AI inom cybersäkerhet, ett dubbelsidigt svärd

En kvalitativ studie om Artificiell Intelligens roll inom cybersäkerhet

Författare: Arvid Marklund, Felix Lundqvist & Simon Sjöman

HT 24
Informatik med systemvetenskaplig inriktning, Kandidatuppsats 15hp
Artificiell Intelligens (AI) & cybersäkerhet
Handelshögskolan vid Örebro universitet
Handledare: Kai Wistrand
Examinator: Ella Kolkowska

Sammanfattning

Cybersäkerhet är ett område som utvecklas kontinuerligt där den pågående digitaliseringen innebär både möjligheter och utmaningar för svenska företag. Traditionella metoder har visat sig otillräckliga mot de avancerade och sofistikerade cyberattacker som företag dagligen utsätts för. Detta har skapat ett behov av innovativa lösningar, där artificiell intelligens (AI) framstår som en av de mest lovande teknikerna. Denna studie syftar till att undersöka hur stora svenska företag implementerar och använder AI för att stärka cybersäkerheten, med fokus på samspelet mellan tekniska, mänskliga och miljömässiga faktorer. Genom kvalitativa intervjuer med företagsrepresentanter, kompletterat av en litteratursökning, analyseras dessa aspekter utifrån ett sociotekniskt ramverk. I resultatet presenteras det exempel på hur AI kan förbättra cybersäkerhetsarbetet, exempelvis genom automatiserad hotdetektion och effektiv hantering av stora datamängder. Samtidigt identifierades utmaningar, såsom bristande förtroende och risker med överberoende av teknologin. Studien understryker vikten av en balanserad integration mellan människa, teknik och miljö för att undvika risker och skapa lösningar som driver cybersäkerhetsarbetet framåt. Slutsatsen är att ett sociotekniskt perspektiv inte bara underlättar implementeringen av AI, utan även optimerar säkerheten och ökar organisationernas resiliens.

Nyckelord: Artificiell intelligens (AI), Cybersäkerhet, Sociotekniska system (STS), Förklarbar AI (XAI), Human-in-the-Loop.

Abstract

Cybersecurity is a continuously evolving field where ongoing digitalization brings both opportunities and challenges for Swedish companies. Traditional methods have proven insufficient against the advanced and sophisticated cyber threats the world faces today. This has created a need for innovative solutions, with artificial intelligence (AI) emerging as one of the most promising. This study aims to explore how large Swedish companies implement and utilize AI to strengthen cybersecurity, focusing on the interplay between technical, human, and environmental factors. Through qualitative interviews with company representatives, complemented by a literature review, these aspects are analyzed within a sociotechnical framework. The results show that AI has the potential to enhance cybersecurity efforts through automated threat detection and efficient management of large datasets. However, challenges such as a lack of trust and the risks of overdependence on technology were also identified. The study emphasizes the importance of a balanced integration of human, technological, and environmental elements to mitigate risks and create solutions that drive cybersecurity efforts forward. The conclusion is that a sociotechnical perspective not only facilitates the implementation of AI but also optimizes security and increases organizational resilience.

Keywords: Artificial Intelligence (AI), Cybersecurity, Sociotechnical Systems (STS), Explainable AI (XAI), Human-in-the-Loop.

Innehållsförteckning

1. Introduktion 2
1.1 Bakgrund 2
1.2 Problemformulering 4
1.3 Syfte och forskningsfråga 5
1.3.1 Frågeställningar 5
1.4 Avgränsning 5
2. Teori 6
2.1 Bakgrund av Socio-Technical Systems Theory 6
2.1.1 Anpassning av teorin 7
2.2 Tidigare forskning 8
2.2.1 Tidigare forskning om Sociala teman 8
2.2.2 Tidigare forskning om Tekniska teman 9
2.2.3 Tidigare forskning om Miljömässiga teman 10
3. Metod 12
3.1 Forskningsstrategi- och design 12
3.1.1 Forskningsstrategi 12
3.2 Datainsamling 13
3.2.1 Insamling av teoretiskt material 13
3.2.2 Utformning av teman och intervjufrågor 14
3.2.3 Urval 14
3.2.4 Insamling av empiriskt material 14
3.2.5 Etiska överväganden 15
3.3 Dataanalys 16
3.3.1 Analysmetod 16
3.3.2 Kvalitetskriterier 16
3.3.3 Metodologiska begränsningar 16
4. Resultat 18
4.1 Hur använder företagen AI? 18
4.2 Vilka effekter har implementationen medfört? 18
4.2.1 Sociala aspekter 18
4.2.2 Tekniska aspekter 21
4.2.3 Miljömässiga aspekter 24
5. Diskussion 27
6. Slutsats 29
6.1 Förslag på framtida forskning 29
7. Källförteckning 31
8. Bilagor 33
Bilaga 1 - Intervjuguide 33

Begreppslista

Black Box AI (BB)
BB refererar till AI-modeller där insyn i beslutsfattandet saknas, vilket gör det svårt för användaren att kontrollera anledningen till besluten som AI-modellen tar (Capuano et al., 2022).

Explainable AI (XAI)
XAI är motsatsen mot BB och avser AI-modeller som kan förklara sina beslut och resonemang för användaren (Capuano et al., 2022).

Human-in-the-Loop (HL).
HL är ett koncept som innebär att en användare inkluderas i en process som delvis har automatiserats med hjälp av AI. Människan har vanligtvis i en övervakande eller beslutsfattande roll för att säkerställa kvalitet och precision (Michael et al., 2024).

Cybersäkerhet
Avser att skydda elektroniska system, nätverk och digital information från cyberattacker, intrång och annan skadlig aktivitet (NIST, 2025).

1. Introduktion

I takt med att världen blir alltmer digitaliserad står företag inför både möjligheter och hot inom cybersäkerhet (Rojko, 2017). Rapporter visar att cyberattacker fortsätter att öka både i omfattning och komplexitet. Cyberattacker kostade företag runt 5,5 biljoner euro på den globala marknaden under 2020 och den siffran förväntas öka inom kommande år (Europaparlamentet). Företag tvingas därför att söka nya och innovativa lösningar för att skydda sina tillgångar och säkerställa trygghet i en alltmer hotfull digital värld (ENSISA, 2024).

Artificiell intelligens (AI) har under senare år framstått som en av de mest lovande teknologierna för att möta dessa utmaningar. Med sin förmåga att analysera stora mängder data, identifiera mönster och upptäcka hot snabbare än traditionella metoder har AI potentialen att revolutionera cybersäkerhetsarbetet (Zeadally et al., 2020). Samtidigt medför AI utmaningar, då tekniken även kan användas för att skapa nya typer av hot som företag står inför idag. Dessutom diskuteras det hur företag bäst kan implementera och utnyttja AI-tekniken för att förbättra säkerheten (ibid). Tidigare forskning har vanligtvis fokuserat på de tekniska aspekterna av AI men har ofta uteslutit de sociala och miljömässiga dimensionerna, vilka kan vara avgörande för långsiktig framgång (Jada & Mayayise, 2024). Detta har lett till en kunskapslucka som denna studie avser bidra till.

