har i studien gjorts flitigt genom att först göra en omfattande litteraturstudie och därefter både enkät och intervjuer.
Creswell (2013) lyfter även hur det är viktigt att forskare i början av studien klargör deras förutfattade meningar och fördomar som kan påverka undersökningen. För denna studie var ämnet helt okänt innan undersökningen påbörjades, varpå förutfattade meningar eller fördomar var låga. Däremot formades meningar under studiens gång och för detta arbete har ett riskfokus valts vilket tenderar att fokusera på negativa bemärkelser. Lika mycket forskning och teorier finns troligen kring AI:s positiva sidor, vilket denna studie inte representerar i så hög grad.

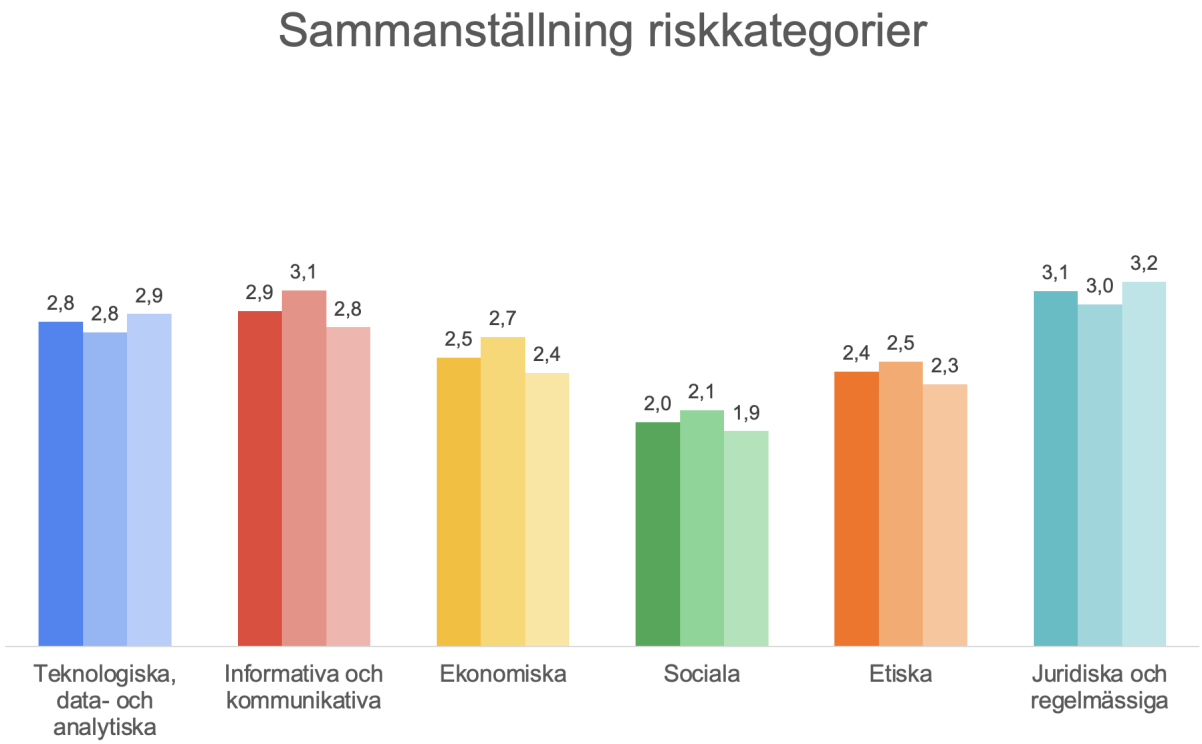
För detta arbete användes AI-verktyget Klang.ai för att underlätta transkribering. Som grund för detta gjordes en utförlig riskvärdering i vad detta kan innebära, främst gällande datahantering. Enligt verktygets policy sparas eller används ingen data på plattformen efter att den har raderats. För att ytterligare säkerställa respondenternas integritet och anonymitet informerades samtliga om att detta AI-verktyg kommer användas samt hur data kommer att hanteras, där ingen respondent hade någon invändning.

4 Resultat

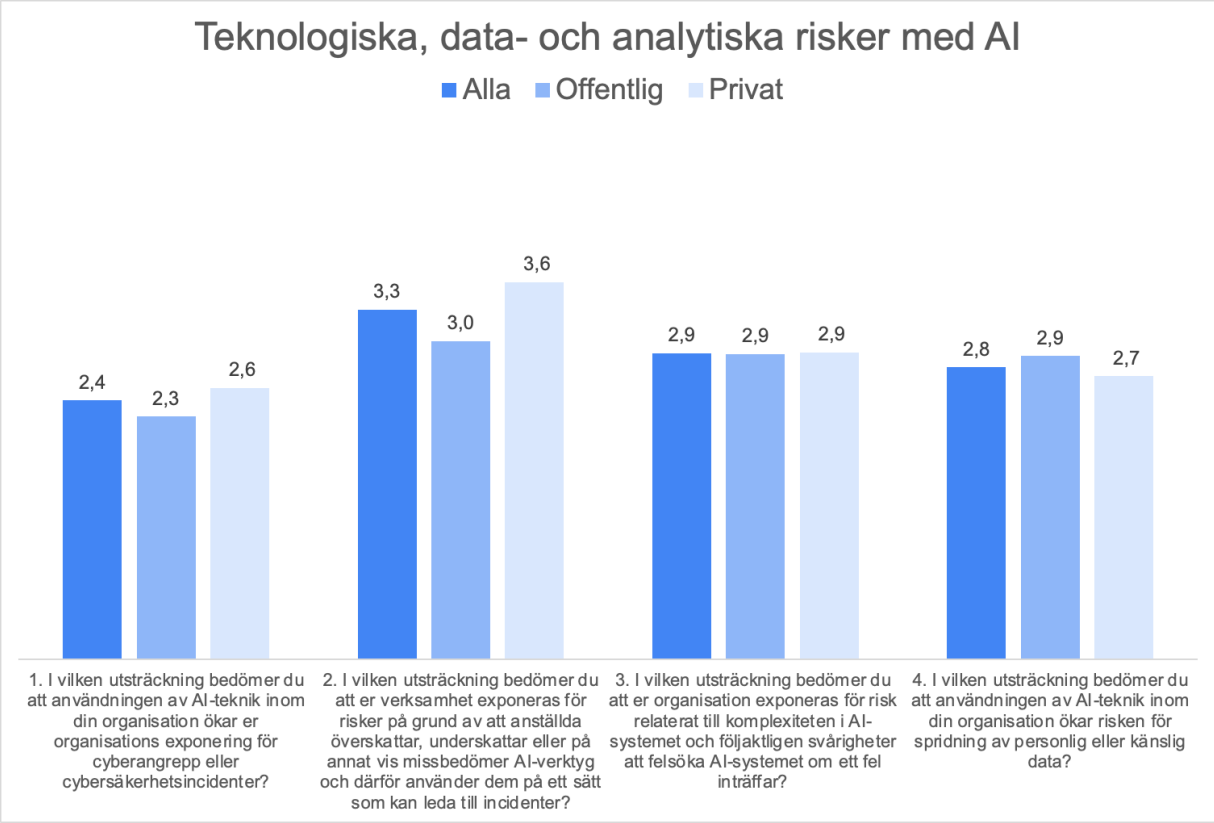
Nedan presenteras enkätsvaren samt resultaten efter analys och tematisering av intervjuerna.

4.1 Enkätsvar

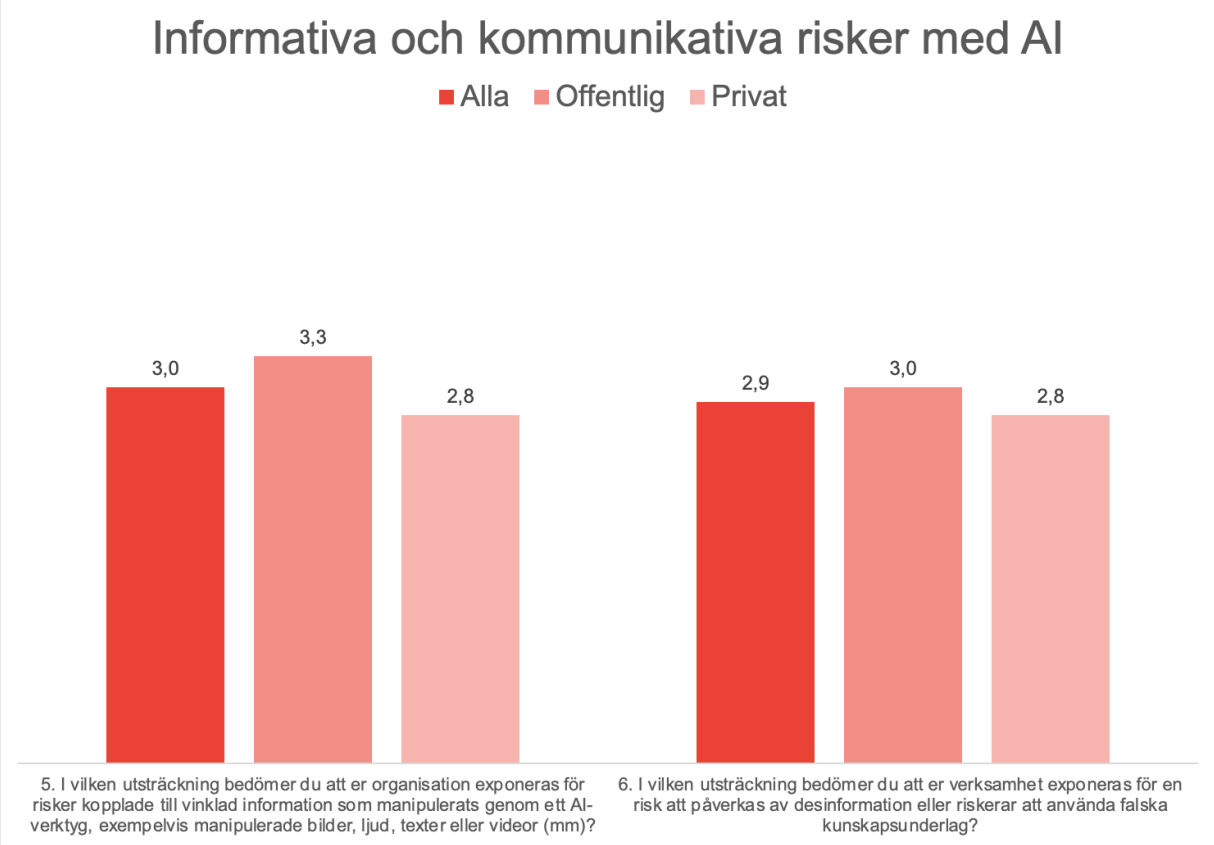
Respondenterna fick ranka respektive risk mellan 1-5, där 1 = liten risk och 5 = hög risk. Bland de högst rankade riskerna fanns juridiska och regelmässiga, teknologiska, data- och analytiska samt informativa och kommunikativa risker med AI. Några av riskkategorierna innehöll mer än en fråga. Resultaten i figur 1 redovisar medelvärdet av samtliga frågor för hela kategorin. Resultaten i figur 2-7 redovisar medelvärdet för varje enskild fråga.



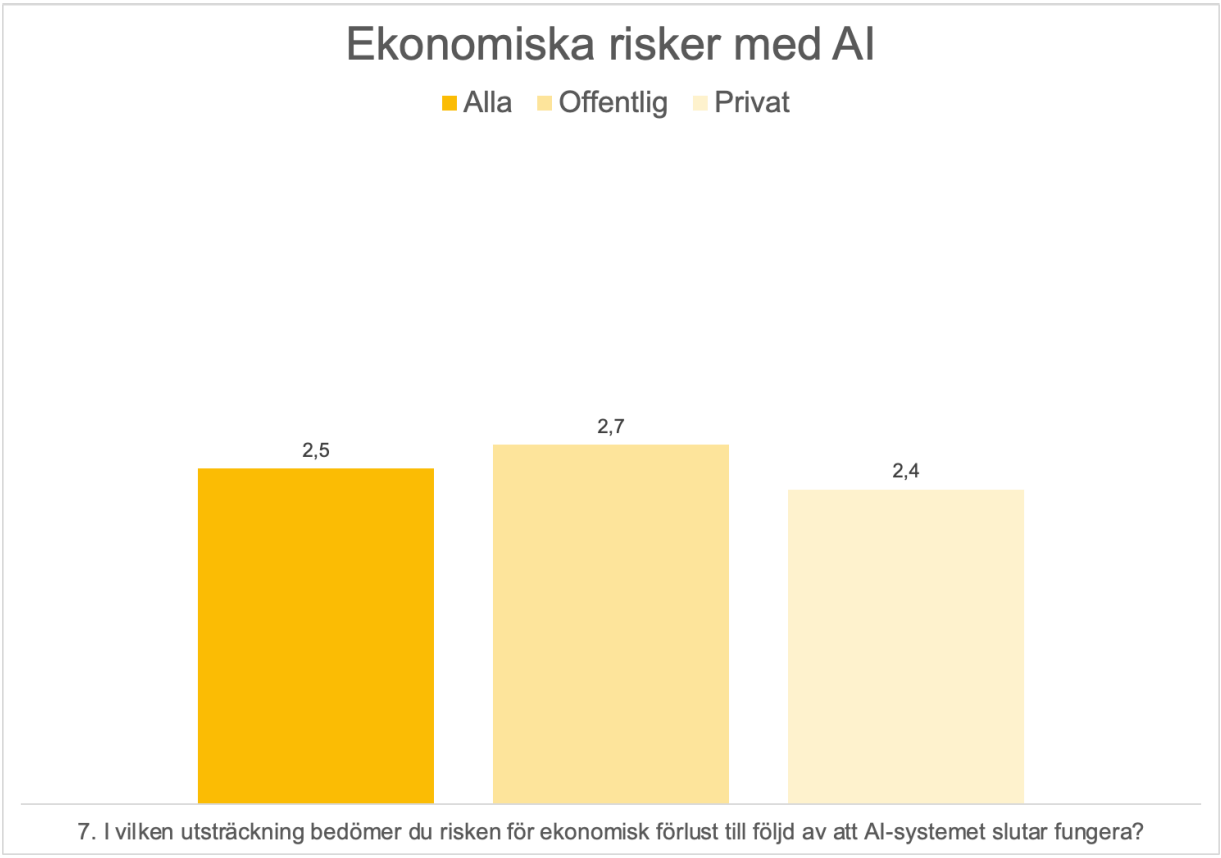
***Figur 1**. Sammanställning av medelvärde på riskkategorierna på en skala 1-5. Resultaten presenteras från vänster till höger i ordning alla, offentlig och privata sektor per grupp av staplar.*



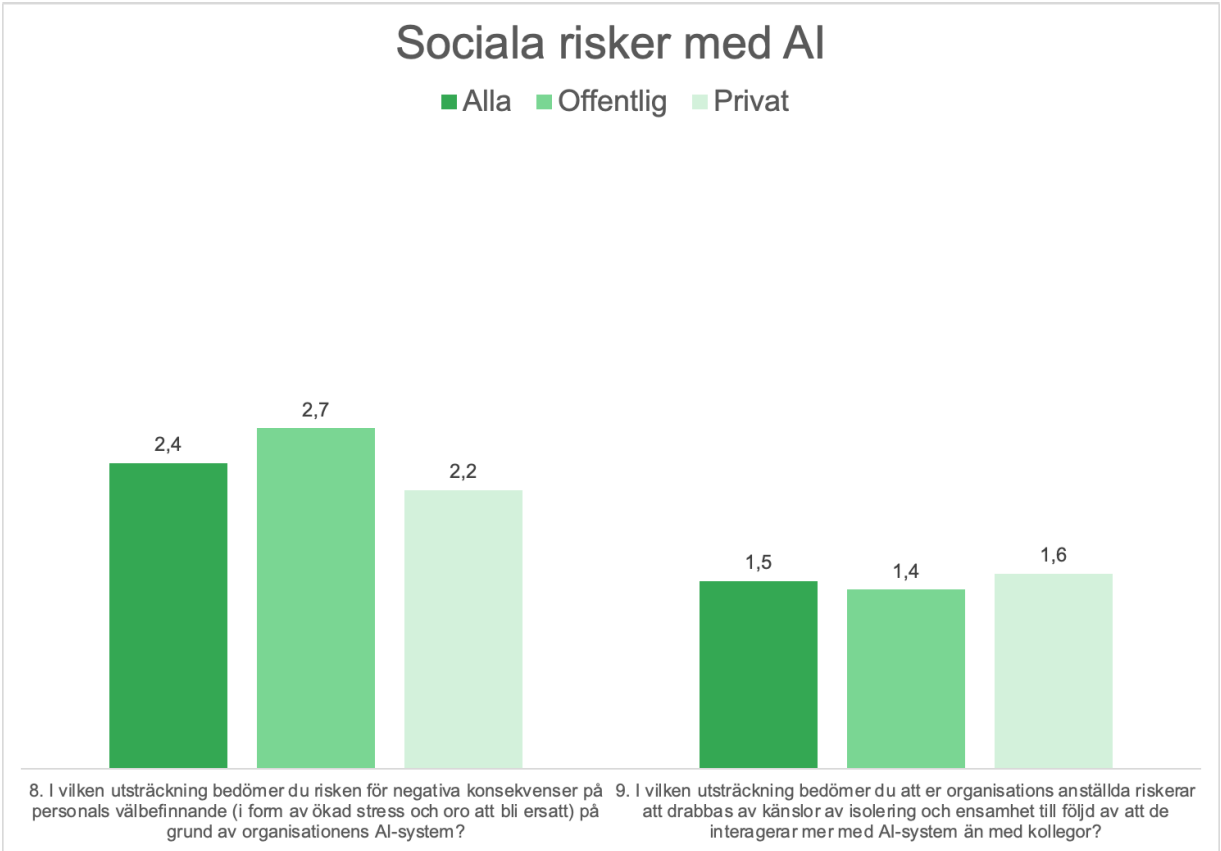
***Figur 2**. Medelvärdet av teknologiska, data- och analytiska på en skala 1-5.*