1.1 Bakgrund

Deep Learning (DL) revolutionerade möjligheterna med AI år 2010 och lade grunden för de komplexa modeller och neurala nätverk som används idag. DL används bland annat för att skapa generativa språkmodeller, Large Language Model (LLM) som är tränade på stora textmängder och kan förstå samt generera mänskligt språk på en avancerad nivå, exempelvis Chat-GPT. Denna teknik används inom flera olika områden, exempelvis hälsovård, röstassistenter, reklam men även för cybersäkerhet. Inom cybersäkerhet så används tekniken för att analysera stora datamängder och identifiera potentiella hot samt attacker, vilket den kan göra snabbare än traditionella metoder (Rudolph, 2023). Men tekniken används även för att automatisera attacker och möjliggör nya former av avancerade hot, vilket leder till att AI kan ses som ett dubbelsidigt svärd inom cybersäkerhet (Zhang et al., 2022).

Zeadally et al. (2020) belyser vikten av nyanserade cybersäkerhetslösningar för att bekämpa samt minska riskerna för en cyberattack eftersom de traditionella metoderna blir otillräckliga mot de ständigt växande och mer sofistikerade hoten. Ett av de hoten som idag nämns som en av de vanligaste är phishing. Phishing används inte sällan med epost och syftet är att övertyga en person att klicka på en viss länk som sedermera ska läcka konfidentiell eller privat information. Attacken sker ofta utifrån ett personaliserat mail i hopp om att övertyga en godtrogen användare att länken är legitim. För att möta utmaningarna med de avancerade hoten har AI blivit en förespråkad teknologi som ska leda vägen. AI och maskininlärning har visat sig väldigt effektiva på att analysera stora datamängder och identifiera avvikelser och dolda mönster som högst troligen är hot, så kallat anomalier (ibid). Författarna diskuterar vidare hur AI kan integreras för att försvåra för de cyberkriminellas nya metoder och bidra med cybersäkerhetslösningar. AI bidrar med betydande fördelar som Intrusion Detection Systems, vilket är en typ av maskininlärning som används till att upptäcka och klassificera nätverkstrafik som antingen legitim eller skadlig. Maskininlärningen kan ske i två olika aspekter, oövervakad och övervakad. Den övervakade inlärningen används för att träna AI-modeller med fördefinierade etiketter. Den oövervakade inlärningen utnyttjar AI:s förmåga att upptäcka dolda mönster och okända hot. Med hjälp av AI:s effektiva förmåga med

automatisering av hotdetektion kan säkerhetsteam agera snabbt, vilket gör AI till en ytterst viktig resurs i kampen mot de cyberkriminella (ibid). Enligt Gupta et al. (2023) kan ChatGPT utnyttjas för att förbättra cybersäkerhetslösningar. ChatGPT kan användas för att automatiskt analysera säkerhetsincidenter eftersom den snabbt kan granska enorma mängder loggdata och nätverkstrafik. Den har då möjlighet att hjälpa säkerhetsteam med hotdetektering i okända mönster eller potentiella attacker, som påskyndar upptäckten av hot och dess åtgärder. ChatGPT kan även skapa automatiska cybersäkerhetincidentssvar, såsom åtgärdsplaner och vägledning för säkerhetsteam genom krishantering (ibid).

Samtidigt finns det utmaningar när man har implementerat AI i cybersäkerthetssammanhang. Zhang et al. (2022) berättar om svårigheter som dataförfalskning, vilket innebär att små förändringar i data kan orsaka fel i de systemen använder AI. Genom att bara ändra några bytes i en fil kan AI uppfatta den som skadlig, vilket gör att säkerheten äventyras. Författarna berättar även om AI:s dubbelsidiga natur, att den är lika effektiv för angripare som för personer med goda intentioner. De menar att de innovativa lösningarna för cybersäkerhet som föds med AI kan även få bakslag där nya lösningar för cyberkriminella kan uppdagas, detta har lett till att det idag är en kapplöpning där både försvar och attacker utvecklas parallellt (ibid). Fortsättningsvis förklarar Gupta et al. (2023) hur ChatGPT kan missbrukas av personer med oetiska intentioner. Social engineering handlar om att utnyttja en svag länk i cybersäkerhetsarbetet, människan. Det kan handla om e-mail med länkar eller frågor om känslig information och med tanke på hur effektivt ChatGPT kan generera texter är det svårt att avgöra ifall de är skrivna av en generativ AI-modell eller en människa (ibid). En specifik metod inom social engineering är phishing där en LLM som ChatGPT ofta utnyttjas till att skicka personaliserade e-mails till anställda. Poängen med metoden är att få förtroende från den som blir utsatt så att hen sedan delger känslig information eller klickar på en länk med virus.

Regeringskansliet informerade i september 2024 att Sverige satsar på att bli en ledande nation inom tillförlitlig och säker AI. För att underlätta en fortsatt framgång inleddes en utredning som undersöker hur Sverige effektivt kan implementera EU:s AI-förordning. Dessa riktlinjer har nu fastställts inom EU och de syftar till att skapa gemensamma riktlinjer som förespråkar en människocentrerad och tillförlitlig AI i syfte att öka innovation och tillväxt (Regeringen, 2024). Sverige har bra förutsättningar för denna satsning enligt SCB (2024) som visar att 43% av stora företag i Sverige har implementerat AI idag samt att det är en ökande trend.

En ytterligare satsning som sker i Sverige är inrättandet av Nationella Cybersäkerhetscentret (NCSC), som har fått i uppdrag av regeringen att rusta upp Sveriges förmågor gällande cybersäkerhet. Detta är ett omfattande arbete som innefattar Försvarsmakten, MSB, Säkerhetspolisen, FRA och Polisen. NCSC beskriver AI:s roll inom cybersäkerhet som dubbelverkande, där AI-tekniken möjliggör nya och effektiva försvarssystem men samtidigt kan det användas för nya och skadliga angrepp av illasinnade aktörer (NSCS, 2024).

Implementeringen av AI inom cybersäkerhet kräver ett holistiskt tillvägagångssätt där både tekniska, mänskliga och miljömässiga aspekter tas i beaktning. Enligt Michael et al. (2024) tenderar cybersäkerhet att prioritera tekniska lösningar, vilket ofta innebär att de mänskliga och miljömässiga aspekterna förbises. Författarna använder ett socio-tekniskt ramverk för att bättre kunna förstå samspelet mellan de olika dimensionerna och lyfter vikten av att involvera flera intressenter i implementationen av AI inom cybersäkerhet.

Vidare utvecklar Schoenherr et al. (2023) resonemanget om behov utöver de tekniska genom att betona behovet av en människocentrerad ansats inom AI-utveckling, där människan står i centrum för design- och utvecklingsprocessen. En sådan ansats syftar till att säkerställa att AI-lösningar inte bara är tekniskt effektiva, utan även transparenta, begripliga och förankrade i mänskliga värderingar. Författarna menar att detta ska öka förtroendet hos användare då de inte kommer känna sig lika osäkra kring hur beslut fattas samt på vilka grunder.

Jada & Mayayise (2023) påpekar i deras studie hur tekniska lösningar ensamma ofta är otillräckliga för att hantera komplexa hot som organisationer står inför. De betonar att inkluderandet av mänskliga faktorer är avgörande för att säkerställa resiliens och långsiktig effektivitet. Samtidigt lyfter de riskerna med att förlita sig för mycket på AI, då de kan utsättas för sårbarheter såsom manipulerade data eller brist på förklarbarhet, vilket kan försvaga säkerheten snarare än att stärka den.

Idag har många företag och organisationer börjat med initiativ för att implementera AI-lösningar till sin cybersäkerhet. Implementeringen har resulterat i många tekniska fördelar som nämnt tidigare, men det har även uppstått en hel del organisatoriska och sociala svårigheter som måste bemötas. Jada & Mayayise (2024) menar att det finns flera problem som kan uppstå ifall det sker ett försök med integrering av AI-lösningar. Ett av dem är att det krävs stora resurser, både ekonomiska och mänskliga för att implementationen ska gå effektivt eftersom det ska integreras in i befintliga system. Det kan även uppstå svårigheter för anställda eftersom AI-lösningar kan förändra arbetsroller och processer på företaget. Vilket i sin tur kan skapa motstånd från medarbetare som inte har förtroende eller kunskap om systemen. En annan sida av AI-implementation är risken för överberoende av AI, vilket innebär att de anställda fullständigt litar på systemen utan att granska det som genererats som kan medföra en säkerhetsrisk (ibid).

Som tidigare visat finns det ett stort intresse av att genomföra en förändring och det sker redan, men det fokuseras för mycket på den tekniska biten av implementationen. För att förstå hur organisationer kan utnyttja implementeringen av AI-lösningar för förbättrad cybersäkerhet förhåller sig denna studie till att analysera data utifrån den sociotekniska teorin. Den sociotekniska teorin ska se till att alla viktiga faktorer tas i beaktande vid implementeringen. Mumford (2000) förklarar att teorin tar hänsyn till både mänskliga och tekniska delar när det implementeras ett nytt system. Människorna som arbetar med systemet och själva systemet i helhet ska utvecklas parallellt för att bidra till en effektiv och god arbetsmiljö. Det är viktigt att de mänskliga behoven inte ignoreras när ett tekniskt system introduceras. Teorin ska även fungera som en ram för att optimera samspelet mellan människan, tekniken och miljön.

1.2 Problemformulering

Utvecklingen av sofistikerade cyberattacker har drivit på framfarten för cybersäkerhetsarbetet då de traditionella cybersäkerhetslösningarna idag i många fall är otillräckliga. Detta har medfört att det idag finns ett behov av nya perspektiv för hur arbetet med cybersäkerheten ska ske och där har konceptet om AI-lösningar vuxit fram (Zeadally et al., 2020). AI erbjuder tekniskt effektiva lösningar genom automatiserad hotidentifiering, snabbare respons och hantering av stora mängder data (ibid). Dock fokuseras majoriteten av forskningen på de tekniska effekterna medan viktiga organisatoriska faktorer tenderar att bli marginaliserade. Jada & Mayayise (2024) menar att de sociala faktorerna är viktiga och spelar en avgörande roll för organisationers möjligheter att implementera AI i cybersäkerhetsarbetet effektivt. De sociala faktorerna handlar om de anställda och innefattar exempelvis förtroende samt

arbetsbörda. De menar även att miljömässiga och externa faktorer som hot eller regelverk bör tas i beaktning för att säkerställa att implementationen senare efterlevs och effektiviserar arbetet med cybersäkerhet för organisationen (ibid). De ovanstående punkterna väcker frågan om hur väl organisationer tar hänsyn till samspelet mellan människan, tekniken och miljön när de väljer att implementera AI-baserade lösningar i cybersäkerhetsarbetet.

Hur stora svenska företag har implementerat AI i sitt cybersäkerhetsarbete är relativt outforskat ur ett socioteknisk perspektiv. Forskningen inom detta område är som tidigare nämnt fokuserad kring den tekniska prestandan hos AI-modellerna istället för faktorer som kommer gynna organisationer i längden. Detta skapar en klyfta mellan teknisk potential och organisatoriskt värde, vilket bör undersökas vidare eftersom den organisatoriska aspekten är central ur ett cybersäkerhetsperspektiv.

1.3 Syfte och forskningsfråga

Syftet med studien är att undersöka hur stora svenska företag har implementerat och använder AI för att förbättra deras cybersäkerhet. Dessutom syftar studien till att skapa en översikt av potentiellt gynnsamma effekter ett socioteknisk förhållningssätt kan medföra, detta görs genom att analysera samspelet mellan tekniska, mänskliga och utomstående faktorer som påverkas av att implementera AI i ett cybersäkerhetssystem.

1.3.1 Frågeställningar

- Hur använder stora svenska företag AI för att förbättra cybersäkerheten?
- Hur kan implementationen av AI inom cybersäkerhet gynnas av ett socioteknisk förhållningssätt för dessa företag?

1.4 Avgränsning

Denna studie avgränsas till Sverige och stora svenska företag. Detta grundades dels i att förenkla datainsamlingen då forskningen genomfördes i Sverige, dels för att en tidigare undersökning inom det aktuella området inte har genomförts i denna kontext. Definitionen av stora företag är enligt SCB (2024) att de har mer än 250 anställda eller en årsomsättning över 500 miljoner kronor och det finns 1200 registrerade stora företag i Sverige 2024. Stora företag valdes på grund av deras kapacitet och goda ekonomiska förutsättningar att implementera AI för cybersäkerhet, vilket är ett relativt nytt samt växande fenomen. Generaliserbarheten kan även tänkas gynnas av denna avgränsning tack vare eventuella kulturella och organisatoriska likheter mellan stora svenska företag.

Slutligen avgränsas företagens användning av AI som berör cybersäkerhet i denna studie, även om företagen som deltagit använder AI-teknik inom flera andra sektorer bedöms detta inte vara relevant för att uppfylla studien syfte.

2. Teori

I följande kapitel introduceras den sociotekniska teorin, som utvecklades för att förstå samspelet mellan teknik och mänskliga faktorer vid implementering av nya system. Fokus ligger särskilt på hur denna teori tillämpas inom cybersäkerhet, där balansen mellan tekniska innovationer och organisatoriska behov är avgörande för att hantera komplexa säkerhetshot. Vidare sammanställs tidigare forskning inom området AI och cybersäkerhet.

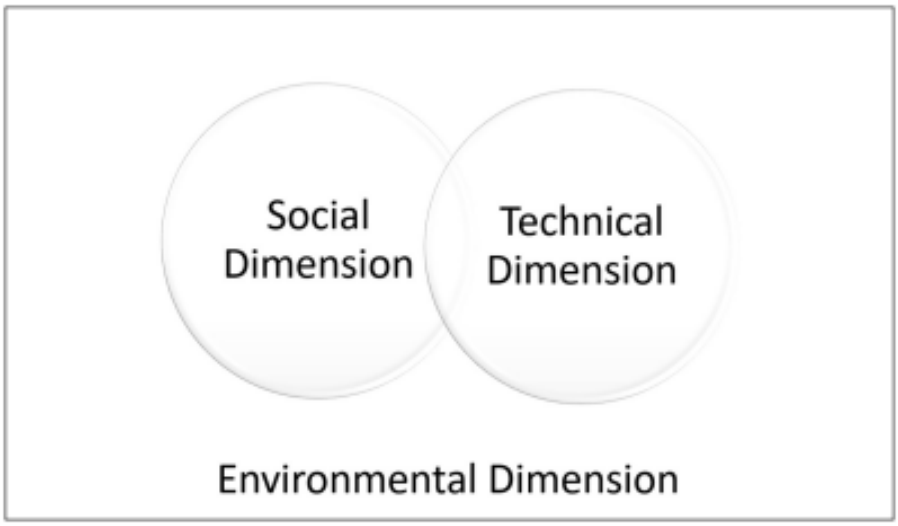
2.1 Bakgrund av Socio-Technical Systems Theory