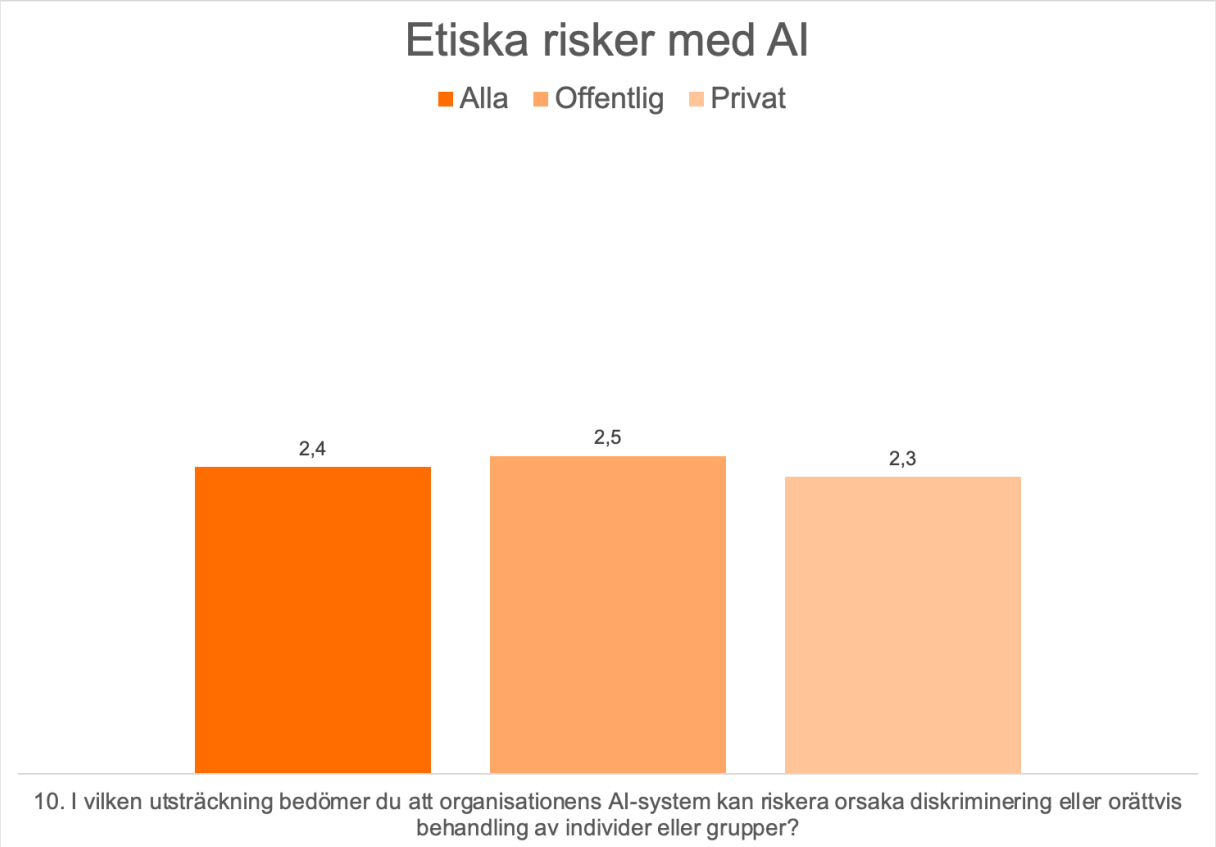
***Figur 3**. Medelvärdet av informativa och kommunikativa på en skala 1-5.*



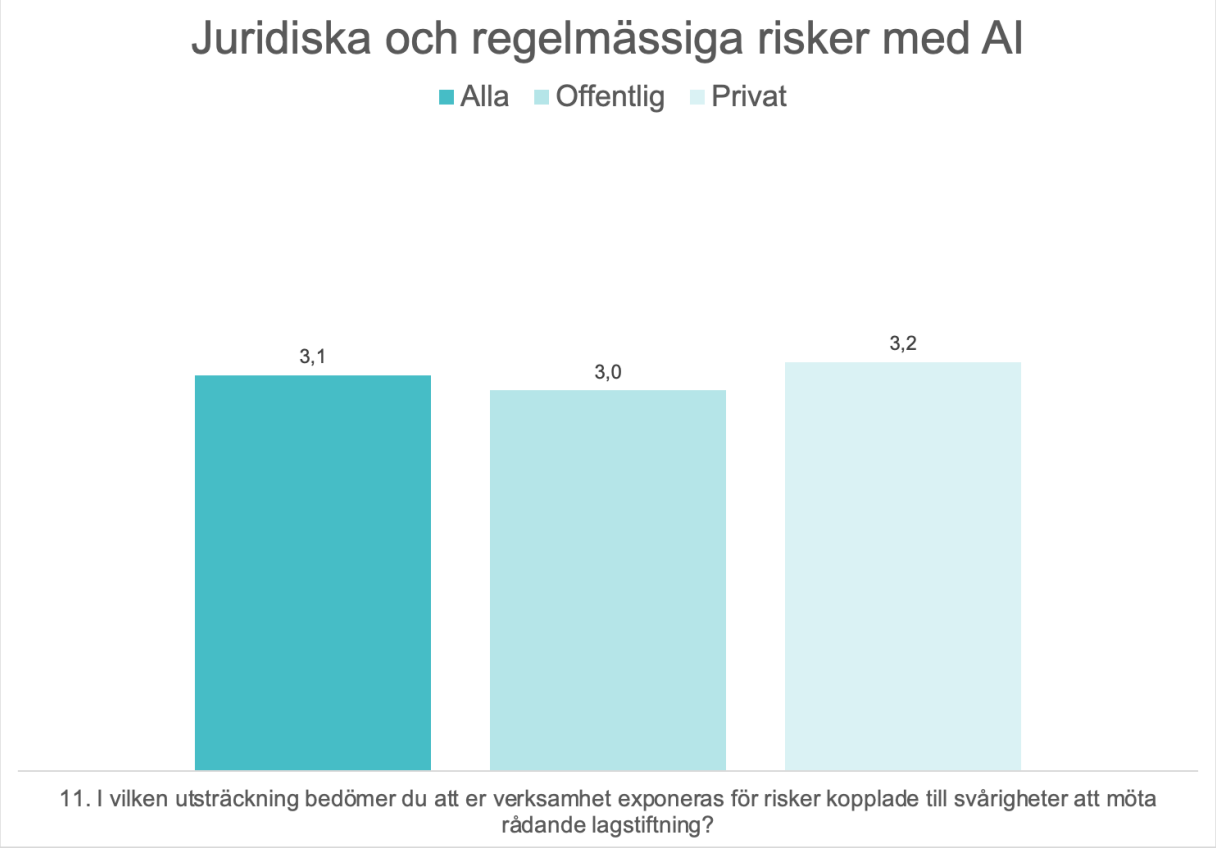
Figur 4. *Medelvärdet av ekonomiska på en skala 1-5.*



Figur 5*. Medelvärdet av sociala på en skala 1-5.*



***Figur 6**. Medelvärdet av etiska på en skala 1-5.*



***Figur 7**. Medelvärdet av juridiska och regelmässiga på en skala 1-5.*

4.2 Intervjusvar

Nedan presenteras en djupare genomgång av intervjusvar baserat på intervjuguidens olika frågor samt riskkategorierna. Respondenter från privat sektor förkortas som PR och från offentlig sektor som OR.

4.2.1 AI-definition

Majoriteten av respondenterna hade sedan innan koll på EU AI Acts definition av ett AI-system, och alla var villiga att ställa sig bakom den för intervjuns skull. Det rådde delade meningar bland de respondenter som hade åsikter kring definitionen, där en respondent från privat sektor beskrev den som "aggressiv" med motiveringen att den kräver mer än vad man normalt har sett kring AI. En annan respondent från privat sektor poängterade också svårigheten kring att definiera AI, och en respondent från offentlig sektor ansåg sig visualisera AI som ett forskningsfält med system som besitter vissa typer av egenskaper. En annan respondent från offentlig sektor utgick från Europaparlamentets definition från 2020 som anses vara "mjukare i kanterna", och poängterade att den och arbetets definition skiljer sig en del åt trots att båda kommer från EU. En respondent från privat sektor menade att definitionen "nog inte följs av de som jobbar med AI", medan en respondent från offentlig sektor menade att den "omfattar AI så som det diskuteras idag".

"Jag tycker nu att det här citatet ni plocka ut där är väldigt generellt och inte särskilt väl beskrivande det som de flesta menar med AI nu och det sista halvåret om man säger så. Det har gått så otroligt fort så det där är mer en allmän AI [...] AI i betydelsen automatiserat beslutsfattande. Det som präglar AI både ur möjligheternas perspektiv och ur risker är den här maskininlärda AI. Det är den som är både möjliggörarna och riskerna nu. Den här gammaldags AI som man har hållit på med i 15 år, den har inte alls de här problemen som alla är oroliga för nu. Utan det är de här diffusa, automatiskt upptränade grejerna som ingen riktigt vet hur de fungerar. Det är de som är risken. Därför tycker jag att den är lite…. Det märks att det kanske är så att den är lite gammal den här definitionen, eller ja, i AI-mått mätt. Att den har ett par år på nacken och att man inte hade insett att de behöver faktiskt snäva in det på de här språk- och bildmodellerna som man har tränat på de här jättestora datamängderna. För de är ju fundamentalt annorlunda mot allt annat vi har haft förut. Det gäller ju inte för de här gamla AI-systemen" - OR

4.2.2 Användning av AI

I början av intervjun ombads respondenterna att redogöra för deras nuvarande användning av AI, deras eventuella framtida användning samt en kortfattad beskrivning av den tekniska uppbyggnaden av dessa AI-system. Detta innebar att de förklarade huruvida systemen baseras på exempelvis ML, DL, språkmodeller eller andra tekniker.

Nästan alla respondenter angav att de använder ML för olika applikationer inom organisationen. Många använder också DL/NN samt språkmodeller där sistnämnda bland de flesta var i teststadiet. Generativ AI och större språkmodeller var den AI som de flesta uppgav var i planeringsstadiet och som flera organisationer funderar på att implementera.

Majoriteten av respondenterna hade använt AI under en längre tid, varvid ML ofta nämns som ett etablerat användningsområde. Endast några få uppgav att de inte hade använt AI tidigare, och samtliga av dessa befann sig i utvecklings- och testfasen för att implementera AI.

Bland respondenterna varierade det om de utvecklade AI-lösningar internt eller om de köpte in dem från externa leverantörer. Hos några av organisationerna förekom intern utveckling av AI, där vissa också använde externa verktyg som komplement. För andra organisationer var det vanligare att köpa in AI-lösningar, eftersom intern utveckling kunde vara utmanande på grund av begränsade resurser.

Användningsområdena av AI varierar beroende på verksamheternas behov och mål, från automatisering av rutinmässiga uppgifter till avancerad dataanalys och kundinteraktioner. De mest utbredda användningsområdena för AI är inom kategorisering och klassificering, där AI sorterar stora informationsmängder för att förenkla vidare hantering. Andra vanliga tillämpningar inkluderade användning av AI för att göra prognoser, transkribering, översättning och bildanalys. Mindre vanliga användningsområden som nämns inkluderar chatbotar för anställda, AI för kundassistans och AI i riskbedömning. Specifika AI-verktyg som nämndes av några respondenter var bland annat från Microsoft (exempelvis Copilot), GitHub och BERT, men dessa verktyg behandlas inte i detalj inom detta arbete.

4.2.3 Risker med AI

Nedan presenteras de sammanställda svaren från intervjuerna, de områden som flest respondenter diskuterade och som ansågs vara nyckelaspekter av riskerna under samtalen.

Teknologiska, data- och analytiska

Responsen från deltagarna i denna kategori visade en bred variation, även om vissa frågor tenderade att framkalla gemensamma teman. Många deltagare uttryckte liknande synpunkter om riskerna med datadelning, specifikt inom cybersäkerhet och missbruk av AI-verktyg av anställda, vilket kan leda till läckage av känslig information. Däremot fanns en del skilda åsikter om dessa frågor. Av den anledningen presenteras resultatet här utifrån varje enskild fråga istället för en sammanfattande tematisk analys över hela kategorin, eftersom det inte var möjligt att integrera alla svar under en enda tematik, trots att många svar liknade varandra.

Cyberangrepp eller cybersäkerhetsincidenter

Majoriteten av respondenterna ansåg inte att de var särskilt exponerade för cyberangrepp eller cybersäkerhetsincidenter. Några av respondenterna berättade att sättet som de använder AI på inte utgör en direkt risk för cyberattacker, vilket delvis berodde på att AI-användningen inte var så avancerad eller omfattande inom deras organisationer, samt att de som implementerade AI endast använde det internt inom organisationen. En respondent från den privata sektorn, vars organisation hade sitt AI-system kopplat till molntjänster (även kallat “molnet”), ansåg dock att detta var säkrare än att ha det internt, tack vare de högre säkerhetsnivåerna som sådana tjänster vanligtvis erbjuder.

Även om flertalet organisationer inte ansåg att de var speciellt exponerade för denna risk upplevde majoriteten ett ökat hot från externa aktörer som använder AI-teknik för att utföra cyberangrepp, och bland dem som trodde att denna risk skulle öka inom organisationen i

framtiden var detta den faktor som låg bakom denna skattning. En respondent beskrev det som en ständig utmaning för organisationen att försöka hålla sig steget före hotaktörer. En annan påpekade risken för att bli attackerad när AI-system importerades utifrån, eftersom en illvillig aktör kunde sabotera kod eller modeller på olika sätt. Detta var en faktor som några respondenter ansåg kunde bidra till en ökad risk, samtidigt som det fanns de som såg möjligheter med AI i att motverka cyberattacker och som menade att det bidrog till en förbättrad cyberförsvarsförmåga.

Komplexitet och svårighet att felsöka

I stora drag kunde det konstateras bland respondenterna att risken inte var särskilt hög inom vare sig privata eller offentliga organisationerna i dagsläget. En betydande faktor var vad för AI-system som används, samt om AI-systemet är utvecklat av organisationen själva eller om det är utvecklat av tredje (extern) part. Vissa respondenter, både från privat och offentlig sektor, argumenterade för att egna system, ofta uppkopplade på interna servrar, är säkrare och att man har mer kontroll på dessa. Vissa egna system som härstammar från hämtad kod saknar samma säkerhetskontroller som en tredje parts-lösning har, och kräver därför intern riskhantering. För de med tredje parts-lösningar var en framträdande inställning gällande svårigheter kring felsökning och komplexiteten att detta är upp till tredje part att lösa, samtidigt som en privat respondent även lyfte hur detta gör att organisationen själv får svårt att lösa de problem som eventuellt uppstår.

Några respondenter yttrade att risken kopplat till komplexiteten i nuvarande AI-system inte är så närvarande inom respektive organisation, medan andra menade att AI inte används på ett sätt att risken i fråga kan komma att bli ett problem. Respondenter från privat och offentlig sektor lyfte även behovet av AI-kompetent personal. En respondent från offentlig sektor menade att organisationen var fullständigt införstådd i komplexitets- samt felsökningsproblematik sen tidigare, samt att även om komplexiteten ökar i framtiden kommer ökad kompetens göra att den totala risken inte ökar.

En respondent från privat sektor poängterade att “Explainable AI” saknas i det vardagliga AI-arbetet och kräver mer fokus i framtiden. Respondenter från både offentlig och privat sektor lyfte även vikten av mänsklig delaktighet i AI-systemet, alltså att AI:n är vägledande, inte beslutsfattande, för ett slutgiltigt, mänskligt beslut.

Något som lyftes från en respondent från offentlig sektor var ett transparenskrav som framförallt myndigheter lyder under, som begränsar vad för AI-implementering som är möjlig att genomföra.

Vad gäller framtidsutsikten för denna risk var även en återkommande kommentar från respondenter inom både privat och offentlig sektor att ju mer system kombineras, integreras och avanceras, desto mer ökar dess komplexitet och försvårar insyn, vilket i sin tur ökar risken kopplat till svårigheten att felsöka. Generativ AI är något som respondenter från både offentlig och privat sektor lyfte som en potentiell implementering, vilket i sin tur medför större svårighet för felsökning, minskad insyn samt förståelse för hur systemen fungerar då dessa AI-system ansågs vara mer avancerade och komplexa.

Över- eller underskattning alternativt missbedömning av AI-verktyg

I stora drag är respondenter från både offentlig och privat sektor överens om att detta främst handlar om risken att anställda *överskattar* AI-verktyg, vilket kan leda till oönskade konsekvenser. Detta grundar sig i en okunskap gällande AI-verktygen och konsekvenserna av felaktig hantering. Denna risk är nära sammankopplad till risken för spridning av känslig eller personlig data, då det är den konsekvens som majoriteten av respondenterna flätar samman med överskattning alternativt missbedömning av AI-verktyg. Detta gäller AI-verktyg i allmänhet, men generativ AI i synnerhet, som näst intill alla respondenterna nämner som det största användningsområdet som riskerar att överskattas eller missbedömmas, där resultatet också kan orsaka spridning av data.

Majoriteten av respondenterna från både offentlig och privat sektor anser att risken har stor förbättringspotential om utbildning av personal prioriteras framgent. AI-policys och riktlinjer lyfts också, där vissa upplever sig ha strikta riktlinjer kring hur personal handskas med AI, medan andra uttrycker att de inte har så mycket riktlinjer. En privat respondent pekar ut säkerhetskultur som en viktig del i detta arbete.

Något som däremot stack ut var att det av några respondenter inom privat sektor även lyftes en oro att AI *underskattas* och inte används tillräckligt, och därmed att många möjligheter som AI-verktyg ger riskeras att missas, något som inte nämndes av någon av de offentliga respondenterna.