Den sociotekniska teorin uppstod efter andra världskriget för att möta de växande utmaningarna med implementeringen av ny avancerad teknik, som ofta inte tog tillräcklig hänsyn till de människor som använde den. Ett tidigt exempel är Trist och Bamforths (1951) arbete vid kolgruvor, där en balans mellan teknikens möjligheter och arbetsförhållandena för de anställda skapades. Resultatet blev en hållbar arbetsmiljö och en mer effektiv kolgruva. Senare anpassade Mumford (1979) dessa lärdomar till systemutveckling genom sina ETHICS-principer, ett ramverk för design av informationssystem där både teknikens potential och användarnas arbetsmiljö beaktades. Detta holistiska synsätt vid design och implementation kallas för ett sociotekniskt förhållningssätt (ibid).

Lyytinen och Newman (2008) undersökte hur implementeringen av nya informationssystem påverkar organisationer ur ett sociotekniskt perspektiv. De konstaterade att effekterna kan vara gradvisa, med mindre påverkan, eller abrupta, där tekniken innebär omfattande förändringar för både användare och organisation. En abrupt teknik förändrar ofta arbetsprocesser, relationer och strukturer på djupgående sätt och kräver en noggrann balans mellan tekniska och sociala faktorer. Denna insikt är central för att förstå behovet av ett integrerat perspektiv vid design och implementering av komplexa system vilket ett sociotekniskt förhållningssätt möjliggör (ibid).

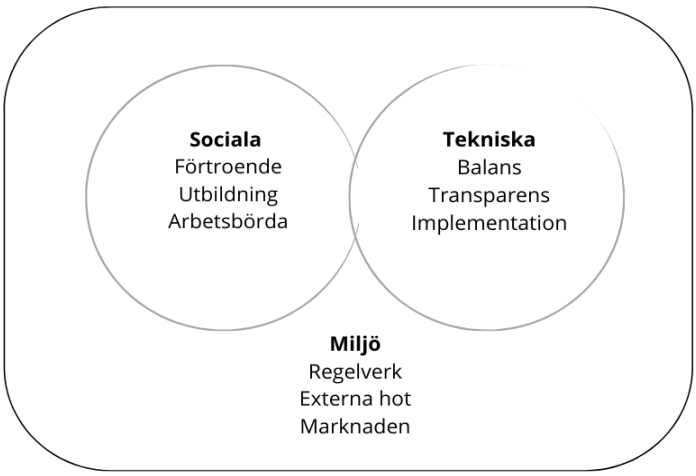
Kolevski och Michael (2015) byggde vidare på dessa principer och definierade hur ett cybersäkerhetssystem bör utformas utifrån tre grundpelare. Den första pelaren handlar om det sociala: människors beteende, samarbete och organisatoriska strukturer. Den andra pelaren täcker de tekniska aspekterna, såsom de verktyg och tekniker som används inom organisationen. Slutligen inkluderar den tredje pelaren den externa kontexten, som påverkar organisationen från omvärlden. Detta ramverk ger en helhetsbild av hur cybersäkerhet kan utformas på ett hållbart sätt genom ett sociotekniskt förhållningssätt (ibid).

2.1.1 Anpassning av teorin



Figur 1. Illustrering av Socio-Technical-Enviromental Dimensions (Källa: Michael, K., Vogel, K. M., Pitt, J., & Zafeirakopoulos, M. (2024). Artificial intelligence in cybersecurity: A socio-technical framing)

Figur 1 visar hur Michael et al. (2024) anpassade teorin när de genomförde en omfattande litteraturöversikt för att kartlägga hur AI används inom cybersäkerhet. Studien, som baserades på den sociotekniska teorin, utvidgade Kolevski och Michaels (2015) tre grundpelare för att specifikt adressera de nya utmaningar och möjligheter som AI innebär. De betonade vikten av att holistiskt integrera tekniska innovationer med sociala och organisatoriska faktorer för att skapa robusta cybersäkerhetssystem som kan hantera dagens komplexa hotlandskap. Studien resulterade i tio identifierade kunskapsgap och rekommendationer för framtida forskning gällande AI, cybersäkerhet och ett socioteknisk förhållningssätt (ibid).



Figur 2. Illustrering av studiens modifierade Socio-Technical-Enviromental Dimensions med tillhörande teman

Figur 2 visar hur denna studie har anpassat STS teorin. Med Michael et al. (2024) som utgångspunkt användes de tre grundpelarna för ramverket: sociala, tekniska och miljömässiga. Vidare identifierades relevanta teman för vardera grundpelare utifrån tidigare

forskning, som sedan itererades i flera steg, se avsnitt 3.2.2. Dessa teman ligger som grund för studiens intervjufrågor (se bilaga 1), vilket möjliggör en gedigen insamling av empiri för kommande sociotekniska analys.

2.2 Tidigare forskning

Tidigare forskning med socio-teknologisk tillämpning visar hur det kan uppstå ett gap där nya tekniska system, särskilt vad gäller komplexa tekniker som AI, ofta misslyckas med att möta sociala krav. Ehsan et al. (2023) lyfter detta gap XAI och föreslår ett ramverk för att förstå hur tekniska och sociala faktorer samspelar i design och användning av AI-lösningar. Ramverket bygger på två huvudsakliga komponenter: tekniska delen, som analyserar aspekter såsom datakvalitet och begriplighet gentemot AI-modellen. Tillsammans med den sociala delen, som fokuserar på användarnas tillit och förmåga att ta till sig, förstå och använda det som genereras för att fatta beslutet (ibid). Kolevski och Michael (2015) presenterar liknande insikter, där vid analys av cybersäkerhetssystem identifieras liknande gap mellan tekniska och sociala krav. De argumenterar för hur sociala, tekniska och miljömässiga aspekter måste balanseras för ett effektivt system.

2.2.1 Tidigare forskning om Sociala teman

Den sociala aspekten omfattar de mänskliga och organisatoriska aspekterna och hur de påverkas implementeringen av artificiell intelligens inom cybersäkerhet. Enligt den socio-tekniska teorin är det avgörande att förstå hur människor interagerar med tekniken för att skapa ett balanserat och fungerande system (Michael et al., 2024). Detta innefattar vikten av att ta hänsyn till anställdas attityder, kompetens och arbetsmiljö vid införande av ny teknik inom organisationen.

Arbetsbörda

Arbetsbördan påverkas av hur AI implementeras i organisationers arbetsflöden. Michael et al. (2024) förklarar att AI kan effektivisera arbetsuppgifter genom att automatisera rutinmässiga moment, vilket frigör tid för mer värdeskapande arbete. Samtidigt beskriver Rahman et al. (2021) att mänskliga misstag ofta är den svagast länken inom cybersäkerhet och att AI kan minska risken för fel genom att ta över enklare samt tidskrävande arbetsuppgifter. De identifierar även ett ökat behov av effektiva cybersäkerhetssystem, där bristande resurser kan åtgärdas med hjälp av AI. Däremot framhäver Rahman et al. (2021) att implementeringen av AI-lösningar kan innebära en initial arbetsbörda, särskilt när det gäller utbildning av medarbetare. AI kan avlasta vissa arbetsuppgifter men behovet för mänsklig kontroll kvarstår. Artikeln understryker behovet av en balansgång mellan teknisk avlastning och mänsklig övervakning (ibid).

Förtroende

Förtroende är centralt för att teknik ska accepteras och användas effektivt i organisationer. Bristande transparens hos AI-lösningar kan skapa osäkerhet och motstånd bland anställda, vilket Michael et al. (2024) lyfter som en betydande utmaning. För att motverka detta kan förklarande AI-modeller (XAI) användas, då dessa modeller kan förklara förslag och beslut som presenteras på ett utförligt sätt, vilket kan bidra till ökad tillit och säkerhet hos organisationen. Capuano et al. (2022) undersöker vilka problem som uppstår i ett cybersäkerhetssammanhnag om AI inte förklarar sina beslut. Egenskapen att förmedla vilka beslut som tas och varför är enligt författarna grundläggande för att skapa förtroende hos

användaren. De betonar även hur XAI kan öka förtroendet hos tekniska liksom otekniska användare (ibid).

Utbildning
Utbildning är avgörande för att anställda ska kunna anpassa sig till och nyttja den fulla potentialen av AI. Vid implementering av nya tekniker, såsom AI, krävs det att organisationer bygger upp nödvändig kompetens hos de anställda för att de ska kunna förstå och hantera tekniken (Michael et al., 2024). Utan tillräcklig utbildning riskerar organisationer att möta både tekniska och sociala utmaningar, såsom bristande kunskap och felaktigt användande av verktygen, vilket kan sänka produktiviteten och öka säkerhetsriskerna. Rahman et al. (2021) understryker behovet av kontinuerlig utbildning hos medarbetare, särskilt inom områden som hotdetektion och användning av AI-lösningar, för att stärka organisationers resiliens mot cyberattacker samt minska risken för mänskliga misstag.

2.2.2 Tidigare forskning om Tekniska teman

Tekniska aspekter och teman

Den tekniska aspekten för AI inom cybersäkerhet behandlar inte enbart tekniken, utan hur den är en del av ett socioteknisk system. De tekniska komponenterna består av hårdvara, mjukvara, de mänskliga användarna och processerna som binder samman dem. AI ses som ett komplement till mänskligt beslutsfattande, inte dess ersättare (Michael et al., 2024).

Balans

En central tanke inom ett socioteknisk cybersäkerhetssystem som använder AI är det ökade stöd som teknologin möjliggör (Michael et al., 2024). AI har förmågan att analysera stora mängder data och identifiera mönster som är omöjliga för en människa att upptäcka. Detta ger ett bredare och mer välgrundat beslutsunderlag. Författarna betonar dock att AI:s främsta roll är att förstärka människors förmågor, inte att själv fatta avgörande beslut. Därför framhålls behovet av mänsklig ledning som en kritisk faktor (ibid).

Michael et al. (2023) diskuterar även "automation paradox", som beskriver det motsägelsefulla förhållandet där ökad automation leder till mer arbete. Inom cybersäkerhet uppstår detta när användningen av AI skapar ett större behov av mänsklig övervakning och analys för att säkerställa att AI:s beslut är korrekta och lämpliga (ibid).

Transparens

Förhållandet mellan XAI och BB-struktur inom AI har diskuterats i flera år. Michael et al. (2024) framhåller att möjligheten att tydligt förklara sina beslut är avgörande för att användarna ska känna förtroende för systemet. Om systemet inte kan förklara sina beslut uppstår en säkerhetsrisk, eftersom det blir omöjligt att kontrollera om besluten är korrekta. En lösning på denna problematik är XAI, som med hjälp av en LLM kan motivera sina beslut och förslag på ett begripligt sätt (ibid).

Zhang et al. (2022a) utforskar hur XAI kan stärka användarnas förtroende och främja cybersäkerheten. Artikeln betonar vikten av transparens i beslutsprocessen för att öka förtroendet. Samtidigt lyfts utmaningen att implementera XAI utan att introducera nya säkerhetsrisker. När användaren får bättre insyn och förståelse för systemet kan dess

funktioner anpassas till verksamhetens behov, vilket bidrar till ett mer robust säkerhetssystem (ibid).

Implementation
Tekniken som skall implementeras bör vara väl anpassat mot existerande system menar Michael et al. (2024). De rekommenderar att utforma ett socio-tekniskt ramverk för att på ett effektivt och hållbart sätt integrera AI mot existerande system. Michael et al. (2024) undersöker hur ett system bäst bör implementeras ur ett socioteknikskt perspektiv. De betonar vikten av att involvera flera intressenter såsom tekniska experter, policyskapare och slutanvändare för att tillse att samtliga behov uppfylls. De förespråkar även en modulär arkitektur för AI-lösningar så att de enkelt kan integreras med befintliga cybersäkerhetssystem (ibid).

Dehghantanha et al. (2024) menar att en effektiv implementering av AI-lösningar bör ske stegvis, det vill säga, manuella, halvautomatiska och slutligen automatiska. Detta för att säkerställa att integrationen med resterande säkerhetssystem fungerar samt att användarna involveras i implementationsprocessen. Författarna förespråkar även att implementering bör ske där AI:s förmåga att analysera stora mängder data kan nyttjas, såsom vid hotdetektion och anomaliövervakning (ibid).

2.2.3 Tidigare forskning om Miljömässiga teman

Den miljömässiga aspekten innefattar de yttre faktorerna som påverkar användningen av AI för cybersäkerhet, bland annat hotbild, utbud och styrande regelverk (Michael et al., 2024).

Hotbild
Michael et al. (2023) beskriver hur AI:s dubbelsidiga natur är ett problem som behöver hanteras i dagens cybersäkerhetsarbete. De menar att AI användas som ett skydd samtidigt som det tekniken används i skadliga syften. De belyser även risken när digitaliserade och komplexa system kan ge upphov till ett beroende av AI vilket kan göra systemen mer sårbara, det vill säga ett överberoende. Samtidigt uppstår risker om användningen av AI är otillräcklig vilket kan resultera i att företaget står oförmögna att bemöta dagens cyberattacker.

Zhang et al. (2022b) förklarar även att det finns olika sätt för angripare att utnyttja AI-lösningar. Det kan vara både med hjälp av AI eller att manipulera AI-modellen. Angripare kan hitta sårbarheter i AI-modellerna under deras utvecklingsfas som dem senare kan utnyttja vid attacker. Med hjälp av AI kan angripare även överbelasta och förfina traditionella metoder för att göra dem svårare att upptäcka.

AI-tekniken medför även nya möjligheter för cyberangrepp och andra former av externa hot. Dessa hot kommer bland annat i form av social engineering och phishing-mail som blir mer sofistikerade med AI (Michael et al., 2024).

Marknad
Zhang et al. (2022b) menar att AI:s tillväxt är väsentlig ur ett säkerhetsperspektiv. Samhällets ökade tekniska sammankoppling och komplexitet ställer höga krav på säkerhetslösningar hos företag. AI:s förmåga att hantera stora datamängder och identifiera hotbilder eller avvikande nätverkstrafik kan bli avgörande för cybersäkerheten. Författarna påpekar att AI erbjuder konkurrensfördelar genom minskad reaktionstid och förbättrad detekteringsförmåga.