Spridning av personlig eller känslig data

Som nämnt ovan vävdes svaren från frågan ovan nästan genomgående ihop med svaren till denna, varför resonemanget överlag var likadant. Något som däremot utmärkte sig var att respondenterna från privat sektor hade delade åsikter gällande hur stor risk de ansåg att spridning av personlig data var, vissa ansåg att det var en stor risk, andra låg, vilket också var branschberoende. Bland respondenter från offentlig sektor ansågs risken hög för spridning av personlig data. Risken för känslig data uttrycktes olika hög beroende på verksamhet och bransch, och bland vissa användes känslig och personlig data som likvärdiga. Här poängterades även att detta inte enbart berör AI utan även är generellt för digitalisering. En del orosmoment lyftes däremot gällande användandet av just generativ AI, främst de "öppna" verktygen likt ChatGPT, som nästan uteslutande användes som exempel där anställda riskerar läcka data.

Här lyfte respondenterna också frågan om molntjänster respektive interna servrar mycket. Det råder delade åsikter om vad som är säkrast bland samtliga verksamheter, men en generell trend är att de offentliga verksamheterna är mycket mer måna om att utveckla interna AI-verktyg för att minska risken för att personlig (och därmed känslig) data sprids av misstag. En annan infallsvinkel var hur svårt det är för (framför allt) offentliga verksamheter att följa med i den tekniska utvecklingen, eftersom att de flesta har för mål att utveckla egna system som ska motsvara de öppna, allmänna verktygen (som exempelvis ChatGPT).

Något som även poängterades var att många respondenter, både från privat och offentlig sektor, upplever en viss rädsla att gå miste om möjligheterna med användandet av AI-verktyg, och att man i rädsla av att riskera göra övertramp låter bli att utforska dessa.

Informativa och kommunikativa

Responsen för båda dessa frågor inom informativa och kommunikativa AI-risker var väldigt lika för många respondenter och berörde ofta samma problemområden. Framförallt uttrycks denna risk som ett samhällsproblem som påverkar och fortsatt kommer att påverka både individer och organisationer. Falska nyheter och spridning av desinformation är idag redan ett utbrett problem och många respondenter anser att AI kommer bidra till att detta amplifieras i framtiden där de direkt eller indirekt kommer påverkas. Den direkta påverkan kan vara att informationen som inhämtas från omvärlden är falsk eller att falsk information skickas direkt till organisationen via mail eller andra kanaler. Indirekt exponering är via de effekter som falsk information kan ha på det politiska klimatet och lagstiftning. Här skilde det sig något mellan privat och offentlig sektor där flera respondenter inom den offentliga sektorn ansåg sig i hög grad vara exponerade för denna risk på grund av dess utbredning i samhället. Ett stort problem med denna riskutbredning i samhället är hur information överlag kommer förlora sitt värde. Denna problematik uttrycktes bland annat som:

"[...] hur kan vi snart knappt lita på någonting som genereras? Antingen kommer folk ta allt för givet att det är i sanning eller så kommer folk vara misstänksamma mot allt" - PR

Respondenter från organisationer som på olika sätt ansvarar för informationsgivning till allmänheten upplevde detta som en stor risk då de kan få utmaningar i att bedriva sin verksamhet på grund av falsk information som sprids av illasinnade till allmänheten, eller då allmänheten börjar misstro informationen som de skickar ut.

Några respondenter ansåg att detta problem inte är specifikt för AI utan att detta är ett problemområde som funnits sedan en längre, men majoriteten tror dock att användningen av AI-verktyg för att generera falsk media och sprida desinformation kommer göra det svårare att avgöra vad som är sann respektive falsk information i framtiden. Här menar några respondenter att man också kan använda AI som ett moteld mot denna problematik i framtiden där AI kan användas för att känna igen material som blivit manipulerat.

Hur exponeringen för denna risk upplevdes skilde sig väldigt markant mellan olika branscher. Några respondenter beskrev hur de exponerats idag där exempel som angavs var att anställda fått AI-manipulerade mail eller blivit uppringda av AI-manipulerade röster där avsändaren önskat få ut värdefull data. Respondenter från organisationer som på olika sätt förlitar sig på information via media känner sig extra utsatta då de använder sådant material som underlag till mycket och har därför utvecklat checklistor för faktakoll för att minimera risken. Inom den offentliga sektorn berättade respondenter att de exponerats för falsk information i form av falska dokument och andra handlingar som skickas som underlag vid beslutsprocesser där avsändaren på olika sätt förfalskat materialet. Här nämns dock inte alltid AI som ett specifikt verktyg vid sådana incidenter, men flera av respondenterna menar att denna redan existerande problematik troligen kommer bli värre med mer avancerade AI-verktyg. Inom offentlig sektor upplevdes det att de är mer utsatta än privat sektor:

"Det är ju jävligt populärt just nu, liksom, att ge sig på, inte minst offentlig sektor. Det är ju där skadan är som störst"- OR

Generellt är risken för exponering bland respondenterna idag lik den de senaste åren men denna förväntas öka och bli svårare att hantera allt eftersom AI, främst generativ AI, blir allt mer avancerad och kan producera allt mer högkvalitativa underlag. Många respondenter upplever att de har svårt att möta denna problematik idag och ser det som en stor utmaning i framtiden, men en respondent upplevde att vi i takt med att AI blir mer avancerat också kommer bli mer medvetna om problematiken.

Kategorins två frågor genererade till större delen samma svar, men viss skillnad fanns gällande exponering där fler ansåg sig inte vara exponerade för manipulerad media i form av bilder, ljud mm men nästan alla uppgav att desinformation, som på grund av dess effekter på samhället, kan påverka organisationen. De respondenter som inte upplevde att de var exponerade för manipulerad media var framförallt inom inom privat sektor, som beskrev att de inte använde sådant material inom organisationen och inte hade eller använde AI-system på sådant sätt som skulle göra dem exponerade. En respondent inom den offentliga sektorn angav låg risk på denna exponering med motiveringen att organisationen har utvecklat bra sätt att hantera denna problematiken idag och därmed inte tror att detta kommer bli mer utmanande i framtiden.

Gällande desinformation var dock flera överens om att man inte kan säga med säkerhet att de är helt oexponerade på grund av dess utbredning i samhället, vilket uttrycktes som:

"Det är inte en fråga om vi är infiltrerade. Det är en fråga hur mycket infiltrerade vi är" - PR

Ekonomiska

I stora drag har respondenterna inom privat och offentlig sektor liknande synsätt, nämligen att ekonomiska risker inte är så närvarande med tanke på att man inte kommit så långt med sin AI-implementering, alternativt att den AI som implementerats inte används på ett sådant sätt att ekonomisk förlust skulle uppstå vid eventuellt avbrott.

"[...] Men det [AI-systemet] är ju ingenting som vi är beroende av. Så om de systemen går ner så kanske vi jobbar lika effektivt som vi gjorde för ett år sedan istället. Så det är ju ingen ekonomisk skada då idag. Om fem år så kanske vi har skapat oss ett beroende" - PR

Enstaka respondenter från både privat och offentlig sektor ansåg sig ha redundans i sina system idag som dämpade sina AI-systems ekonomiska påverkan till näst intill ingen. En vanlig uppfattning bland respondenterna var att ekonomiska riskerna med AI idag inte skiljer sig markant från de med generell IT eller automatisering:

"[...]rent allmänt att vi är väldigt beroende av att IT-systemen är uppe. Så att jag tror inte att det är större risk med AI-system än ett annat IT-system." - PR

Bland de respondenterna från privat sektor nämndes även andra ekonomiska risker som straffavgifter/sanktioner till följd av AI-orsakade fel, men även mer storskaliga risker som exempelvis skada på organisationens varumärke, vilket i slutändan skulle kunna utgöra en

ekonomisk risk. En annan synpunkt från respondenter från både privat och offentlig sektor rör hur en ekonomisk risk är att inte använda AI och därmed inte realisera eventuella besparingar om man inte använder AI på ett smart sätt. Några av de respondenterna från offentlig sektor lyfte en annan ekonomisk farhåga, nämligen att investeringar rörande AI inte ger ekonomisk vinning till organisationerna, utan snarare till andra:

"Vi får en kostnad för AI-lösningen [...] men effekterna hamnar någon annanstans. För oss blir det bara dyrare" - OR

Framgent ansåg dock en majoritet av respondenterna från både offentlig och privat sektor att den ekonomiska risken kan komma öka i takt med att AI-systemen blir mer och mer integrerade, och att organisationer utvecklar ett beroende av dessa, varpå ett avbrott i systemen skulle få stora negativa effekter.

Sociala

Huruvida respondenterna upplevde att deras anställdas välbefinnande påverkades av implementationen av AI berodde till stor del på den allmänna attityden hos anställda, dels den generella attityden mot förändring men också specifikt attityden mot AI. De respondenter som rapporterade att AI inte har eller kommer ha inverkan på personalens välbefinnande lyfte bland annat kulturen på arbetsplatsen gentemot AI som en viktig faktor där man genom utbildning och information om AI:s påverkan på arbetslivet ser till att de anställda inte påverkas negativt. Även de anställdas nyfikenhet och vilja till utveckling bidrog till att AI generellt inte upplevdes som något orosmoment.

Bland de respondenter som upplevde att AI har negativ inverkan på personalens välbefinnande lyftes personals generella motstånd till förändring, särskilt när det gäller införandet av nya tekniska verktyg som inte alla är lika bekväma med kan leda till ökat motstånd och ökad stress. Det fanns även vissa yrkesgrupper som respondenterna upplevde var mer utsatta och som till större del hade uttryckt oro för att bli ersatta. Detta var framförallt arbeten av administrativ karaktär, anställda som arbetar med programmering och anställda som utför repetitiva arbetsuppgifter. Vidare ansågs brist på kunskap vara en bidragande orsak till sämre välbefinnande där dålig kunskap kring AI och dess effekter ansågs bidra till stress och oro att bli ersatt.

"Vi tänker att om man inte vet vad det är, då har folk en tendens att bli lite oroliga för saker som är okända" - PR

Utöver positiv attityd hos anställda lyftes andra faktorer som ansågs minska denna risk. En faktor var då organisationerna inte använde eller planerade använda AI på ett sådant sätt att det skulle kunna ersätta arbetsuppgifter. Flera respondenter upplever inte att AI kommer ersätta arbetare och specifika yrken utan att det snarare kommer leda till förändring och utveckling. Några respondenter gillade att jämföra detta med tidigare samhällsförändringar och menar att AI därmed inte är unikt och troligen kommer påverka yrken på liknande sätt.

"Alltså det är många som säger att nu kommer AI definitivt ta våra jobb. Även white collar-jobb, inte bara taxichaufförer och kassamedarbetare. Men min uppfattning är att så har vi

alltid sagt. [...] När traktorn kom istället för häst och plog. Alltså den typen av diskussioner har vi alltid. Men vi har lyckats" - PR

En respondent motsatte sig dock detta och menar att AI kommer ersätta arbeten och ha en större påverkan än vad vi sett tidigare i historien. Det var också bland dessa respondenter, som tror att AI har potentialen att ersätta yrken i framtiden, där risken för ökad stress och oro bland anställda förväntades stiga.

"Ja, det beror ju på hur man menar med ersatta. Enskilda jobb, absolut. Men de blir väl också mycket omdefinierande, som väl har skett i 300 år. Samtidigt så tror jag att det här är kanske mer omvändande än mycket annat. [...] Så jag är fortfarande ödmjuk inför.. jag tror inte att det är exakt samma sak som att man kan symaskiner " - PR

Flera respondenter anser att AI-implementering är viktigt för organisationen. En respondent inom den privata sektorn såg AI som essentiellt för organisationens överlevnad:

"Alltså om vi bestämmer att "nej men vi ska inte hålla på med AI-system", då tappar vi ju all vår… Då har vi ju inte möjlighet att konkurrera mot andra företag. Så då går vi i konkurs på tio års sikt [...] Så om man vill värna om personalen och ha kvar personalen, då måste man ju hänga med och använda de här verktygen. För annars blir man ju helt irrelevant".

Inom den offentliga sektorn ansågs AI-implementering vara viktig för att kunna möta rådande utmaningar såsom arbetsbelastning och brist på personal. Positiva aspekter som lyftes var bland annat att personalen får mer tid över och även slipper arbeta med monotona arbetsuppgifter, en respondent uttryckte hur denna avlastning så småningom kanske kan leda till att vi kan gå ner till fyra arbetsdagar i veckan. Det var dock en respondent från offentlig sektor som väckte viss oro kring denna effektivisering och menade att detta inte nödvändigtvis kommer leda till att personal blir avlastad:

"Samtidigt som det å andra sidan kan ge upphov till att vi tvingar människan att vara för fokuserad, koncentrerad hela tiden och aldrig får enkla uppgifter som kan vara lite avlastande och balansera arbetet under en hel vecka."

Risken gällande personalens upplevelser av isolering ansågs idag vara låg eller obefintlig bland samtliga respondenter. Några få respondenter ansåg att denna risk eventuellt kan bli förverkligad i framtiden men då om många år och inte inom vår livstid. Respondenter inom den offentliga sektorn lyfte distans och hemarbete som en större riskfaktor för upplevelsen av isolering.

Etiska

I stora drag upplevde respondenter från både privat och offentlig sektor att etiska AI-risker till stor del är sammankopplat med komplexitet där fenomenet "black box" lyfts återkommande. Som exempel lyfts att AI-system kan sätta ihop egna, i vissa fall diskriminerande, parametrar som är svåra för människor att förstå, eller att människor pga black box-problematik inte förstår eller upptäcker rådande diskriminering i AI-system. Generellt pratar respondenterna

om denna risken mer på samhällsnivå än organisationsspecifikt. Generativ AI lyfts som ett stort fokusområde även för den etiska riskkategorin. En återkommande notering bland respondenterna är att AI-systemen får inbyggda bias i sin träningsdata eftersom att människor är biased. I samma anda lyfter även respondenterna att AI-systemen löper risk att reproducera gamla samhällsstrukturer, exempelvis vid bildgenerering, där ett exempel är att AI vid förfrågan att generera bild på en VD skulle generera en bild på en man.

Några respondenter menar sig ha en hög etisk standard inom organisationen och att AI-systemen inte ska utgöra ett särskilt problem. Policys och ramverk kopplat till etik och specifikt diskriminering lyftes som stödfunktion för anställda bland respondenter från både privat och offentlig sektor. Förbättringsprotential kan även ses om man ökar kunskapen, då människor i vissa fall inte uppmärksammar diskriminering i AI-system. En offentlig respondent berättade om ett etiskt råd som vägleder organisationen i frågor likt diskriminering och orättvis behandling.

Ett återkommande skräckexempel inom etiska risker är ifall AI används inom rekrytering för exempelvis gallring av ansökningar, men ingen av respondenterna uppger att detta används idag. Framgent råder det delade meningar om risken ökar eller minskar, där vissa respondenter menar att etiska AI-risker kan öka i takt med att AI-systemen blir mer storskaliga och avancerade. Å andra sidan menar vissa respondenter att man blir bättre på att handskas med diskrimineringsfrågor, alternativt att lagstiftning som exempelvis EU AI Act kommer att hjälpa till att "sanera" diskriminering från AI-systemen. Något som också lyfts är att AI-system har potential att kunna motverka diskriminering i framtiden, och därmed minska risken.

"Det finns ju många nivåer av diskriminering som inte vi själva ens medvetna om eller ser och den kommer ju lära sig alla dem [...] Om vi inte känner till dem så kan vi inte heller tvätta bort dem. Men det kanske går att lösa med ytterligare AI. Det är möjligt att det går att liksom träna någon på att hitta diskrimineringsmönster eller oegentligheter [...]" - OR

En respondent inom den offentliga sektorn lyfte även den etiska risken att *inte* använda AI, att de som organisationen tjänar har rätt till den nya tekniska utrustningen som finns och inte förtjänar att hamna efter för att organisationen vill vara försiktig med nya AI-implementeringar.

Juridiska och regelmässiga

Många av respondenterna lyfte här den kommande EU-lagstiftningen EU AI Act som planeras träda i kraft om cirka två år, trots att frågan var formulerad att representera dagens rådande lagstiftning. Därmed speglar många av svaren en kombination av nutid och framtid, vilket redovisas nedan.

Utmaningar med rådande och kommande lagstiftning

En gemensam faktor som delades mellan många respondenter var hur lagstiftning gällande AI, och då främst datahantering, upplevdes som otydlig och svår att begripa. Aspekter som AI:ns bristande transparens gör att det inte går att avgöra om AI:n själv drar slutsatser som kan vara diskriminerande eller att man felklassificerar data via AI som därmed kan leda till

lagbrott genom dataläckage, där felklassificering gör att data hamnar på fel ställe. En respondent lyfte frågan kring hur man ska hantera upphovsrätt om material används från en AI vars träningsdata och därmed output genererats av olika verk från andra kreatörer.

Några respondenter lyfte även här den tidigare nämnda problematiken i att använda AI som är anslutet till molnet. Många upplevde att de bästa AI-tjänster som finns tillgängliga idag finns i molnet och om organisationerna själva skulle försöka skapa liknande verktyg inom organisationen skulle dessa modeller inte bli lika effektiva då molntjänsterna är mer omfattande och svåra att återskapa i mindre kontexter. Framförallt kände respondenter inom den offentliga sektorn att de blev begränsade i sådan användning då de har data vars reglering och hantering lyder under striktare lagkrav och som vid spridning kan få rättsliga konsekvenser. De upplevde att lagstiftning kan ha en hämmande effekt på deras tekniska utveckling och innovation. Detta var dels på grund av datatyp men också de upplever att de har flera lagar de måste förhålla sig till som kan krocka med AI-lagstiftning. Dessutom nämndes det att aktörer inom den offentliga sektorn ofta "tar höjd" när det gäller att följa praxis för att inte riskera lagbrott och hur denna försiktighet kan resultera i att de går miste om många positiva fördelar och inte kommer kunna dra full nytta av AI på samma sätt som andra:

"Man vågar inte göra någonting för att man är rädd för att bryta lag eller osäkerhet kring lag. Alla i offentlig sektor känns det som att vi sitter och väntar på att någon ska ta första steget och testa saker. För att man inte vet om man får göra så eller inte, får man däng för det eller inte och då gör man ingenting. Då får man ingen AI alls och då kommer vi inte klara våra utmaningar. [...] Jag tycker det är en risk om man inte använder AI faktiskt. Att vi inte kan lösa våra välfärdsutmaningar" - OR

En annan respondent inom offentlig sektor lyfte hur detta påverkar oss globalt som hänvisar till kommande EU-lagstiftning där länder såsom USA inte jobbar i samma utsträckning med att reglera AI:

"EU är ju jätteduktiga på att verkligen jobba med att reglera AI till skillnad från till exempel USA. Och det sätter ju oss i ett prekärt läge där vi då har en supermakt som ständigt ökar sin kompetens i området medan vi sätter lite krokben för oss själva i Europa med den lagstiftningen och regleringen som vi nu inför"

Trögrörlighet och svårigheter för snabb förändring var också faktorer som ansågs vara utmaningar i att möta rådande (och kommande) lagstiftning. En respondent inom den privata sektorn beskrev hur det alltid är utmanande att se till att ny lagstiftning etableras inom stora organisationer där det kommer att bli utmanande att få allt i ordning i tid. Några respondenter beskrev att deras juridiska team inte verkar hålla sig uppdaterade, att det uppstår utmaningar då de juridiska representanterna ska samarbeta med de tekniska samt att de idag har svårt att hantera rådande lagstiftning på grund av personal- och kompetensbrist och att detta inte förväntas bli lättare framöver då EU-lagstiftningen träder i kraft.