Capuano et al. (2022) undersökte hur marknaden för AI och cybersäkerhet kommer att utvecklas de närmaste åren. De konstaterade att intresset för området förväntas öka markant, vilket skapar attraktiva investerings- och utvecklingsmöjligheter. Kombinationen av AI och cybersäkerhet ger utrymme för innovation och stärker skyddet mot avancerade cyberattacker.

Supply chains utgör ett komplext system av sammankopplade organisationer och aktörer som är beroende av varandra. Detta medför en stor angreppsyta där en lyckad cyberattack kan få omfattande konsekvenser för hela ekosystemet. Dessa risker ställer höga krav på säkerhetslösningar och är en drivande faktor bakom utvecklingen av avancerade tekniker och regelverk för cybersäkerhet (Michael et al., 2024). Konkurrens är också en central drivkraft, eftersom företag med välfungerande cybersäkerhet kan vinna fördelar på marknaden. Kundernas förväntningar på hög säkerhet och konfidentialitet blir därmed en nyckelfaktor för framgång (ibid).

Regelverk

Capuano et al. (2022) diskuterar hur regelverk kan försvåra implementeringen och användningen av XAI i cybersäkerhetsarbetet. Idag stöter många organisationer på problem vid användandet av AI eftersom det finns en bristande transparens och det är en stor utmaning kring regelbestämmelser.

Zhang et al. (2022a) förklarar att EU:s GDPR ställer krav på förklarade beslut av AI-modeller när det handlar om påverkan på individer, då förklarande av beslut är en kärnaspekt av XAI är den högaktuellt. Enligt Recital 71 i GDPR har individen rätt att bestrida automatiska beslut och kräva en förklaring till dem. Författarna förklarar även att drivkraften att integrera XAI i sina AI-lösningar handlar om att efterlevnaden till olika regelverk som rättvisa och förklarbarhet uppfylls (ibid).

Michael et al. (2024) understryker vikten av att reglera hur AI ska implementeras och vilka säkerhetskrav de ska uppfylla. Dessa regelverk finns i form av EU:s nya riktlinjer AI Act och genom rekommendationer från säkerhetsinstitut som National Institute of Standards and Technology (NIST) (ibid).

3. Metod

I följande kapitel diskuteras och förklaras tillvägagångsätten som användes i studien. Metoddelen ger även insikt i hur empirin och det insamlade materialet analyserades och teman utformades samt vilka åtgärder som vidtogs för att säkerställa en högkvalitativ studie.

3.1 Forskningsstrategi- och design

3.1.1 Forskningsstrategi

För att besvara forskningsfrågorna har en kvalitativ forskningsstrategi använts. Enligt Bryman (2018) är kvalitativ forskning lämplig när forskare vill förstå hur människor tolkar och upplever specifika fenomen i dess naturliga kontext. Denna strategi passar studien väl då det undersöks hur representanter från svenska företag uppfattar och beskriver sina erfarenheter av att implementera AI inom cybersäkerhet.

Vidare har en deduktiv ansats använts som utgångspunkt för studien. Bryman (2018) förklarar att deduktion används när forskare utgår från en befintlig teori för att formulera forskningsfrågor, strukturera datainsamlingen och analysera resultaten. I denna studie har Socio-Technical Systems Theory anpassats och använts som utgångspunkt för att belysa samspelet mellan människa, teknik och organisation. Den deduktiva ansatsen säkerställer att dessa processer inte enbart beskrivs, utan även ge möjlighet att förstå och tolka dem i relation till ett befintligt teoretiskt ramverk.

En viktig aspekt vid valet av kvalitativ forskning är ämnets karaktär. Implementering av AI för cybersäkerhet är ett relativt nytt fenomen och ett område som präglas av hög sekretess. Det är endast ett fåtal nyckelpersoner, ofta Chief Information Security Officers (CISO), som är direkt involverade i och har insikter om implementeringsprocessen av nya tekniker kopplat till cybersäkerhet. Enligt Bryman (2018) används mindre och mer fokuserade urval i kvalitativ forskning för att möjliggöra djupgående analyser av enskilda fall, vilket passar denna studie då syftet att är förstå dessa nyckelpersoners perspektiv och erfarenheter.

I kontrast kräver kvantitativ forskning ofta ett större urval för att kunna påvisa statistisk signifikans och generalisera resultaten till en större population (Bryman, 2018). Eftersom studien riktar sig mot en mindre grupp med nödvändig expertis inom organisationer, är det inte lämpligt att samla in data från ett stort antal respondenter. Istället fokuserar studien på att fånga upp djup och komplexitet i respondenternas erfarenheter.

Genom användandet av kvalitativa metoder kan centrala aspekter av AI-implementeringen utforskas, såsom beslutsfattande, organisatoriska utmaningar och samspelet mellan teknik samt människa. Strategin säkerställer en detaljerad och nyanserad förståelse av fenomenet, vilket är avgörande för att uppnå studiens syfte samt besvara forskningsfrågorna.

Denna design gör det möjligt att undersöka fenomenet på djupet och relatera det till den valda teorin. Genom att inkludera respondenter med god expertis från flera företag kan studien skapa en mer nyanserad förståelse av fenomenet.

3.2 Datainsamling

3.2.1 Insamling av teoretiskt material

Det teoretiska materialet som ligger till grund för denna studie har valts utifrån dess relevans till forskningsfrågorna och den teoretiska ramen. Insamlingen genomfördes genom sökningar i databaserna Scopus, ACM Digital Library och Google Scholar, som är framträdande inom forskning kring teknik och informationssäkerhet. Dessa databaser valdes för att säkerställa tillgång till forskning som är både relevant och av hög kvalitet. För att finna relevanta och heltäckande artiklar användes sökord som: ”Cyber security”, ”AI” och ”Human factors” samt närliggande synonymer. Se tabell 1 nedan för en mer specifik bild av använda sökord.

Sökord	**Synonymt begrepp**
AI	Artificial Intelligence, Machine learning & Deep learning
Cyber security	Information security & security
Human factors	Social factors
Implementation	Adoption
Challenges	Risks
Socio-technical	STS, Socio-technical systems theory & Socio-technical framework

Tabell 1. Lista över sökord som använts i olika databaser samt tillhörande synonyma begrepp.

I det initiala skedet av sökningen användes relativt snäva kombinationer av dessa sökord för att hitta så relevanta och heltäckande artiklar som möjligt. Dessa sökord tillsammans med närliggande synonymer utgör centrala nyckelord inom forskningsområdet och reflekterar studiens fokus på både tekniska och sociala aspekter. Däremot resulterade de väldigt inriktade sökningarna till ett begränsat antal träffar. För att överkomma detta och få en rättvis spegling av dagsläget, justerades strategin på ett bredare och mer strukturerat sätt. I detta skede användes mindre inriktade söksträngar i enlighet med Bryman (2018) som föreslår att använda olika kombinationer av sökord för att vidga omfattningen av relevanta artiklar. Genom att iterativt optimera söksträngarna till delområden av ämnet kunde ytterligare relevanta artiklar identifieras, även om de inte behandlade samtliga delar av forskningsområdet.

Vid sökandet av material applicerades följande inkluderings- och exkluderingskriterier för att säkerställa att sökresultaten var av god kvalitet och relevans:

1. Artiklar som berör artificiell intelligens är publicerade år 2020 eller senare. Detta är med grund för att LLM:s blivit alltmer framträdande i olika branscher på senare år, vilket gör denna avgränsning viktig för studiens trovärdighet. Däremot inkluderades äldre artiklar som behandlar teorin för att ge en tydlig bild av dess uppkomst och vidare hur den appliceras samt operationaliseras i denna studie.

2. Det användes enbart artiklar skrivna på engelska för att säkerställa en bred internationell relevans.

3. För att materialet skulle hålla god vetenskaplig kvalitet användes endast artiklar som är peer-reviewed, vilket innebär att de har skrivits och granskats av välrenommerade forskare innan publicering.

4. Artiklar med stort fokus på teknikens funktionalitet har exkluderats då dessa inte ansågs bidra till att besvara forskningsfrågorna.

5. Artiklar med ett högt antal citeringar har prioriterats då detta tyder på att de är omtalade och av hög kvalitet.

Under litteratursökningen utvärderades artiklarna genom en noggrann granskning av abstract och slutsats för att snabbt avgöra innehållets relevans till studiens syfte. Därefter genomfördes en mer omfattande läsning av de artiklar som bedömdes relevanta. Vidare sammanställdes punktlistor av artiklarnas innehåll i ett separat dokument för att snabbt kunna få en tydlig överblick av likheter och skillnader i befintlig litteratur.

3.2.2 Utformning av teman och intervjufrågor

Vid utformandet av intervjufrågorna användes den teoretiska grunden från Kolevski och Michaels (2015) samt Michael et al. (2024), se avsnitt 2.1.1, där tre centrala grundpelare användes. Dessa användes sedan under litteratursökningens gång för att välja ut relevanta studier som därefter sammanställdes och kodades tematiskt för att slutligen placeras under tillhörande grundpelare. Dessa teman var grunden för utformandet av intervjufrågorna för att möjliggöra kommande analys av resultatet och besvarandet av studiens frågeställningar.

3.2.3 Urval

I studien genomfördes ett strategiskt urval av respondenter till studien där flera års erfarenhet, arbete på ett stort svenskt företag samt goda kunskaper inom cybersäkerhet och AI prioriterades. Detta resulterade i en smalare omfattning av potentiella respondenter men detta ansågs nödvändigt för att få tillräcklig insikt för att besvara studiens frågeställning. Bryman (2018) menar att ett strategiskt urval innebär att deltagare väljs ut baserat på hur väl deras kunskap kan bidra till studiens forskningsfokus.

Chefen för informationssäkerhet har en unik inblick i både tekniken och den organisatoriska strukturen inom de stora svenska företagen. Deltagarna var lämpade för undersökningen och bidrog med en unik insikt i det sociotekniska samspelet studien syftar till att undersöka.

Respondent	Yrkesroll	Erfarenhet	Intervjutid
1	Lead IT Security Architect	22 år	75 minuter
2	CISO	25 år	60 minuter

Tabell 2. Lista över respondenter som innefattar benämning, yrkesroll, antal år i branschen samt tidslängden på intervjun.

3.2.4 Insamling av empiriskt material

I denna studie bildades en intervjuguide inför intervjuerna (se bilaga 1). För att utforma frågorna användes de teman som tagits fram från teorin. Bryman (2018) berättar om riktlinjer att följa vid designen av frågor. Den första är att formulera frågor som är ospecificerade för

att ge utrymme till respondenten att ha alternativa idéer och synsätt. De andra är att inte ställa ledande frågor då detta begränsar respondentens frihet i dennes svar. Utifrån de råden och de analyserade teman från teorin kunde intervjufrågor utformas. Utformningen ledde i slutskedet till att det blev elva frågor om de aktuella teman och tre generella frågor som grundades. Bryman (2018) påpekar även att det bör finnas en röd tråd i frågorna gällande individuellt tema för tema. Detta för att frågorna ska följas upp på ett naturligt sätt och för att kunna förhålla sig till något ifall respondenten byter riktning under en fråga.

Bryman (2018) påpekar att intervjuer är den vanligaste metoden vid kvalitativ forskning och i denna studie ämnade sig den semistrukturerade intervjun som den lämpligaste. I en semistrukturerad intervju bör respondenten ges stor frihet att utforma sina svar. Det är också viktigt att intervjuaren är öppen och använder intervjuguiden som ett stöd för att beröra specifika teman, men att frågorna inte behöver ställas i en fast ordning. Följdfrågor som anknyter till respondentens svar kan också ställas i en semistrukturerad intervju (ibid).

Anledningen till att den semistrukturerade intervjun ansågs mest lämplig var för att den överensstämmer väl med en kvalitativ forskningsansats, vilket var en nyckelaspekt för författarna vid valet av intervjumetoden. Metoden är dessutom utformad till en friare tankegång, vilket är fördelaktigt för denna studie eftersom respondenterna då styrs mot att ge svar som speglar deras egna uppfattningar och perspektiv. Att ge dem frihet i svaren är avgörande för studien, eftersom respondenterna är sakkunniga med lång erfarenhet inom det studerade ämnet, samtidigt som deras svar kan vara svåra att förutse (ibid). Vidare bedömdes vissa av studiens frågeställningar vara av känslig eller konfidentiell karaktär, vilket innebär en risk att vissa frågor inte besvaras. För att minimera denna risk var den semistrukturerade intervjun särskilt lämplig, eftersom den gav respondenterna möjlighet att avvika från den ursprungliga frågan och istället svara på det de känner sig bekväma med (ibid).

För att intervjun skulle bli så framgångsrik som möjligt togs Brymans (2018) lista över tio krav för en lyckad intervju i beaktande. Exempel på egenskaper som anses viktiga för en intervjuare är att denne är öppen, tydlig, insatt och sensitiv. Utifrån dessa kriterier kunde intervjun ge studien den meningsfulla och goda responsen som behövdes för genomförandet.

Bryman (2018) påpekade vikten av att avgöra huruvida en transkribering var möjlig utifrån tidsramen för undersökningen. Med tanke på att det rörde sig om två intervjuer och eftersom respondenterna tillät att intervjuerna spelades in, beslutades det även att en transkribering skulle genomföras. Anledningen till detta var för att underlätta analysen av respondenternas svar i förhållande till studiens utformade teman och eftersom det var av vikt att kunna återkomma till de transkriberade intervjuerna för en noggrannare granskning.

3.2.5 Etiska överväganden

Eftersom studiens datainsamlingsmetod är baserad på intervjuer är det av vikt att följa och överväga etiska riktlinjer. Enligt Bryman (2018) finns det några grundläggande etiska principer som forskning bör förhålla sig till. De fyra omfattande principerna som är relevant för studien är uppdelade i olika krav:

1. Informationskravet innebär att deltagarna ska informeras om studiens syfte och de olika delarna som ingår i undersökningen. De ska även informeras om att deras deltagande är helt frivilligt och att de får hoppa av när som helst utan att ange anledning.