De som inte upplevde svårigheter att möta rådande lagstiftning och som inte uttryckte oro för framtiden var enbart några få respondenter inom den offentliga sektorn. Dessa uppgav att de

hade starka juridiska team, en bra hantering av rådande lagar och en bra kompetens och kunskap inom organisationen och upplevde därmed inte utmaningar med att möta kommande lagstiftning då de anser sig vara väl förberedda.

De flesta respondenter upplevde dock att den kommande EU-lagstiftningen kommer att vara svår att tyda initialt då där inte finns någon praxis och då eventuella fall kommer behöva prövas i EU-domstol. Dock verkar majoriteten vara överens om att denna utmaning kommer bli bättre med tiden allt eftersom vi drar lärdomar från händelser och motgångar. En respondent inom offentlig sektor föreslog att man kan använda AI för att underlätta tolkning av lagtext:

"[...] om vi skulle låta AI läsa igenom våra lagtexter så tror vi att de skulle kunna göra en derivering av lagtexterna rätt så ackurat. Så då kunde vi säga att AI skulle kunna fatta ett juridiskt korrekt beslut så att vi inte blir exponerade för några risker [...] den kanske skulle kunna vara mycket mer objektiv än vad vi är när vi läser in material."

EU AI Act

Utmaningar med denna lagstiftning som nämnts ovan samt som svar på hur lagstiftningen kommer påverka organisationerna samt om den medför utmaningar har bland annat varit risk för hämmande effekt av utveckling, att lagstiftningen riskerar krocka med annan lagstiftning samt att organisationer på grund av trögrörlighet i juridiska frågor och personalens kompetens kan uppleva utmaningar i implementering av lagstiftningen.

Potentiella utmaningar som lyfts under detta avsnitt men inte redovisats ovan var bland annat hur tolkningen av vilken riskklass man tillhör kan bli utmanande samt hur samarbete med andra länder som inte lyder under sådan lagstiftning kan bli problematisk eller att länder inom EU kommer tolka lagstiftningen olika. I den mån respondenterna kunde svara ombads de även ange vilken riskkategori de tror de kommer hamna under och här angav majoriteten att de troligen kommer tillhöra högrisk-kategorin, några att de troligen kommer ha AI som faller under samtliga kategorier och ingen angav att de kommer ha förbjuden AI. Då lagstiftningen inte trätt i kraft ännu var detta dock mest spekulationer och antaganden baserade på rådande information som finns tillgänglig vilket kan komma att ändras.

Några respondenter inom både privat och offentlig sektor uppgav att organisationen kommer behöva utveckla data- och AI-styrning för att säkerställa att lagstiftningen följs korrekt. Vidare är uppfattningen att lagstiftningen kommer att medföra ökat krav på dokumentering och rapportering, men flera respondenter inom både privat och offentlig sektor tycker idag att det behövs tydligare riktlinjer och ser positivt på den kommande lagstiftningen.

Två största riskerna idag

I slutet av intervjuerna då eventuella oklarheter kring frågorna hade klarats upp, fick respondenterna möjlighet att på nytt reflektera över vilka risker de ansåg vara mest betydande, för att fånga eventuellt förändrade åsikter. För att hålla svaren kortfattade fick respondenterna svara på vilka två risker de tror är mest närvarande idag samt vilka två de tror kommer vara störst i framtiden. Den största risken idag enligt respondenter från både offentlig och privat

sektor anses vara de som ingår i riskkategorien *Teknologiska, data- och analytiska*. Här svarade båda sektorerna relativt lika, se Figur 8 och 9.

Respondenterna från privat sektor rankade juridiska och regelmässiga risker som de näst största idag, följt av informativa och kommunikativa samt ekonomiska risker. Ingen av respondenterna uppgav att etiska och sociala risker hörde till de två mest framträdande riskerna i nuläget.

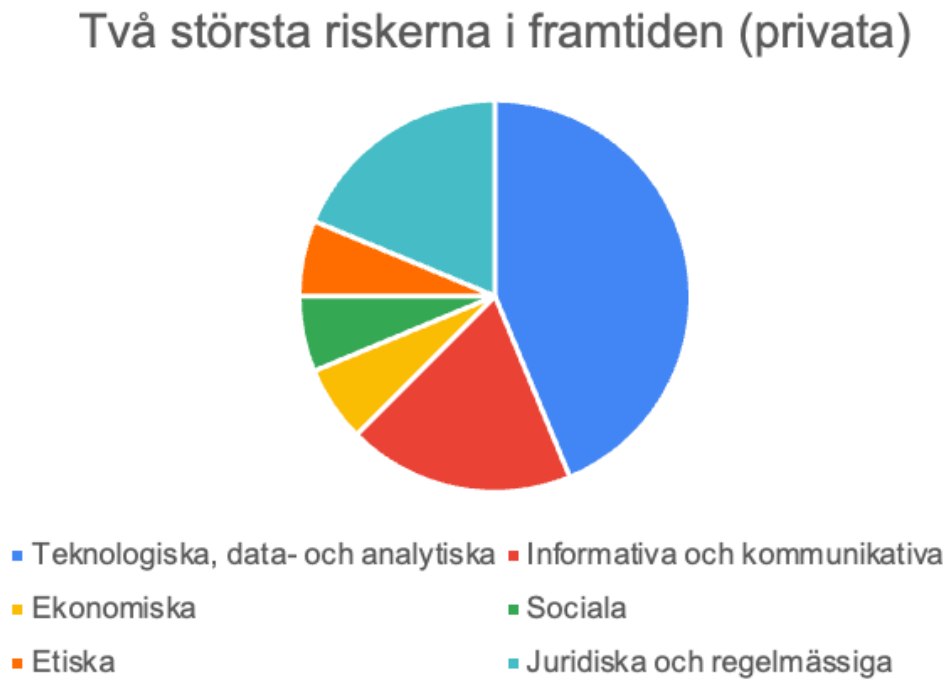
Respondenterna inom offentlig sektor ansåg att alla risker var närvarande idag. Precis som inom privat sektor, rankades juridiska och regelmässiga risker som näst störst (efter teknologiska risker) följt av informativa och kommunikativa, etiska, sociala samt ekonomiska risker.

Figur 8. Cirkeldiagram över de två största riskerna idag från privat sektor.

Figur 9. Cirkeldiagram över de två största riskerna idag från offentlig sektor.

Två största riskerna i framtiden

Flera respondenter från privat sektor förutspår att teknologiska, data- och analytiska risker kommer att vara bland de två största riskerna i framtiden. Inom offentlig sektor pekade många respondenter på informativa och kommunikativa risker som en av de framtida toppriskerna. Bland respondenterna från privata sektor var det också vanligt att nämna juridiska och regelmässiga risker som en av de två största, en kategori som inte ansågs lika framträdande av någon respondent inom den offentliga sektorn. Se Figur 10 och 11 för detaljerad jämförelse.



Figur 10. Cirkeldiagram över de två största riskerna i framtiden från privat sektor.

Figur 11. Cirkeldiagram över de två största riskerna i framtiden från offentlig sektor.

Irrelevanta riskkategorier i framtiden

Majoriteten av respondenterna från både privat och offentlig sektor poängterade att alla risker kommer vara relevanta i framtiden, men några respondenter lyfter främst tre riskkategorier som de tror kan spela mindre roll i framtiden: etiska, sociala samt juridiska och regelmässiga AI-risker. De *etiska* motiverades av respondenter från privat sektor inte utgöra en framtida risk på grund av organisationens branschtillhörighet, alternativt att verksamheten har strikta riktlinjer att utrymme inte finns för eskalerande etiska risker. Sist men inte minst nämndes att det genom AI kan komma en gemensam modell:

"Men skapar vi svenska modeller eller [...] mer geografiska EU-modeller kanske så kommer vi nog kunna hålla vad vi tycker är mest politiskt korrekt information. Så att ja, den kommer nog på sikt kanske försvinna" - PR

De *sociala* lyftes av respondenter från både privat och offentlig sektor, där en respondent från offentlig sektor menade att AI eventuellt kan fylla tomrum som människor upplever idag. Sist men inte minst lyftes även de *juridiska och regelmässiga* från enbart respondenter inom offentlig sektor, där ståndpunkten är att det antingen är ett område man har så pass bra koll på att det inte ska behöva bli en risk, alternativt att det kommer att klarna upp i framtiden.

Övriga risker

En respondent från privat sektor lyfte den dystopiska risken att AI kommer att utrota mänskligheten, och en respondent från offentlig sektor lyfte även frågan om hur man vet att AI företräder människor:

"[...]har vi lyckats skapa någonting som är intelligentare än oss själva så kommer den kunna skapa någonting som är intelligentare än oss själva och sen så till slut kommer människan att

vara som en myra i förhållande till den och undrar varför ska vi vara där. Och har man då inte lyckats och skapa ett mål för AI som är för mänskligheten goda på en otroligt bred front så kan ju den underminera för oss att fortsätta leva. De brukar ju ge sådana exempel som att ja men vi vill förhindra svält. Okej men om vi dödar mänskligheten så finns det ingen svält. De har uppnått sitt mål men kanske inte på det sättet som vi tänkte att de skulle göra.Och där håller de ju på och resonerar om hur i hela världen ska man formulera mål som gör att det blir allt igenom gott. Och den är ju otroligt svår och de har ju inte lyckats lösa den grejen än.“

En respondent inom offentlig sektor lyfte risken kopplat till “vad innebär det att vara människa?”, alternativt vad människovärdet är i en tid där teknisk utveckling tar över/ersätter mänskliga roller allt mer. En annan respondent inom offentlig sektor menade även att den största risken är att vi inte gör någonting med AI-tekniken, och att det finns en slags övertro till att människor alltid gör rätt, trots att individuella bedömningar kan skilja sig markant. Sist men inte minst nämnde även en respondent inom offentlig sektor risken att låta stora aktörer (likt OpenAI och andra teknikjättar) få för mycket inflytande och makt över AI-utvecklingen vilket kommer påverka dess utveckling samt de som försöker ta sig in på marknaden.

“På lång sikt kan man ju mycket väl tänka sig att det till exempel blir så att det blir precis som Google har varit väldigt dominerade i sökmotorer så kommer det vara en risk att vi hamnar i ett läge där några få väldigt stora AI-operatörer har ett otroligt stort inflytande i världen. Och vad är det för risk? Jag vet inte vilken risk det är [...] Att vi hamnar liksom i greppet på de här OpenAI och så vidare. Ännu mer än vi har varit tidigare. För att de här när man börjar bygga in AI, när man börjar bygga samhällsprocesser på AI så kommer vi vara helt fast och inte kunna backa längre. [...]“ - OR

4.2.4 Risker på samhällsnivå (perspektiv från DIGG och MSB)

Efter att två respondenter intervjuats gjordes ett medvetet val att kategorisera deras svar separat. Dessa respondenter arbetar med AI i stor skala, där de assisterar med eller bedömer risker ur ett bredare perspektiv. Därför tematiserades svaren under den egna kategorin "Risker på samhällsnivå". Detta då svaren inte speglade användningen inom den egna organisationen, utan snarare användningen och riskexponeringen bland andra organisationer som de arbetat med eller baserat på deras uppfattning och kunskaper kring AI i omvärlden. Respondenterna gav sitt godkännande till att bli avanonymiserade, eftersom deras bakgrunder anses vara avgörande för att förstå de samhällsomfattande aspekterna av ämnet.

DIGG är en relativt nystartad myndighet vars roll i stora drag är att stödja den offentliga förvaltningens digitalisering, samt att stödja eller rådgöra regeringen i digitaliseringspolitiken. Representanten från DIGG heter Mats Snäll, med bakgrund som jurist, men även lång erfarenhet inom digitalisering. Han har även tidigare ansvarat för den avdelning inom DIGG som ansvarar för framkantsteknologier och utforskande verksamhet. Idag arbetar Snäll som senior rådgivare med uppdrag att se över DIGGs AI-utveckling samt verksamhetsutvecklingen inom den offentliga förvaltningens digitalisering där AI-frågor utgör ett delmoment.

Representanten från MSB heter Joachim Elevant och arbetar som handläggare inom enheten Säkerhet och cyberfysiska system, med ansvarsområdena energi, hälso- och sjukvård och artificiell intelligens. Hans roll innebär en stödjande funktion till samhällsviktiga verksamheter inom området cybersäkerhet.

Snäll framhöll att förtroende är en av de största samhällsutmaningarna relaterade till AI. Under de inledande faserna av implementeringen är det avgörande att säkerställa att allmänheten känner sig trygg och har förtroende för AI-systemet. Om detta inte uppnås kan det hämma systemets utveckling och leda till möjliga förluster inom de områden där AI annars kan ha positiva effekter. DIGG har bland annat tagit fram en förtroendemodell med fokus på att stärka förtroendet för AI-system. Snäll lyfte också juridiskt ansvar och hur utmaningar med detta bör hanteras, samt att detta anknyter starkt till vikten av förtroende hos framför allt offentliga organisationer och rättssystemet i stort:

"Men jag tror inte att man ska tillförskriva [AI] något ansvar i sig utan det måste ju finnas någon som backar upp det ansvaret. Och för offentliga organisationer är det här helt avgörande och för ett helt rättssystem är det helt avgörande. I en rättsstat, demokratiskt system så är ju det ansvar, det strikta ansvar som vi tar på oss för våra beslut är helt avgörande för att man ska tilltro till systemet."

Enligt Snäll blir även förtroende aktuellt gällande ekonomiska risker, eftersom ett minskat förtroende i slutändan kan resultera i en ekonomisk skada. Snäll kommenterar även EU AI Act och hur denna kan komma bli svår att bedöma, vilket i sin tur kan göra att organisationer nedvärderar sina system för att komma undan krav. Han varnar också för risken att vissa organisationer kan bli högre klassade än vad de borde, vilket med tanke på de höga bötesbeloppen kan orsaka betydande skador. Elevant poängterade även hur det i EU AI Act kommer finnas krav på att kunna tillhandahålla en förklaring för AI-systemets handling, vilket kommer utgöra en stor utmaning för verksamheter. Här lyfts även problematiken med komplexiteten i AI-system där Snäll lyfte att de offentliga organisationerna här kan ha större utmaningar eftersom de troligen till högre grad än de privata kommer behöva köpa in externa tjänster istället för att utveckla egna. Enligt Elevant tar denna problematik ytterligare en dimension då AI börjar användas mer i kritisk infrastruktur, och komplexiteten i AI-systemen utgör därmed en svårighet för samhället i stort. Han menar även på att i takt med att AI-systemen blir fler och komplexare ökar risken, och man kommer behöva ha en viss kontroll över komplexiteten.

Vad gäller den ekonomiska risken ansåg Elevant att "[...]AI har större potential att bidra till ekonomisk utveckling än att den bidrar till ekonomiska risker för landet". En framtidsvision gällande de sociala riskerna från Elevants sida var att AI kommer kunna skapas i ganska god balans med hur människor vill ha det.

Både Snäll och Elevant pekade ut *teknologiska, data- och analytiska* samt *informativa och kommunikativa* AI-risker som de största riskerna i samhället i dagsläget. Både Snäll och Elevant pratade även om svårigheten med AI-manipulerad information, som skapar stor osäkerhet i samhället. Anledningen till att Elevant tyckte att dessa två riskkategorier var störst

är för att AI-utvecklingen är en "teknisk utveckling", samt att informationsmanipulation idag redan är ett stort problem i samhället.

Framgent menade Elevant att *"[...]de allvarliga konsekvenserna går att hitta i kategorier som är mycket närmre människor. Men källan till problemen ligger mycket i tekniken".*

På längre sikt menade Snäll att man kanske bör reflektera över människans roll i framtiden och hur man kommer påverkas av ett AI-samhälle. Risken är att man kanske tappar bort denna fråga då man istället fokuserar på de mest framträdande riskerna, såsom teknologiska och informativa risker.

Både Snäll och Elevant lyfte att det juridiska riskfokuset är för starkt. Snäll poängterade att de juridiska och regelmässiga (samt etiska) riskerna får mycket mer uppmärksamhet i allmänhet, eftersom samhället i stort förstår problematiken och därför alltid landar i de frågorna. Snäll påpekade att det finns en stark fokusering på riskerna med AI idag, men underströk att det också är viktigt att inte förbise de möjligheter som AI erbjuder, vilka lätt kan hamna i skymundan på grund av denna riskcentrering. I viss mån menade även Snäll att AI kommer vara helt avgörande - exempelvis inom vården där AI kan effektivisera och täcka upp för personalbristen som råder. Elevant lyfte även att AI kommer kunna hjälpa oss motarbeta vissa risker framåt.

4.2.5 Personlig inställning till AI

Som avslutande del på intervjuerna frågades respondenterna om deras inställning till AI var positiv eller negativ. Överlag rådde det en positiv inställning bland samtliga respondenter, där det sågs enorm potential och förbättringsmöjligheter. Däremot påpekade de flesta av respondenterna även respekt, ödmjukhet alternativt oro inför (framförallt) framtida svårigheter och risker, främst på samhällsnivå. En respondent inom offentlig sektor menade att man inte kan avgöra huruvida man är positiv eller negativ till AI i dagsläget, för att man helt enkelt inte vet tillräckligt. En respondent inom privat sektor placerade sig på ytterligheterna, dvs både extremt positiv och rädd", med motiveringen att AI kan ses som ett verktyg som vid "rätt" användning är jättebra men det blir farligt om det används på ett felaktigt, "dumt" sätt, en åsikt som delas av flera respondenter från båda sektorerna.

> *"Jag tycker teknologin är positiv men jag är orolig för hur den ska användas. För att all teknologi eller all typ av saker kan ju användas för onda avsikter av antagonister" - PR*

En annan inställning som framkom från både privata och offentliga respondenter är att man inte har något val eller alternativ gällande AI:

> *"[...]det är som en kraft som inte går att stoppa lite grann. Man får istället försöka fokusera på hur kan vi göra det bästa av situationen" - PR*

Både representanterna från DIGG och MSB hade en generellt positiv inställning till AI, men Snäll är rädd att pendeln väger över till att bli negativ och återhållsam bland gemene man.

5 Diskussion

Syftet med studien var att undersöka hur samhällsviktiga verksamheter i Sverige upplever sig exponerade, både idag och i framtiden, för de sex största AI-riskerna inom litteraturen, samt om andra risker förekommer. Syftet var även att undersöka eventuella skillnader mellan privat och offentlig sektor. Nedan redovisas de essentiella delarna som lyfts av litteraturen och respondenterna i studien, kring de sex olika riskerna. Skillnader mellan offentlig och privat sektor redovisas löpande under varje riskkategori för de risker där störst skillnad förelåg.

5.1 Teknologiska, data- och analytiska

Denna riskkategorins frågor berörde till viss del samma områden med olika vinkling. På frågan rörande cyberangrepp eller -säkerhetsincidenter ansågs risken vara lägst. Spännande nog beskrivs detta vara en av de stora riskerna med AI (Marr, 2023; Nigmatov & Pradeep, 2023), men respondenterna verkar inte känna sig speciellt exponerade för detta idag. Däremot ansåg respondenterna att AI förstärker hotet utifrån, där tekniken underlättar för antagonister att angripa organisationerna. Resultaten stämmer med annan forskning som funnit att detta är en av utmaningarna med att AI är en dual-use teknologi och därmed tillgänglig för alla (World Economic Forum, 2023). Denna rädsla väcks möjligtvis av att man inte vet kapaciteten hos dessa antagonister eller att man inte vet exakt vilka förmågor AI har. Detta understryker behovet av fortsatt forskning och utveckling inom AI-säkerhet för att stärka skyddet mot och förståelsen av sådana hot.

Ett intressant fynd var att AI kan användas för att motverka cyberincidenter och därmed bidra till en förbättrad cyberförmåga. Detta överensstämmer med vad Jawhar et al. (2024) och Li (2018) fann, att AI förbättrar förståelsen, utredningen och utvärderingen av cyberhot. Dock är AI-system enligt Li (2018) också sårbara för cyberhot som kan störa urvals-, inlärnings- och beslutsprocesser. Dessa resultat kräver därför en försiktig tolkning, eftersom det finns både positiva och negativa aspekter gällande användningen av AI inom cybersäkerhet.

Den risken som ansågs vara absolut högst var över- eller underskattning alternativt missbedömning av AI-verktyg, där framförallt överskattning och felaktig hantering var den största problematiken. Resultaten från denna studie tycks överensstämma med annan forskning som funnit att det finns en tendens till att överskatta AI (Keding & Messinger, 2021). Faktorer som överskattning och missbedömning ses även i svaren på frågan gällande spridning av data, som var en av konsekvenserna av att anställda felaktigt använde AI-verktyg. Återkommande nämndes hur anställda riskerar läcka data via externa tjänster, och här ansågs generativ AI vara den mest bidragande sortens AI. Detta bekräftar också "Shadow AI"-problematiken som Campbell och Jovanovich (2024) beskriver och verkar således vara närvarande inom samhällsviktiga verksamheter.

Det verkar finnas problematik gällande förtroendet vid AI-människa-interaktioner, där människors tendens till överskattning orsakar organisatorisk skada. Detta är ett intressant fynd som kan ha flera orsaker. En potentiell förklaring är att människor med stort förtroende till sin egen prestation (även kallat Dunning-Kruger-effekten) tenderar att förlita sig mindre på AI-

system (He et al., 2023). Möjligtvis gäller därför det omvända scenariot, att anställda med lägre AI-kompetens inte har så stort förtroende för sin egen förmåga, och därför förlitar sig mer på AI-system. Denna bristande kompetens är inte förvånande med tanke på teknologins snabba utveckling och frånvaron av formell utbildning inom området på många arbetsplatser. Resultaten pekar på ett behov av förbättrad AI-kompetens bland anställda, vilket bör ses som en organisatorisk prioritet för att förhindra oönskad spridning av data och stärka förtroendet för AI-användningen. Även om kompetenshöjning kan bidra till att minimera denna risk, visar forskning att problemet kan vara mer komplicerat än så.

Prest (2023) lyfter problemen med att människor förmänskligar AI, vilket kan skapa utmaningar vid anställdas hantering av tekniken, trots kompetenshöjande åtgärder. Även Schneier (2023) menar att detta är särskilt påtagligt med generativ AI, eftersom den uppvisar de mest "mänskliga" egenskaperna. Människor tenderar att behandla AI-assistenter som betrodda vänner eftersom kommunikationen sinsemellan sker med ett naturligt språk, vilket oundvikligen leder till att de tillskrivs mänskliga egenskaper (Schneier, 2023). En framtida utmaning kan därför bli hur man ska hantera denna övertro eller risk för missbedömning, då risken är att anställda, men också samhället i stort, börjar förlita sig blint på AI. Det är avgörande att anställda lär sig att AI inte besitter mänsklig intuition eller omdöme och att de tekniska besluten alltid bör granskas kritiskt. Utbildning med sådan inriktning skulle inte bara stärka teknisk kompetens utan även förmedla en kritisk medvetenhet om de psykologiska effekterna av interaktioner med AI, vilket är avgörande för att säkra en ansvarsfull och informerad användning av teknologin.

Litteraturen belyser "black-box"-problematiken med AI-system som på grund av ökad komplexitet försvårar insyn. Ett intressant fynd är att trots att "black box"-fenomenet framhävs som en av de största riskerna med komplex AI-teknologi, verkar många respondenter uppleva riskerna relaterade till systemets komplexitet som relativt låga idag. Detta kan bero på den allmänt låga implementeringsgrad av högkomplexa, lågtransparanta AI-system bland organisationer idag. Det är dock tydligt att problemen med transparens är en avgörande fråga, då dessa svårigheter identifieras som en kritisk riskfaktor för framtiden, särskilt gällande generativ AI. Generativ AI är en av de teknologier som många ser enormt stor potential med och sådana system kommer troligen implementeras inom samhällsviktiga verksamheter som är i stort behov av att effektivisera exempelvis inom sjukvård, utbildning och infrastruktur. Ett sätt att möta black-box problematiken verkar vara att upprätthålla en hög grad av transparens och spårbarhet i AI-system. Detta innebär att varje beslut kan spåras och granskas, och att det finns tydliga protokoll och riktlinjer för hur data samlas in, används och skyddas. Genom att säkerställa transparenta system kan samhällsviktiga verksamheter dra full nytta av AI-teknologins potential utan att kompromissa med säkerhet eller förtroende.

För att lyckas med detta pekar resultaten från denna studie på ett specifikt område inom AI, nämligen explainable AI (också kallat XAI). Denna metod fokuserar på att genom hela processen vidhålla transparens och spårbarhet (Holzinger et al., 2018). Explainable AI kan således bidra med att stärka förtroendet för AI-systemen.

5.2 Informativa och kommunikativa

Resultaten indikerar att offentliga organisationer kan vara mer sårbara på organisationsnivå. Respondenterna från offentlig sektor uppfattade att denna riskkategori i framtiden kommer vara större jämfört med de från privat sektor. Denna skillnad kan bero på att respondenter från offentlig sektor har erfarenheter av att ha mottagit förfalskade underlag, samt att det finns ett ökat intresse från antagonister att rikta sina attacker mot offentlig sektor.

Problematiken beskriven inom denna riskkategori är inte nödvändigtvis associerad direkt till AI, utan ses snarare som ett redan befintligt samhällsproblem. Wirtz et al. (2022) analyserar även dessa risker främst från ett samhällsperspektiv, vilket även reflekteras i respondenternas svar som ofta gavs utifrån en samhällssynvinkel. Fenomen som "fake news" och "deep fakes" som beskrivits av Hammonds (2023) och Srivastava 2024) anses bidra till en generell misstro mot information i samhället. Wagner och Blewer (2019) uttrycker detta som att "[...]we may exist in a moment in which non-reality is as real as reality".

Utmaningen ligger i hur AI, särskilt generativ AI, förvärrar denna redan etablerade problematik genom att ytterligare underlätta skapandet av falskt material. För samhällsviktiga verksamheter, framförallt bland de som ansvarar för informationsutgivning, skapar denna misstro till information stora utmaningar. Detta kan delvis förklara varför respondenter inom offentlig sektor ansåg att denna risk kommer vara den största i framtiden, då flera har sådan verksamhet med uppgift att informera allmänheten, framförallt i situationer av kris där misstro till information kan ha förödande konsekvenser. Pinnell (2024) beskriver hur AI möjliggör sofistikerade deep fakes och hur dessa underminerar information, vilket i sin tur kan ha negativa effekter på allmänhetens förtroende för institutioner, källor och fakta. En andra sida av en tid med desinformation, som författaren väljer att kalla "lögnarens utdelning" beskriver hur förekomsten och risken för syntetiska medier, som deep fakes, kan missbrukas för att underminera trovärdigheten hos äkta information. Pinell (2024) understryker att det krävs förberedelse, samarbete och kunskap från olika samhällsaktörer för att hantera problematiken. Ett sätt att göra detta är att fokusera på spårbarheten av informationen, vilket åter knyter an till behoven av att vidhålla transparens som nämns ovan under teknologiska AI risker. Istället för att tala om för allmänheten vad man ska lita på och vad man inte ska lita på, bör man tillhandahålla transparent information för att låta publiken bestämma sig själv.

Det är avgörande att förstå och adressera den underliggande misstron mot information som teknologin kan förstärka. Trots att riskkategorin enligt Wirtz et al. (2022) var en av de mindre utforskade områdena tyder studiens resultat på att den utgör en av samhällets största utmaningar både idag men framför allt i framtiden. Detta är en oroande utveckling som bör tas på stort allvar. Det krävs därför ett fortsatt samarbete mellan olika samhällsaktörer, inte minst mellan privata och offentliga aktörer, och en ökad satsning på utbildning, transparens och spårbarhet av information för att publiken själva ska kunna göra informerade bedömningar om informationens tillförlitlighet. Denna proaktiva och inkluderande strategi är avgörande för att mitigera riskerna med AI och säkerställa en trygg och informerad allmänhet.

5.3 Ekonomiska

Denna kategori var en av de lägre rankade, och respondenterna menar att risken för ekonomisk förlust vid avbrott som Campbell och Jovanovic (2024) beskrivit är låg, eftersom beroendet av AI-systemen hittills inte gör organisationerna mer sårbara. Detta kan däremot komma att öka framgent när beroendet successivt ökar. Den ekonomiska riskkategorin är bland den minst utforskade enligt Wirtz et al. (2022), vilket också kan förklara varför den ekonomiska risken inte ansågs vara unik för AI-system utan snarare kunde likställas med traditionella IT-system som organisationerna generellt är beroende av.

Den enda ekonomiska risken som direkt kunde kopplas till AI var istället risken för de straffavgifter/sanktioner som AI-orsakade fel kan medföra efter införanden av EU AI Act, vilket beroende på överträdelsen och företagets storlek kan ligga på belopp mellan 35 miljoner euro (eller 7 % av den globala omsättningen) till 7,5 miljoner euro (eller 1,5 % av omsättningen) (Europaparlamentet, 2023d).

Litteraturen har visat att AI kommer kunna effektivisera och öka produktionen hos många organisationer. Resultaten från denna studie indikerar en risk för förlust av sådan möjlighet, då organisationer i rädsla för att råka göra fel kanske väljer att avstå från att använda AI. Detta understryker vikten av att framtida lagstiftning måste vara klar och tydlig, för att organisationer ska kunna avgöra när ett fel faktiskt begås. Detta är en betydande utmaning eftersom denna process fortfarande är i ett tidigt skede, utan etablerad praxis eller tidigare erfarenheter att förlita sig på.

5.4 Sociala

Wirtz et al. (2022) beskriver denna risk som arbetslöshet till följd av AI-automatiserade processer och motståndet som uppstår mot AI bland anställda till följd av detta. Studiens respondenter verkar inte uppleva särskilt stor oro, eftersom riskkategorin blev lägst rankad av samtliga kategorier. Den låga rankningen kan kopplas till det faktum att yrkesgrupper som Hammond (2023) pekat ut som särskilt sårbara inte var markant representerade i denna studie, då respondenterna var nischade mot AI. Vidare kunde inte dessa respondenter svara för potentiellt missnöje bland övriga anställda, vilket också märktes då respondenter med mer av en chefsposition eller med en roll som gav dem ett bredare organisatoriskt perspektiv gav mer välutvecklade svar.

Resultat från denna studie går alltså inte i linje med tidigare forskning kring ämnet som identifierar detta som en av de stora riskerna med AI. Att respondenterna är en dålig representation av de yrkesgrupper som kanske känner sig mest drabbade kan vara en orsak, men det finns också viss osäkerhet i området i sig. Hur pass stor inverkan AI som fenomen kommer ha på yrken och individer är idag svårt att avgöra. Detta belyser Khogali och Mekid (2023) vars studie presenterar en översikt över hur automatisering och AI kan påverka företag och jobb. En av deras slutsatser är just att det idag är omöjligt att veta effekterna och att AI:s påverkan på mänskligt liv är ett brett forskningsområde. AI på arbetsplatsen kan underlätta, förenkla och fria människor från andra arbetsuppgifter och AI jämförs med tidigare framsteg inom automatisering som höjt produksitonsstandader, arbetsspecialisering och värdet av

“mänskliga egenskaper” som kreativitet, problemlösning och matematisk skicklighet. Det lyfts hur det idag finns lite bevis för att AI-verkligen kan ersätta människor och författarna anser att AI, då det är en kärnkomponent i datorinlärning, är avgörande för mänskligheten framtid (Khogali & Mekid, 2023). Effekterna av AI är således fortfarande osäkra, och därmed kan det ändå vara viktigt för arbetsgivare att beakta de sociala riskerna vid implementering av AI i organisationer.

5.5 Etiska

När det gäller de största upplevda riskerna idag, rankas etiska risker överraskande lågt, där ingen av respondenterna från privat sektor identifierade etiska risker som en betydande oro, samtidigt som etiska bland offentlig sektor utgjorde en av de minst prioriterade riskerna. Respondenterna uttryckte en hög medvetenhet om dessa risker, inklusive kunskap om hur AI kan innehålla inbyggda fördomar, vilket stämmer överens med tidigare litteratur. Wirtz et al. (2022) beskriver även detta som den mest debatterade risken inom vetenskaplig litteratur. Det faktum att etik är ett flitigt diskuterat ämne reflekteras också i att många respondenter redan har implementerat policys och ramverk för att hantera dessa frågor inom organisationerna, vilket delvis kan förklara den låga rankingen av denna risk, eftersom mitigeringsåtgärder redan är integrerade. En annan förklaring till varför denna risk betraktas som låg kan vara då många respondenter inte använder AI på sådant sätt som anses innebära höga etiska risker enligt litteraturen, som i rekryteringsprocesser, låneansökningar eller bildigenkänning (Europakommissionen, 2020; Nigmatov & Pradeep, 2023).

Bland vissa respondenter förväntas denna risk öka i framtiden, där särskilt generativ AI utgör en utmaning, eftersom sådana system kan ha inbyggda fördomar som blir svåra att identifiera på grund av systemets komplexitet och brist på transparens, där AI:n blir en “black box” (Campbell & Jovanovic, 2024). Som tidigare litteratur även påpekat finns här en risk att AI självständigt skapar diskriminerande parametrar (Europakommissionen, 2020; Du & Yuan, 2024). Resultaten bekräftar denna problematik där oro finns för att oavsiktligt bryta mot lagar, exempelvis diskrimineringslagstiftningen. Problem likt dessa kan komma att öka då såväl teoretiska som empiriska resultat tyder på att etiska risker kan förstärkas i takt med att AI-systemen blir allt mer komplexa. Ett förslag på hur man kan bli bättre på att tackla dessa “black-box”-risker är genom utbildning och en respondent exemplifierade hur dennes organisation satt ihop ett etiskt råd för att adressera etiska frågeställningar i sina AI-system.

De respondenter som inte anser att risken kommer att öka betonar AI Act som ett medel för att minska diskriminering, eftersom denna lagstiftning behandlar etiska frågor. Dessutom kan risken minskas genom att AI potentiellt kan identifiera fördomar som människor kanske inte är medvetna om, vilket överensstämmer med vad Srivastava (2024) och Statskontoret (2024) påpekar. AI har således potential att uppmärksamma omedvetna fördomar som vi idag inte ser.

Resultaten indikerar därmed att det inte finns ett entydigt svar på om AI är negativt ur ett etiskt perspektiv, även om den rådande litteraturen ofta hävdar detta.

Ett aspekt som inte har behandlats i denna studie, men som framkommer i litteraturen, är risken att vi *inte* lyckas införliva mänskliga värderingar i AI-system (Wirtz et al., 2022). Konsekvensen av detta kan vara att AI-systemen inte innehåller de etiska aspekter som är inneboende i mänskligt beteende. I en dystopisk framställning kan detta leda till det scenario som Bales et al. (2024) beskriver, där AI, om den inte är i harmoni med mänskliga värderingar, kan utgöra ett hot mot mänskligheten.

Denna risk är frekvent diskuterad och etik är ständigt aktuellt gällande AI-system. Resultaten från aktuell studie belyser flera dimensioner av detta, där AI inte bara är ett verktyg som riskerar diskriminera utan också kan utgöra en resurs för att hantera diskriminering. Det kan vara värt att överväga om denna riskkategori tillmäts för stor uppmärksamhet och om det finns en risk att andra, potentiellt mer samhällspåverkande risker, förbises som ett resultat.

5.6 Juridiska och regelmässiga

Tidigare litteratur belyser problem med att det råder otydligheter kring ansvar gällande juridik och AI-system samt hur tvister kring AI-orsakade fel ska hanteras (Wirtz et al, 2022; Loudiyi, 2021; Europeiska kommissionen, 2020; Jarvenpaa et al., 2023; Berggren et al., 2023). Detta kan förklara varför respondenter inom både offentlig och privat sektor rankade denna risk som en av de två största riskerna idag. Dessutom verkar de juridiska riskerna vara tätt sammankopplade med flera andra riskområden, vilket innebär att de även utgör stora integrationsutmaningar.

Resultaten visar att det finns oro för juridiska och etiska övertramp vid användning av AI-system genom exempelvis oavsiktlig diskriminering, vilket härsammar från riskerna associerade med komplexitet, där brist på transparens försvårar insynen i huruvida algoritmerna skapat diskriminerande beslutsprocesser. Detta försvårar även frågor om juridiskt ansvar och spårbarhet i AI-system. Resultaten visade också en rädsla för ekonomisk förlust, till följd av oavsiktligt brott mot kommande EU-lagstiftning. Svårigheter i att möta kommande lagstiftning, som primärt gäller AI Act, upplevdes mer närvarande hos respondenter inom privat sektor, som angav risken som den näst största i framtiden, medan respondenter inom den offentliga sektorn inte alls angav juridiska och regelmässiga som en framtida risk.

Brist på kunskap anses vara den faktor som ökar den juridiska risken mest enligt både respondenter och litteratur (Jarvenpaa et al., 2023; Berggren et al., 2023). Detta kan förklara varför det förelåg skillnad mellan privat och offentlig sektor, där respondenter inom offentlig sektor angav att de har god kompetens inom organisationen, vilken i sin tur förklarar varför de ansåg denna risk vara nästintill obefintlig i framtiden.

Orsaken till att offentliga respondenter bedömer risken som låg i framtiden kan även vara samma som väcker deras oro för att bli begränsade av framtida lagstiftning (EU AI Act). Den juridiska apparaten inom offentliga organ uttalas vara mer omfattande och mindre flexibel, vilket möjligtvis indikerar att de offentliga aktörerna redan hanterar komplexa, juridiska frågeställningar dagligen. Detta talar för en högre juridisk kompetens och kan orsaka att den kommande lagstiftningen inte upplevs som en överväldigande utmaning. Dock är det en

utbredd oro bland dessa respondenter att de potentiella fördelarna med AI kan gå förlorade på grund av den omfattande juridiska belastningen som riskerar att antingen fördröja eller helt stoppa implementeringen av ny teknik. AI inom offentliga organisationer beskrivs av Neumann et al. (2024) som ett "dubbeleggat svärd", där ena sidan är teknikens otroliga potential att förbättra det interna arbetet och samtidigt förbättra kvaliteten på de offentliga tjänsterna. Den andra sidan är istället hur implementering av AI är mer komplex än andra IT-innovationer, och att de offentliga verksamheterna kommer att stöta på sektor-specifika problem. Det rekommenderas att AI används för rätt ändamål, samt att offentliga verksamheter inte ska tappa viktiga delar av AI som ansvarsskyldighet och rättvisa (Neumann et al., 2024).

Oavsett om AI-implementering ska ske inom offentlig eller privat sektor är det troligtvis typen av data som är den främsta faktorn kopplat till juridiska begränsningar. Mellan offentliga och privata aktörer finns det ofta stora skillnader i vilken typ av data som används för att träna AI-systemen. Offentliga verksamheter hanterar i regel data som innehåller personuppgifter, vilket innebär att de måste navigera i ett komplext landskap av dataskyddslagstiftning. I kontrast till detta äger ofta privata aktörer sin egen data, som används för att träna deras AI-system. Denna data kan vara mer generell objektinformation snarare än känslig individinformation, vilket kan ge privata företag större flexibilitet och färre juridiska hinder jämfört med offentliga verksamheter. Detta är dock starkt beroende av vilken typ av verksamhet som bedrivs inom de berörda organisationerna, och det är inte möjligt att göra en entydig åtskillnad mellan vilka typer av data som hanteras inom offentliga och privata sektorer. Följaktligen är det verksamheter vars data påverkas av ett flertal lagstiftningar som sannolikt kommer att stå inför de största utmaningarna.

Resultaten indikerar att det för den offentliga sektorn kan uppstå hämmande effekter vid AI-implementering på grund av organisationernas administrativa tröghet och strikta lagkrav. Men litteraturen talar också för en stark vilja i samhället att investera i AI inom den offentliga sektorn, och DIGG har exempelvis grundats med målet att underlätta digitaliseringen av offentlig verksamhet. Dessutom finns en tydlig ambition från regeringen att främja utvecklingen och användningen av AI (Regeringskansliet, 2023). Dessa initiativ speglar en önskan att övervinna strukturella hinder och dra nytta av de möjligheter AI erbjuder.

Utmaningarna med AI ligger till stor del i dess snabba utvecklingstakt, vilket utgör betydande hinder för lagstiftare, och enligt Larsson (2023) är det detta "taktproblem" som utmanar genomförandet av EU AI Act och integrationen av generativ AI. Författaren betonar behovet av ökad flexibilitet i lagstiftningen för att hantera den hastiga teknologiska framfarten. En lösning som föreslås är införandet av regulatoriska sandlådor, som innebär möjliggörande av experiment med ny teknik under kontrollerade förhållanden, även om detta tillvägagångssätt fortfarande har sina brister och behöver utvecklas. Vikten av att myndigheterna, om de planerar att nyttja dessa sandlådor, börjar planeringen omgående understryks också (Larsson, 2023).

Ett annat sätt att adressera taktproblemet är att tillsynsmyndigheterna tillhandahåller praktisk juridisk vägledning. Enligt Statskontoret (2024) råder det dock brister i teknisk och juridisk AI-kompetens bland statliga myndigheter, vilket försvårar situationen. Vidare är

myndighetssamverkan kritisk då AI-tillämpningar och regleringar ofta involverar flera olika områden. Denna samverkan är nödvändig för att effektivt hantera AI:s komplexa och tvärvetenskapliga natur.

Kompetensbrist har konsekvent identifierats som en av de största utmaningarna med AI-implementering inom den offentliga sektorn, enligt både respondenter och litteratur (SCB, 2023; Statskontoret, 2024). Wirtz et al. (2024) beskrev även brist på AI-talang/-strategi som en organisatorisk risk. I framtiden kan den största utmaningen för framgångsrik implementering av AI-system inom både offentlig och privata sektor handla om var den mest kvalificerade personalen finns, ett lopp som privat sektor verkar leda. En rapport från SCB (2016) visar på betydande löneskillnader mellan offentlig och privat sektor bland yrkesgrupper som dataspecialister, datatekniker och jurister, där den privata sektorn erbjuder högre löner. Återigen kommer samarbete och kunskapsdelning mellan sektorerna bli viktigt för att mitigera risker, se till att Sverige når sina mål om att vara en aktiv del i den fortsatta utvecklingen och användningen av AI och kanske även för att förhindra att ojämlikhet uppstår där få stora företag eller aktörer får kontroll över AI-utvecklingen.

Även om privata och offentliga verksamheter till viss del hanterar AI olika, så finns det också många likheter mellan sektorerna. Riskerna anses i stor utsträckning vara ungefär lika stora i dagsläget, och liknande risker med AI lyfts. Flera av riskkategorierna är de även överens om kommer bli större i framtiden. Medeiros (2020) understryker att det privat-offentliga samarbetet gällande AI är avgörande för att skapa och möjliggöra styrning som både är innovativ och anpassningsbar i takt med teknikens utveckling, för att säkerställa mänskliga rättigheter och sociala värden, och samtidigt främja innovation och kommersiell tillämpning.

5.7 Generativ AI

Många respondenter lyfte att just generativ AI kommer öka risken markant framöver inom nästan alla riskkategorier, mycket på grund av komplexitets- och transparenssvårigheter. Inom etiska riskkategorin yttrycktes oro för att sådan teknologi självmant konstruerar diskriminerande algoritmer, som sedan spillde över på juridiska och regelmässiga med risk för att oavsiktligen bryta mot lagar, vilket i sin tur spillde över på ekonomiska med risk att få böter då man oavsitkligen bryter mot lagar. Detta belyser tydligt hur sammankopplade och flerdimensionella dessa risker är.

Resultaten visar vidare en ökad oro för att anställda, genom användning av generativ AI, sprider känslig/personlig data. Inom informativa och kommunikativa ansågs bland annat denna teknologi kunna leda till bättre “deep fakes” och “fake news”. Resultatet ger därmed en tydlig indikation på att denna typ av AI kommer bidra till att nästan samtliga risker ökar.

Samtidigt visar också denna teknologi enorm potential. Generativ AI beskrivs kunna hjälpa verksamheter att exempelvis omformulera, förenkla och individualisera sina kundupplevelser (Harvard Business Review et al., 2024). Många respondenter angav också att de funderade på att implementera generativ AI, dels för effektiviseringspotentialen men också för att eventuellt avvärja anställdas användning av externa (generativa) AI-tjänster. Oavsett så kommer organisationer behöva ta ställning till hur de ska hantera den generativa AI:ns intåg,

vare sig det är på samhällsnivå eller inom den egna organisationen. Rekommendationen som Harvard Business Review et al. (2024) ger gällande vad för generativ AI en verksamhet ska implementera är att väga riskerna (hur sannolikt samt skadligt är det att osanningar och felaktigheter genereras och sprids?) mot efterfrågan (vad är det faktiska och hållbara behovet, bortom den nuvarande “hypen”?). På samma sätt behöver organisationer som är rädda för ekonomisk förlust till följd av implementering av AI-system också överväga riskerna mot möjligheterna.

6 Slutsats

Resultaten indikerar att studien framgångsrikt har lyckats identifiera flera av de risker som för närvarande anses vara särskilt relevanta inom området AI baserat på respondenternas svar. Dessa risker betraktas för närvarande som relativt låga bland de samhällsviktiga verksamheter som ingick i denna studie. Flera av riskerna förväntas öka i framtiden på grund av större komplexitet och minskad transparens, inte minst till följd av introduktionen av generativ AI. Offentlig och privat sektor uppvisar i stort mycket likheter för hur pass exponerade de är av dessa risker idag, men även vissa skillnader i uppfattningen om vilka risker som kommer att bli mest framträdande i framtiden, där framförallt juridiska och regelmässiga ansågs vara en större utmaning för privata respondenter medans informativa och kommunikativa ansågs vara störst inom den offentliga sektorn.

AI och dess associerade risker är extremt mångfacetterat och komplext. Att täcka alla potentiella risker i en enda studie är svårt, om inte omöjligt. Området kring AI och dess risker är dynamiskt och kommer kontinuerligt att utvecklas och förändras över tid. Det innebär att förståelsen av risker också behöver vara flexibel och uppdateras i takt med att nya teknologier och tillämpningar utvecklas. Det är viktigt att försöka hålla jämna steg med både tekniska framsteg och förändrade omvärldsförhållanden. På så sätt säkerställs det att nuvarande risker hanteras, samt att man förbereds på framtida utmaningar som kan uppstå i takt med att AI-teknologin fortsätter att avancera. I MSBs (2013) *Handlingsplan för skydd av samhällsviktig verksamhet* anges samarbete mellan privata och offentliga aktörer som ett led i att dessa verksamheter ska fungera optimalt. Detta etablerade nätverk och samarbetsförfarande kan bli ännu viktigare framöver när det gäller hanteringen av risker med AI. Dessa verksamheter kan således ta en ledande roll i hur risker hanteras inom andra sektorer, men detta medför också ett ansvar att redan från början, i AI:s tidiga skeden, agera korrekt. I dagsläget behöver man framförallt etablera en strategi för hur detta ska hanteras, och kanske även nå en konsensus om AI då det är ett område som idag är mångsidigt och saknar tydlig definition. Detta behov av definition och struktur kan potentiellt underlättas av den kommande EU AI Act, som syftar till att skapa en tydligare ram för AI-användning. Däremot medför även EU-lagstiftningen utmaningar och resultaten i denna studie visar att det finns en oro bland framförallt offentliga respondenter, att denna kommer hämma AI-implementering.

Även om denna studies syfte var att diskutera risker med AI var majoriteten positiva till denna teknik och respondenten från DIGG uttryckte att man inte bör bli för negativ och återhållsam. AI har potential att revolutionera många aspekter av samhället, från att förbättra effektiviteten i vården till att öka produktiviteten i näringslivet. Det är därför viktigt att försöka identifiera och hantera riskerna för att inte hämma framtida innovation.

7 Förslag framtida forskning

Under arbetets gång har det observerats att riskerna med AI varierar beroende på organisationernas typ av verksamhet, vilket indikerar att framtida studier med ett mer avgränsat och branschspecifikt urval skulle kunna identifiera fler och mer detaljerade risker. Förutom detta skulle ett framtida forskningsområde kunna omfatta en analys av åtgärder som implementeras för att hantera dessa risker, eftersom denna studie huvudsakligen fokuserar på att kartlägga riskerna. Ett ytterligare forskningsfält vore att utforska hanteringen av de risker som identifierats i denna studie i ljuset av den kommande EU AI Act, en aspekt som endast utgör en mindre del av studien men som förväntas bli en betydande framtida utmaning. Resultaten från studier om AI och dess associerade risker är även starkt beroende av tiden. Framtida forskning kan därför rikta in sig på longitudinella studier för att undersöka hur AI-risker och -strategier utvecklas över tid. Det vore även värdefullt att utforska skillnader i riskuppfattning och hantering mellan olika länder och kulturer för att utveckla mer globalt tillämpbara lösningar på risker relaterade till AI.

Referenser

Alhosani, K., & Alhashmi, S. M. (2024). Opportunities, challenges, and benefits of AI innovation in government services: a review. *Discover Artificial Intelligence*, *4*(1). https://doi.org/10.1007/s44163-024-00111-w

Bales, A., D'Alessandro, W., & Kirk-Giannini, C. D. (2024). Artificial Intelligence: Arguments for Catastrophic Risk. *Philosophy Compass*, e12964. https://doi.org/10.1111/phc3.12964

BBC. (2018, 10 oktober). *Amazon scrapped 'sexist AI' tool.* https://www.bbc.com/news/technology-45809919. [Hämtat: 2024-03-29].

Bell, J. (2006). *Introduktion till forskningsmetodik.* Studentlitteratur.

Berente, N., Bin Gu, Recker, J., & Santhanam, R. (2021). Managing Artificial Intelligence. *MIS Quarterly*, *45*(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Berggren, R., Kronblad, C., & Pregmark, J. E. (2023). Difficulties to digitalize: ambidexterity challenges in law firms. Journal of Service Theory and Practice, 33(2), 217–236. https://doi-org.ludwig.lub.lu.se/10.1108/JSTP-05-2022-0120

Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Stone, P., Saxenian, AL., Shah, J., Tambe, M., Teller. A. (2016). *Artificial Intelligence and Life in 2030.* One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University, Stanford, CA. Doc: http://ai100.stanford.edu/2016-report.

Brynjolfsson, E., McAfee, A., (2019). *Understanding AI and Machine learning.* Section 1 in Artificial Intelligence : The Insights You Need From Harvard Business Review. Harvard Business Review Press.

Buehler, K., Dooley, R., Grennan, L., & Singla, A. (2021). Getting to know—and manage—your biggest AI risks. *Mckinsey and Company*.

Bäck, J. (2023, 26 maj). *Vad är AI för något?* Internetstiftelsen. https://internetkunskap.se/artiklar/grundkurs-i-ai/vad-ar-ai-for-nagot/?gclid=EAIaIQobChMIoYGvs7L4ggMVBE-RBR0IFAEnEAAYASAAEgIm6PD_BwE [Hämtat: 2024-03-13]

Campbell, M., & Jovanović, M. (2024). Directing AI: Charting a Roadmap of AI Opportunities and Risks. *Computer*, *57*(2), 116-120.

Casey, B., A. Lemley, M. (2020). *You Might be a Robot.* 105 Cornell L. Rev. 287. https://scholarship.law.cornell.edu/clr/vol105/iss2/2

Center for AI Security [CAIS]. (u.å.). *Statement on AI Risk*. CAIS. https://www.safe.ai/work/statement-on-ai-risk [Hämtat: 2024-03-26]

Chan, B. (2023). Black-Box Assisted Medical Decisions: AI Power vs. Ethical Physician Care. Medicine, *Health Care and Philosophy, 26*(3), 285–292.

Cipu, D. D. (2023, 3 juli). *The current digital workplace: AI's positive and negative effects, ethical impacts, legal considerations, and future expectations*. https://www.linkedin.com/pulse/current-digital-workplace-ais-positive-negative-effects-daniel-d-cipu/ [Hämtat: 2024-02-22]

Coursera. (2024). *Machine Learning vs. AI: Differences, Uses, and Benefits.* https://www.coursera.org/articles/machine-learning-vs-ai?utm_medium=sem&utm_source=gg&utm_campaign=b2c_emea_ibm-data-science_ibm_ftcof_professional-certificates_arte_feb_24_dr_geo-multi_pmax_gads_lg-all&campaignid=21041942377&adgroupid=&device=c&keyword=&matchtype=&network=x&devicemodel=&adposition=&creativeid=&hide_mobile_promo&gad_source=1&gclid=CjwKCAiArfauBhApEiwAeoB7qN20KDofdxgHI1_5-Eap05iRS1JuChhqPxPFWeTLh9V5T2Gc5oip4RoCpbsQAvD_BwE [Hämtat: 2024-03-14]

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches (3rd ed.)*. Thousand Oaks and London: Sage Publications

Danielsen, M. (2023). The Emotional Risk Posed by AI (Artificial Intelligence) in the Workplace. *Norsk Filosofisk Tidsskrift*, *58*(2–3), 106–117. https://doi.org/10.18261/nft.58.2-3.4

De Cooman, J. (2022). Humpty dumpty and high-risk AI systems: the ratione materiae dimension of the proposal for an EU artificial intelligence act. *Mkt. & Competition L. Rev.*, *6*, 49.

Dekker, S. (2011). Drift into failure. from hunting broken components to understanding complex systems. Ashgate Pub.

Du, Y., & Yuan, C. (2022). A review of Artificial Intelligence risks in Social Science research. *Atlantis Highlights in Computer Sciences/Atlantis Highlights in Computer Sciences*, 273–293. https://doi.org/10.2991/978-94-6463-016-9_30

DIGG. (2024, 31 januari). *Introduktion till AI.* DIGG. https://www.digg.se/ai-for-socialtjansten/introduktion-till-ai#h-Vadarartificiellintelligens [Hämtat: 2024-03-21]

DIGG. (2023, 20 januari). *Uppdrag att främja offentlig förvaltnings förmåga att använda artificiell intelligens. I2021/01825*. DIGG. https://www.digg.se/download/18.5b30ce7218475cd9ed39384/1674479294670/Slutrapport%20Uppdrag%20att%20fr%C3%A4mja%20offentlig%20f%C3%B6rvaltnings%20f%C3%B6rm%C3%A5ga%20att%20anv%C3%A4nda%20AI%20I2021-01825.pdf [Hämtat: 2024-02-12]

EU Artificial Intelligence Act [EU AI Act]. (2024a, 19 maj). *Article 3: Definitions*. EU AI Act. https://artificialintelligenceact.eu/article/3/ [Hämtat: 2024-03-04]

EU Artificial Intelligence Act [EU AI Act] (2024b, 27 februari). *High-level summary of the AI Act.* EU AI Act. https://artificialintelligenceact.eu/high-level-summary/ [Hämtat: 2024-04-22]

Europaparlamentet (2023a, 27 juni). *Vad är artificiell intelligens och hur används det?* Europaparlamentet. https://www.europarl.europa.eu/topics/sv/article/20200827STO85804/vad-ar-artificiell-intelligens-och-hur-anvands-det [Hämtat: 2024-01-23]

Europaparlamentet (2023b, 27 juni) *Artificiell intelligens: Möjligheter och risker*. Europaparlamentet. https://www.europarl.europa.eu/topics/sv/article/20200918STO87404/artificiell-intelligens-mojligheter-och-risker [Hämtat: 2024-02-22]

Europaparlamentet (2023c, 19 december). *EU AI Act: first regulation on artificial intelligence.* Europaparlamentet. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence [Hämtat: 2024-04-22]

Europaparlamentet (2023d, 12 september). *Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI.* Europaparlamentet. https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai [Hämtat: 2024-05-09]

Europeiska rådet. (2024, 11 januari). *Timeline - Artificial intelligence*. Europeiska rådet. https://www.consilium.europa.eu/en/policies/artificial-intelligence/timeline-artificial-intelligence/?taxonomyId=271842c3-2535-4f5a-a049-bcdab2758865&taxonomyId=6b7901c5-1094-4713-add8-3364400eee98 [Hämtat: 2024-04-22]

Europeiska kommissionen. (2020, 19 januari). *White Paper On Artificial Intelligence - A European approach to excellence and trust.* Europeiska kommissionen. https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en [Hämtat: 2024-03-30].

Europeiska kommissionen. (2024a, 20 februari). *AI Act*. Europeiska kommissionen. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai [Hämtat: 2024-02-22]

Europeiska kommissionen. (2024b). *EU artificial intelligence Act. Final Draft (2024).* https://artificialintelligenceact.eu/the-act/ [Hämtat: 2024-02-22]

Eurostat. (2023, 8 april). *Use of artificial intelligence in enterprises.* https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use_of_artificial_intelligence_in_enterprises [Hämtat: 2024-02-12]

Ferdman, C., Nilsson, S., (2016) *Löneskillnader mellan offentlig och privat sektor.* SCB. https://www.scb.se/contentassets/b1ae4493ffd1404987a4d32cbf213ae5/loneskillnader-mellan-offentlig-och-privat-sektor.pdf

Goodson, M. (2021, 28 april). *Why is it so difficult to define artificial intelligence?* Evolution AI. https://www.evolution.ai/post/a-single-future-proof-definition-of-ai [Hämtat: 2024-03-13]

Graziano, A, Raulin, M. (1989). *Research methods. A process of Inquiry*. Printer and binder: R.R Donnelly & Sons Company, New York

Hammond, G. (2023, July 20). What are the risks of using AI in business? *Financial Times*. https://www.ft.com/content/4ded4c1d-5e99-42bf-9aa8-ea6e634aa060 [Hämtat: 2024-02-21]

He, G., Kuiper, L., & Gadiraju, U. (2023). *Knowing About Knowing: An Illusion of Human Competence Can Hinder Appropriate Reliance on AI Systems*. https://doi.org/10.1145/3544548.3581025

Holme, I.M. & Solvang, B.K. (1997). *Forskningsmetodik: om kvalitativa och kvantitativa metoder*. (2., [rev. och utök.] uppl.) Lund: Studentlitteratur.

Holzinger, A., Kieseberg, P., Weippl, E. and Tjoa, A.M., 2018. Current advances, trends and challenges of machine learning and knowledge extraction: from machine learning to explainable AI. In *Machine Learning and Knowledge Extraction: Second IFIP TC 5, TC 8/WG 8.4, 8.9, TC 12/WG 12.9 International Cross-Domain Conference, CD-MAKE 2018, Hamburg, Germany, August 27–30, 2018, Proceedings 2* (pp. 1-8). Springer International Publishing.

Höst, M., Regnell, B., & Runeson, P. (2006). *Att genomföra examensarbete* (6 uppl.). Studentlitteratur AB, Lund.

Internetstiftelsen. (2021). *Algoritmer*. Internetkunskap. https://internetkunskap.se/artiklar/ordlista/algoritm/ [Hämtat: 2024-02-12].

IBM. (u.å.). *What is artificial intelligence? Deep learning vs. machine learning.* IBM. https://www.ibm.com/topics/artificial-intelligence [Hämtat: 2024-03-22]

Jamshed S. (2014). Qualitative research method-interviewing and observation. *Journal of basic and clinical pharmacy*, *5*(4), 87–88. https://doi.org/10.4103/0976-0105.141942

Jarvenpaa, S. L., Selander, L., & Kronblad, C. (2023). Awakening to Algorithmic Transgressions: Non-User Discovery of Algorithmic Decision Making. Academy of Management Annual Meeting Proceedings, 2023(1), 1–6. https://doi-org.ludwig.lub.lu.se/10.5465/AMPROC.2023.17bp

Jawhar, S., Miller, J., & Bitar, Z. (2024, February). AI-Driven Customized Cyber Security Training and Awareness. In *2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC)* (pp. 1-5). IEEE.

Kreps, S., McCain, R. M., & Brundage, M. (2022). All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation. *Journal of Experimental Political Science*, *9*(1), 104–117. https://doi.org/10.1017/XPS.2020.37

Kaminski, M. E. (2024). Regulating the Risks of AI. *Forthcoming, Boston University Law Review*, *103*.

Keding, C. and Meissner, P., (2021). Managerial overreliance on AI-augmented decision-making processes: How the use of AI-based advisory systems shapes choice behavior in R&D investment decisions. *Technological Forecasting and Social Change*, *171*, p.120970.

Khogali, H. O., & Mekid, S. (2023). *The blended future of automation and AI: Examining some long-term societal and ethical impact features.* Technology in Society, 102232. https://doi-org.ludwig.lub.lu.se/10.1016/j.techsoc.2023.102232

Kleven, T. (1995) *Reliabilitet som pedagogisk problem. Rapport Nr 9 1995*. Universitet i Oslo, Pedagogisk forskingsinstitut, Oslo

Krafft, P. M., Young, M., Katell, M., Huang, K., & Bugingo, G. (2020, February). Defining AI in policy versus practice. I *Proceedings of the AAAI / ACM Conference on AI, Ethics, and Society* (s. 72-78).

Larsson, S., (2023). *Reglering av AI: För lite för sent eller för mycket för tidigt? En rapport om generativ AI och AI-förordningen.* Myndigheten för tillväxtpolitiska utvärderingar och analyser. Tillväxtanalys. Rapport2023:17 Larsson_Rapport_2023_17_Reglering_av_AI_-_F_r_lite_f_r_sent_eller_f_r_mycket_f_r_tidigt.pdf (lu.se) [Hämtat: 2024-05-05]

Li, J. (2018). Cyber security meets artificial intelligence: a survey. *Frontiers of Information Technology & Electronic Engineering/Frontiers of Information Technology & Electronic Engineering, 19*(12), 1462–1474. https://doi.org/10.1631/fitee.1800573

Loudiyi, M. (2021). *Vem har koll när algoritmerna styr?* Chefstidningen. https://chefstidningen.se/ledarskap/vem-har-koll-nar-algoritmerna-styr/ [Hämtat: 2024-01-31]

Lunds universitet. (2020, 15 september). *Allt du skulle vilja veta om AI.* Lunds universitet. https://www.lu.se/artikel/allt-du-skulle-vilja-veta-om-ai [Hämtat: 2024-03-15]

Marr, B. (2023, 30 juni). *Why companies are vastly underprepared for the risks posed by AI*. https://www.linkedin.com/pulse/why-companies-vastly-underprepared-risks-posed-ai-bernard-marr/ [Hämtat: 2024-02-27]

Malatji, M., & Tolah, A. (2024). Artificial intelligence (AI) cybersecurity dimensions: a comprehensive framework for understanding adversarial and offensive AI. *AI and Ethics*, 1–28. https://doi-org.ludwig.lub.lu.se/10.1007/s43681-024-00427-4

Mayer, A.-S., Strich, F. and Fiedler, M. (2020) 'Unintended Consequences of Introducing AI Systems for Decision Making', *MIS Quarterly Executive*, 19(4), pp. 239–257. doi:10.17705/2msqe.00036.

McGuire, B., Smith, C., Huang, T., Yang, G. (2006). *The History of Artificial Intelligence.* University of Washington. https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf [Hämtat: 2024-02-01]

Medeiros, M. (2020). Public and private dimensions of AI technology and security. *Modern conflict and artificial intelligence*, 20.

Mellouli, S., Janssen, M., & Ojo, A. (2024). *Introduction to the Issue on Artificial Intelligence in the Public Sector: Risks and Benefits of AI for Governments.* Digital Government: Research and Practice, 5(1), 1–6. https://doi.org/10.1145/3636550

MSB. (2013, december). *Handlingsplan för skydd av samhällsviktig verksamhet*. MSB. https://www.msb.se/contentassets/d8fca23b124c4686a629970fd2c1aa31/handlingsplan-for-skydd-av-samhallsviktig-verksamhet.pdf [Hämtat: 2024-02-20]

MSB. (2023, 1 december). *Identifiera samhällsviktig verksamhet.* MSB. https://www.msb.se/sv/amnesomraden/krisberedskap--civilt-forsvar/samhallsviktig-verksamhet/identifiera-samhallsviktig-verksamhet/#vad_ar_samhallsvktig_verksamhet [Hämtat: 2024-02-20]

MSB. (2020). *Introduktion till samhällsviktig verksamhet* [PowerPoint-presentation]. MSB. https://www.msb.se/contentassets/75e789d780c741cd9c8621eac846ec21/introduktion-till-samhallsviktig-verksamhet.pptx [Hämtat: 2024-02-20]

Neumann, O., Guirguis, K., & Steiner, R. (2024). Exploring artificial intelligence adoption in public organizations: a comparative case study. *Public Management Review, 26*(1), 114–141. https://doi.org/10.1080/14719037.2022.2048685

Nigmatov, A., & Pradeep, A. (2023, September). The Impact of AI on Business: Opportunities, Risks, and Challenges. In *2023 13th International Conference on Advanced Computer Information Technologies (ACIT)* (pp. 618-622). IEEE.

Nolan, D., Maryam, H., Kleinman, M., (2024). *The Urgent but Difficult Task of Regulating Artificial Intelligence.* Amnesty Tech. https://www.amnesty.org/en/latest/campaigns/2024/01/the-urgent-but-difficult-task-of-regulating-artificial-intelligence/ [Hämtat: 2024-03-13]

OECD.AI. (2024). https://oecd.ai/en/data [Hämtat: 2024-02-22]

O'Shaughnessy, M. (2022, 16 oktober). *One of the Biggest Problems in Regulating AI Is Agreeing on a Definition.* Carnegie Endowment for International Peace. https://carnegieendowment.org/2022/10/06/one-of-biggest-problems-in-regulating-ai-is-agreeing-on-definition-pub-88100 [Hämtat: 2024-03-13]

Pinnell, J. (2024). Ai, Disinformation and Democracy: The Need for Parliaments to Act. Parliamentarian, 105(1), 36–38.

Ponto J. (2015). Understanding and Evaluating Survey Research. *Journal of the advanced practitioner in oncology*, *6*(2), 168–171.

Prest, M. (2023, 4 juli). *Med AI blir eget omdöme ännu viktigare.* Åbo Akademi. https://www.abo.fi/nyheter/med-ai-blir-eget-omdome-annu-viktigare/ [Hämtat: 2024-02-16]

Raynovich, R. S. (2023, 22 juni). *The top five real risks of AI to your business.* Forbes. https://www.forbes.com/sites/rscottraynovich/2023/06/22/the-top-five-real-risks-of-ai-to-your-business. [Hämtat: 2024-02-22]

Realtid. (2017, 23 maj). *AI-expert: Rädslan för massarbetslöshet överdriven.* https://www.realtid.se/ai-expert-radslan-massarbetsloshet-overdriven/ [Hämtat: 2024-02-12]

Regeringskansliet. (2023, 8 december). *Regeringen tillsätter en AI-kommission för att stärka svensk konkurrenskraft*. Regeringskansliet. https://www.regeringen.se/pressmeddelanden/2023/12/regeringen-tillsatter-en-ai-kommission-for-att-starka-svensk-konkurrenskraft/ [Hämtat: 2024-04-01]

Rockwell, A. (2017, 28 augusti). *The History of Artificial Intelligence.* Science in the News. https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/ [Hämtat: 2024-02-06]

Sætra, H.S., Danaher, J. (2023). Resolving the battle of short-vs. long-term AI risks. *AI and Ethics*, pp.1-6.

SCB. (2019). *Artificiell intelligens i Sverige.* SCB. https://www.scb.se/contentassets/4d9059ef459e407ba1aa71683fcbd807/nv0116_2019a01_br_xftbr2001.pdf [Hämtat: 2024-02-12]

SCB. (2023). *AI-användning i företag och offentlig sektor.* SCB. https://www.scb.se/contentassets/ea0e9cccd58343e7a07fe4c055f8fad2/nv0116_2022a01_br_nvftbr2301.pdf [Hämtat: 2024-02-12]

Schoonenboom, J., & Johnson, R. B. (2017). How to Construct a Mixed Methods Research Design. *Kölner Zeitschrift fur Soziologie und Sozialpsychologie*, *69*(Suppl 2), 107–131. https://doi.org/10.1007/s11577-017-0454-1

Schneier, B. (2023). Trustworthy AI Means Public AI [Last Word]. IEEE SECURITY & PRIVACY, 21(6), 95–96. https://doi-org.ludwig.lub.lu.se/10.1109/MSEC.2023.3301262

Sisask, R. (2023, 22 december). *Stämde Göteborg för stor teknikmiss – förlorade*. Kvalitetsmagasinet. https://kvalitetsmagasinet.se/stamde-goteborg-for-stor-teknikmiss-forlorade/ [Hämtat: 2024-01-31].

Sasaki, R. (2023, October). AI and Security-What Changes with Generative AI. In *2023 IEEE 23rd International Conference on Software Quality, Reliability, and Security Companion (QRS-C)* (pp. 208-215). IEEE.

Srivastava, S. (2024, 12 februari). *Beneath the Code: Dissecting AI's Fundamental Risks and Their Countermeasures*. Appinventiv. https://appinventiv.com/blog/ai-risks/ [Hämtat: 2024-02-22]

Statskontoret. (2024). *Myndigheterna och AI - En studie om möjligheter och risker med att använda AI i statsförvaltningen.* Statskontoret. https://www.statskontoret.se/publicerat/publikationer/publikationer-2024/myndigheterna-och-ai---en-studie-om-mojligheter-och-risker-med-att-anvanda-ai-i-statsforvaltningen/?publication=true#_Toc162247475 [Hämtat: 2024-04-01]

Stockholms universitet. (2023, 30 oktober). *Export control and dual-use products.* Stockholms universitet. https://www.su.se/staff/services/emergency-crisis/export-control-and-dual-use-products [Hämtat: 2024-04-04]

Tartaro, A., Panai, E. & Cocchiaro, M.Z. AI risk assessment using ethical dimensions. *AI Ethics* (2024). https://doi.org/10.1007/s43681-023-00401-6

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind, 59*(236), 433–460. http://www.jstor.org/stable/2251299

Wagner, T. L., & Blewer, A. (2019). "The word real is no longer real": Deepfakes, gender, and the challenges of ai-altered video. *Open Information Science*, *3*(1), 32-46.

Wirtz, B. W., Weyerer, J. C., & Kehl, I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. *Government Information Quarterly*, *39*(4), 101685.

World Economic Forum. (2023, 26 juli). *Chief Risk Officers Outlook: July 2023* World Economic Forum. https://www3.weforum.org/docs/WEF_Chief_Risk_Officers_Outlook_2023.pdf [Hämtat: 2024-02-22]

Wörsdörfer, M. (2023). Mitigating the adverse effects of AI with the European Union's artificial intelligence act: Hype or hope? *Global Business and Organizational Excellence*. https://doi.org/10.1002/joe.22238

Zhou, J., Pei, K., Qiu, X., Huang, M., & Zhang, J. (2023). ChatGPT: potential, prospects, and limitations. *Frontiers of Information Technology & Electronic Engineering/Frontiers of Information Technology & Electronic Engineering*, *25*(1), 6–11. https://doi.org/10.1631/fitee.2300089

Bilagor

Bilaga A: Samhällsviktiga sektorer från MSB (2023)

Samhällssektor	Exempel på viktiga samhällsfunktioner
Energiförsörjning	Produktion och distribution av el, fjärrvärme och bränslen och drivmedel
Finansiella tjänster	Betalningar, tillgång till kontanter, centrala betalningssystemet samt värdepappershandel
Handel och industri	Bygg- och entreprenadverksamhet, detaljhandel samt tillverkningsindustri
Hälso- och sjukvård samt omsorg	Akutsjukvård, läkemedels- och materielförsörjning, omsorg om barn, funktionshindrade och äldre, primärvård, psykiatri, socialtjänst samt smittskydd för djur- och människor
Information och kommunikation	Telefoni (mobil och fast), internet, radiokommunikation, distribution av post, produktion och distribution av dagstidningar, webbaserad information, sociala medier
Kommunalteknisk försörjning	Dricksvattenförsörjning, avloppshantering, renhållning samt väghållning
Livsmedel	Distribution-, primärproduktion-, kontroll- och tillverkning av livsmedel
Offentlig förvaltning	Lokal, regional och nationell ledning, begravningsverksamhet samt diplomatisk och konsulär verksamhet
Skydd och säkerhet	Domstolsväsendet, åklagarverksamhet, militärt försvar, kriminalvård, kustbevakning, polis, räddningstjänst, alarmeringstjänst, tullkontroll, gränsskydd och immigrationskontroll samt bevaknings- och säkerhetsverksamhet
Socialförsäkringar	Allmänna pensionssystemet samt sjuk- och arbetslöshetsförsäkringen
Transporter	Flyg-, järnvägs-, sjö- och vägtransport samt kollektivtrafik

Bilaga B: Riskkategorisering

Riskkategori	Källa
Teknologiska, data- och analytiska Teknologiska, data- och analytiska risker fokuserar på de risker som skapas av AI-systemet på grund av dess utformning och design. Dessa risker omfattar utmaningar knutna till hur systemen är byggda, deras komplexitet, och potentiella svagheter som kan leda till oväntade problem eller säkerhetsbrister.	(Hammond, 2023) (Marr, 2023) (Raynovich, 2023) (Cipu, 2023) (Europaparlamentet, 2023b) (Europeiska kommissionen, 2024) (Campbell & Jovanovic, 2024) (De Cooman, 2022). (Nigmatov & Pradeep, 2023) (World Economic Forum, 2023) (Buehler et al., 2021) (Kaminski, 2023) (Du & Yuan, 2022) (Wirtz et al., 2022).
Informativa och kommunikativa Informativa och kommunikativa risker handlar om farorna med manipulerad information genom AI. Dessa risker uppstår när AI-systemens algoritmer, som är tränade av utomstående att leverera vinklad information, kan utgöra en risk för organisationen. Om sådan felaktig information används i beslutsfattande, som grund för bedömningar eller liknande processer, kan det få negativa konsekvenser.	(Hammond, 2023) (Marr, 2023) (Raynovich, 2023) (Cipu, 2023) (Srivastava, 2024) (Europaparlamentet, 2023b) (Europeiska kommissionen, 2024) (Campbell & Jovanovic, 2024) (De Cooman, 2022) (World Economic Forum, 2023) (Wirtz et al., 2022).
Ekonomiska Ekonomiska risker berör risker med ekonomisk förlust till följd av beroendet/förlitan till AI-system.	(Raynovich, 2023) (Srivastava, 2024) (Europaparlamentet, 2023b) (Wirtz et al., 2022)
Sociala Sociala risker handlar om AI-systemets påverkan på den anställda individens välbefinnande.	(Hammond, 2023) (Marr, 2023) (Cipu, 2023) (Srivastava, 2024) (Europaparlamentet, 2023b) (Europeiska kommissionen, 2024) (De Cooman, 2022). (Nigmatov & Pradeep, 2023) (Du & Yuan, 2022) (Wirtz et al., 2022).

Etiska Etiska risker inom AI rör diskriminering, segregation, och utanförskap, drivna av algoritmisk-och databias som kan leda till orättvis behandling av individer eller grupper.	(Marr, 2023) (Raynovich, 2023) (Cipu, 2023) (Srivastava, 2024) (Europaparlamentet, 2023b) (Europeiska kommissionen, 2020) (Europeiska kommissionen, 2024) (Campbell & Jovanovic, 2024) (De Cooman, 2022). (Nigmatov & Pradeep, 2023) (World Economic Forum, 2023) (Buehler et al., 2021) (Kaminski, 2023) (Du & Yuan, 2022) (Wirtz et al., 2022). (Statskontoret, 2024) (Mellouli, et al., 2024)
Juridiska och regelmässiga Juridiska och regelmässiga risker inom AI omfattar ansvarsfördelning, tillämpning av sanktioner och styrning, där centrala utmaningar inkluderar hanteringen av ansvar vid fel orsakade av AI och etableringen av en effektiv AI-styrning genom ramverk, policyer, lagar och standarder.	(Hammond, 2023) (Raynovich, 2023) (Cipu, 2023) (Srivastava, 2024) (Europaparlamentet, 2023b) (Europeiska kommissionen, 2024) (Campbell & Jovanovic, 2024) (De Cooman, 2022). (Wirtz et al., 2022).

Bilaga C: Enkätfrågor

Riskkategori	Frågor
Teknologiska, data- och analytiska AI-risker	1. I vilken utsträckning bedömer du att användningen av AI-teknik inom din organisation ökar er organisations exponering för cyberangrepp eller cybersäkerhetsincidenter? 2. I vilken utsträckning bedömer du att er verksamhet exponeras för risker på grund av att anställda överskattar, underskattar eller på annat vis missbedömer AI-verktyg och därför använder dem på ett sätt som kan leda till incidenter? 3. I vilken utsträckning bedömer du att er organisation exponeras för risk relaterat till komplexiteten i AI-systemet och följaktligen svårigheter att felsöka AI-systemet om ett fel inträffar? 4. I vilken utsträckning bedömer du att användningen av AI-teknik inom din organisation ökar risken för spridning av personlig eller känslig data? *Är det någon annan risk inom teknologiska, data- och analytiska som du tycker saknas i detta formulär?*
Informativa och kommunikativa AI-risker	5. I vilken utsträckning bedömer du att er organisation exponeras för risker kopplade till vinklad information som manipulerats genom ett AI-verktyg, exempelvis manipulerade bilder, ljud, texter eller videor (mm)? 6. I vilken utsträckning bedömer du att er verksamhet exponeras för en risk att påverkas av desinformation eller riskerar att använda falska kunskapsunderlag? *Är det någon annan risk inom informativa och kommunikativa som du tycker saknas i detta formulär?*
Ekonomiska AI-risker	7. I vilken utsträckning bedömer du risken för ekonomisk förlust till följd av att AI-systemet slutar fungera? *Är det någon annan risk inom ekonomiska som du tycker saknas i detta formulär?*
Sociala AI-risker	8. I vilken utsträckning bedömer du risken för negativa konsekvenser på personals välbefinnande (i form av ökad stress och oro att bli ersatt) på grund av organisationens AI-system? 9. I vilken utsträckning bedömer du att er organisations anställda riskerar att drabbas av känslor av isolering och ensamhet till följd av att de interagerar mer med AI-system

	än med kollegor? *Är det någon annan risk inom <u>sociala</u> som du tycker saknas i detta formulär?*
Etiska AI-risker	10. I vilken utsträckning bedömer du att organisationens AI-system kan riskera orsaka diskriminering eller orättvis behandling av individer eller grupper? *Är det någon annan risk inom <u>etiska</u> som du tycker saknas i detta formulär?*
Juridiska och regelmässiga AI-risker	11. I vilken utsträckning bedömer du att er verksamhet exponeras för risker kopplade till svårigheter att möta rådande lagstiftning? *Är det någon annan risk inom <u>juridiska och regelmässiga</u> som du tycker saknas i detta formulär?*

Bilaga D: Intervjuguide

<table>
<tr><td colspan="3">Del 1. Organisationsspecifika frågor
I detta avsnitt önskar vi veta om, och i så fall hur, ni använder AI i er organisation samt vilka risker ni ser med detta, alternativt vad för risker ni ser med potentiell användning av AI i er organisation.</td></tr>
<tr><td>Område</td><td colspan="2">Frågor</td></tr>
<tr><td>Definition av AI</td><td colspan="2">Enligt EU Artificial Intelligence Act definieras AI-system som “ett maskinbaserat system som är utformat för att fungera med varierande grad av autonomi och som kan uppvisa anpassningsförmåga efter driftsättning och som, för uttryckliga eller underförstådda mål, från den indata det tar emot drar slutsatser om hur man ska generera utdata såsom prognoser, innehåll, rekommendationer eller beslut som kan påverka fysiska eller virtuella miljöer”.

För den här intervjun, är detta en definition du kan ställa dig bakom?</td></tr>
<tr><td rowspan="2">Användning av AI inom organisationen</td><td colspan="2">Använder ni AI inom er organisation eller planerar ni att göra detta?</td></tr>
<tr><td>Om ja,
Vilken funktion använder ni AI:n till?

Varför tyckte ni att AI var lämpligt just i det användningsområde ni valt?</td><td>Om nej,
Hur kommer det sig?</td></tr>
<tr><td>Riskkategorisering

Innan intervjun skickade vi ut ett formulär där vi bad er ranka risker inom er organisation. Här kommer vi gå in på era svar i enkäten vi skickade innan intervjun och be er utveckla det ni svarat. Här öppnar vi upp för diskussion.</td><td colspan="2">Kan ni utveckla era valda placeringar?

Här är vi särskilt intresserade av att höra om det finns några risker ni tycker saknas bland dem som vi lyft i formuläret?</td></tr>
<tr><td colspan="3">Del 2. Framtid och risker med AI
I det här avsnittet uppmanar vi er att reflektera över framtida risker med AI med</td></tr>
</table>

<table>
<tr><td colspan="3">utgångspunkt från de riskkategorier vi presenterat tidigare.</td></tr>
<tr><td>Framtida risker</td><td colspan="2">Av de 6 kategorierna som nämnts tidigare, hur tror du dessa kommer uttrycka sig/påverka framtiden?

Är det någon/några risk(er) ni tror kommer få större betydelse i framtiden?</td></tr>
<tr><td>Två största riskerna</td><td colspan="2">Enligt din bedömning, vilka är de två största riskkategorierna i organisationen idag?

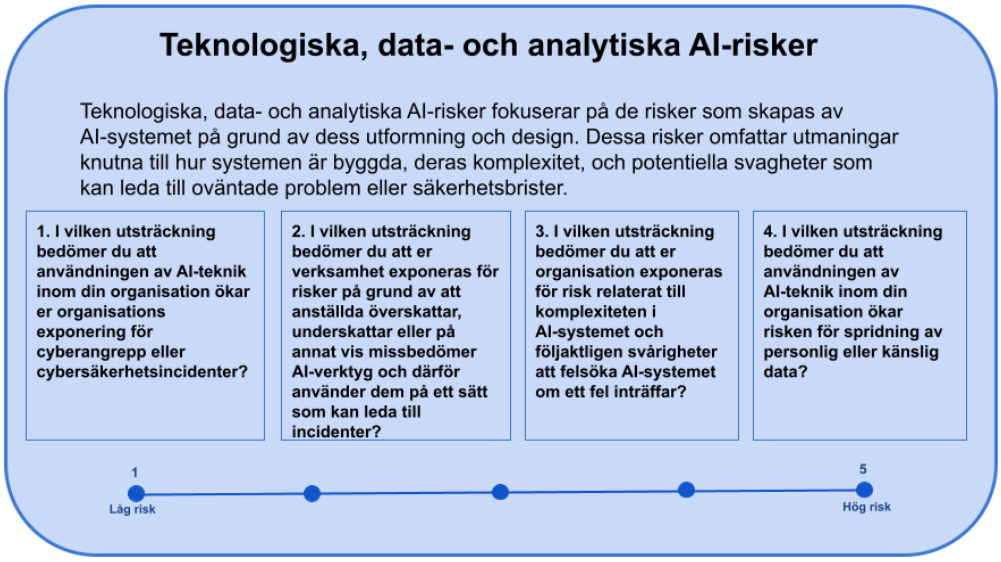
Enligt din bedömning, vilka är de två största riskkategorierna i organisationen i framtiden?</td></tr>
<tr><td rowspan="2">EU AI Act (kommer i slutet av 2025/början av 2026)</td><td colspan="2">Känner ni till detta?</td></tr>
<tr><td>Om ja,

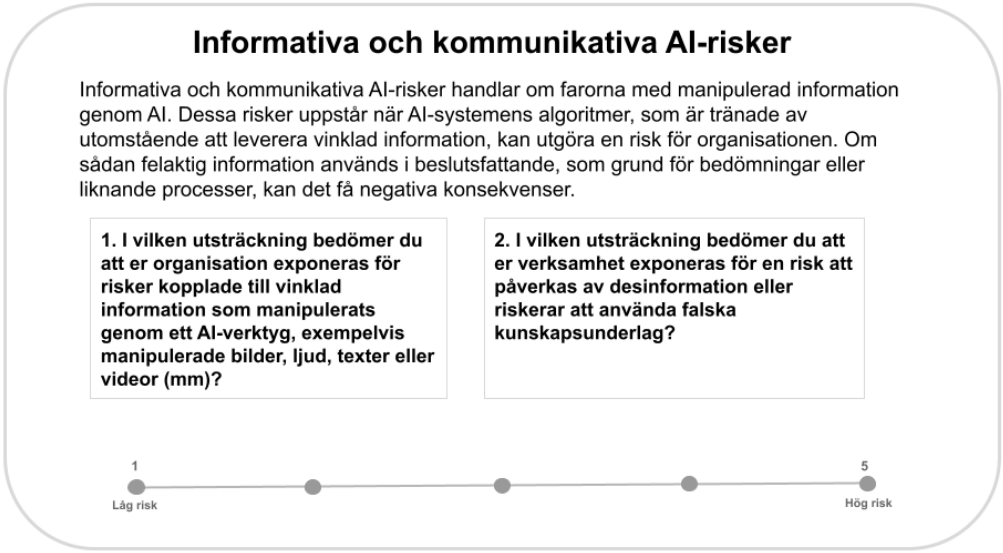
Vilken risknivå tror ni att ni hamnar på?

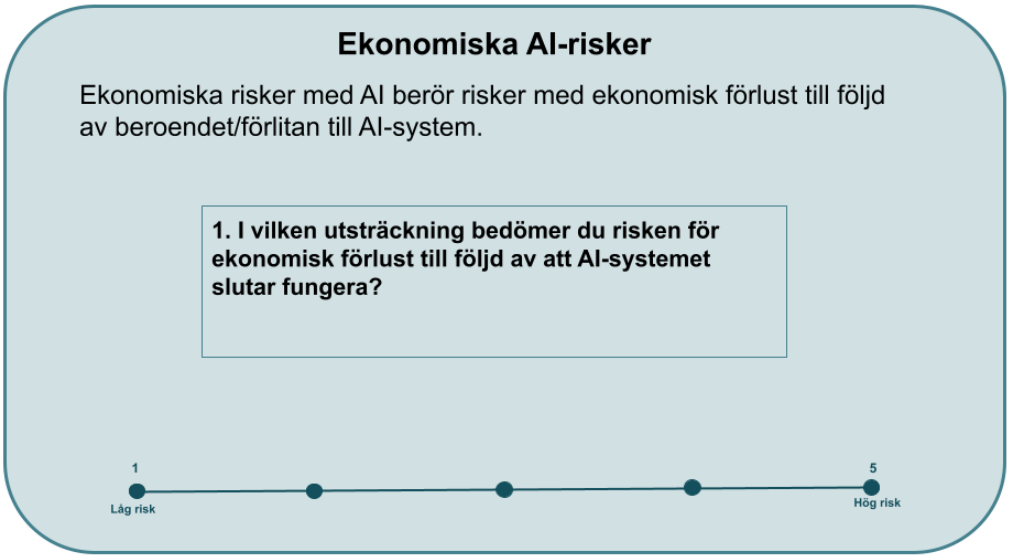
Ser ni några utmaningar med detta?</td><td>Om nej,

-</td></tr>
<tr><td>Skulle du säga att du är generellt positiv eller negativt inställd mot AI?</td><td colspan="2"></td></tr>
</table>

Bilaga E: Visuellt verktyg inför intervjuer







Sociala AI-risker

Sociala risker med AI handlar om AI-systemets påverkan på den anställda individens välbefinnande.

1. I vilken utsträckning bedömer du risken för negativa konsekvenser på personals välbefinnande (i form av ökad stress och oro att bli ersatt) på grund av organisationens AI-system?

2. I vilken utsträckning bedömer du att er organisations anställda riskerar att drabbas av känslor av isolering och ensamhet till följd av att de interagerar mer med AI-system än med kollegor?

Etiska AI-risker

Etiska risker inom AI rör diskriminering, segregation, och utanförskap, drivna av algoritmisk-och databias som kan leda till orättvis behandling av individer eller grupper.

1. I vilken utsträckning bedömer du att organisationens AI-system kan riskera orsaka diskriminering eller orättvis behandling av individer eller grupper?

Juridiska och regelmässiga AI-risker

Juridiska och regelmässiga risker inom AI omfattar ansvarsfördelning, tillämpning av sanktioner och styrning, där centrala utmaningar inkluderar hanteringen av ansvar vid fel orsakade av AI och etableringen av en effektiv AI-styrning genom ramverk, policyer, lagar och standarder.

1. I vilken utsträckning bedömer du att er verksamhet exponeras för risker kopplade till svårigheter att möta rådande lagstiftning?