2. Samtyckeskravet innebär att deltagarna har individuell rätt att påverka och bestämma över deras medverkan.
3. Konfidentialitetskravet innebär att personuppgifter som samlas in från de medverkande ska förvaras under största möjliga konfidentialitet.
4. Nyttjandekravet innebär att informationen som samlas in om individuella personer får enbart användas i det specifika forskningssyftet.

Innan intervjuerna upplystes respondenterna om dessa principer mer konkret för att göra de medvetna om studien och öka deras trovärdighet gentemot den. Vidare gav detta en god grund för genomförandet av intervjuerna att lyckas.

3.3 Dataanalys

3.3.1 Analysmetod

Studien använder en kvalitativ dataanalysmetod, vilket innebär att empirin analyserades och data delades upp tematiskt inför den kommande diskussionen. Empirin från genomförda intervjuer var inledningsvis ljudfiler som sedan manuellt transkriberades av författarna. Denna data är grunden för artikelns tematiska analys vilket Bryman (2018) menar är en passande metod för att identifiera, analysera och rapportera teman i kvalitativa data.

Urvalet av relevant data som sedan analyseras och presenteras i resultatet genomfördes deduktivt utifrån de teman som tidigare presenterats i teoriavsnittet (ibid). Data som inte passade in tematiskt men bidrog till studien adderades till resultatet och möjliggjorde en djupare diskussion. Detta tillvägagångssätt menar (ibid) stärker en studies trovärdighet då data inte inkluderas enbart efter hur väl den passar teorin.

Kodningsprocessen genomfördes i flera steg, inledningsvis skapades koder för att kategorisera empirin från den första intervjun under de tre teoretiska kategorier. Dessa koder generaliseras sedan till tre teman per kategori. Denna iterativa process möjliggör en mer direkt koppling mellan teman, forskningsfrågorna och data (Bryman, 2018).

3.3.2 Kvalitetskriterier

Bryman (2018) skriver om trovärdighet och överförbarhet som viktiga kriterier vid kvalitativ forskning för att öka studiens tillförlitlighet. För att säkerställa trovärdighet i studien har den genomförts med erkända kvalitativa metoder för datainsamling och analys. Den tematiska analysen har därefter diskuterats och granskats av handledare och andra studenter för att säkerställa korrekt tolkning av datamaterialet. Respondenterna har även erbjudits möjligheten att granska data innan publiceringen för att bekräfta att deras svar har uppfattats korrekt. Överförbarheten har säkerställts genom att en detaljerad beskrivning av forskningskontexten, tidigare forskning och de metodologiska val i form av intervjuguide och det teoretiska ramverk som använts vid analysen. Denna beskrivning underlättar för framtida forskare att bedöma resultatets relevans och användbarhet samt möjligheten att reproducera studien i framtida forskning.

3.3.3 Metodologiska begränsningar

Det begränsade urvalet av intervjudeltagare kan påverka studiens generaliserbarhet, vilket är typiskt för kvalitativ forskning (Bryman, 2018). Trots denna begränsning är metoden relevant för studiens syfte tillsammans med de åtgärder som tagits för att säkerställa kvaliteten ansågs

det att empirin är ett tillräckligt underlag för en djup analys och diskussion som kan besvara frågeställningen. Anledningen till detta är för att intervjuerna var båda mer än 60 minuter, vilket genererade en stor mängd analyserbart material. Det ska även nämnas att respondenterna hade en hög grad av expertis inom området, vilket gav en väldigt nyanserad bild av ämnet. Respondenternas gemensamma längd av erfarenheter inom ämnet gav även studien det djup den behövde och eftersom de hade liknande roller på sina respektive stora företag kunde generaliserbarheten för empirin stärkas. Studiens avgränsning som handlade om att studien enbart undersökte stora företag är även av vikt för relevansen eftersom det eventuellt finns organisatoriska och kulturella likheter i denna företagskategori, vilket styrker generaliserbarheten för studiens ändamål.

4. Resultat

Studiens empiriska resultat presenteras genom en beskrivning av hur AI används inom cybersäkerhet, följt av en tematisk sammanställning av de effekter som implementeringen av AI har för att förbättra cybersäkerheten.

4.1 Hur använder företagen AI?

Den empiriska undersökningen visar att samtliga deltagare i studien använder AI inom sina verksamheter och belyser fördelar och utmaningar med tekniken. AI implementeras för att effektivisera arbetsflöden och stärka cybersäkerheten genom automatisk hotdetektion, men tekniken påverkar även den sociala kulturen och organisatoriska strategier för att bemöta utomstående hot och förväntningar. Resultatet möjliggör en djupare analys runt samspelet mellan de sociala, tekniska och miljömässiga aspekterna.

4.2 Vilka effekter har implementationen medfört?

4.2.1 Sociala aspekter

Arbetsbörda

Respondent 2 anser att AI har effektiviserat många arbetsprocesser utan att arbetsbördan har ökat. Ett exempel som lyfts från respondenten är ett AI-stöd vid loggövervakning som ska upptäcka anomalier i nätverkstrafik. Där har kvaliteten och säkerheten förbättrats med en oförändrad arbetsbörda. Samtidigt belyser respondent 2 vikten av att det fortfarande behövs ett par mänskliga ögon för att göra den slutliga bedömningen men att själva granskningen eller analysen kan göras av en AI-modell.

"Det som har hänt är att kvaliteten har ökat med egentligen samma insats. I och med nu fångas mönstren. Vår implementation bygger på att all data skickas till en tjänst som har ett AI-stöd som kan hitta anomalier. Sen går det här vidare till en SOC som fångar upp det här och gör en egen slutlig bedömning. Vad tror vi om det här? Är det på riktigt? Är det en false positive? Är det någonting som vi verkligen borde fånga upp?" (Respondent 2).

Båda respondenterna är gemensamt överens om att AI har ökat effektiviteten i aspekter som att hitta angripare snabbare, men båda är fortfarande övertygade om att den mänskliga övervakningen kommer vara väsentlig många år framåt eftersom AI-modellerna inte ännu går att lita på fullt ut. Respondent 1 betonar att de har haft ett HL tillvägagångssätt när det kommer till AI-lösningar för att säkerställa att besluten görs enligt säkerhetsriktlinjer och att kvaliteten är god. Dock finns det nackdelar med det eftersom effektiviteten hade varit mer betydande ifall det gjordes automatiskt.

"Det är ju så att human in the loop är någonting som vi stått på mycket. Och just den här balansen med mänsklig övervakning, då kanske man tappar lite effektivitet jämfört med om det är automatiskt." (Respondent 1).

Respondent 1 nämner även en annan aspekt av arbetsbörda, vilket är att det finns anställda som aktivt jobbar med att hitta användningsfall för användning och implementering av AI i verksamheten.

Förtroende

Respondent 1 anser att en modell för förklarliga AI-beslut är en bra väg framåt när det kommer till AI-lösningar. Vidare pratar respondenten om utvecklingen i ChatGPT och hur den har förbättrats genom att den kan förklara vart den fått sina källor ifrån.

"J*a men om vi kollar på första släppet av ChatGPT. Så hade ju inte den källreferenser. Det har den ju nu. Det är ju ett perfekt exempel på, nu kan jag i alla fall dubbelkolla vilka källor som den plockade data ifrån. Och ibland är det ju ren jävla dynga liksom. Det är någon bloggpost som en idiot har skrivit någonstans. Men då vet jag ju det i alla fall. Den transparensen fanns inte tidigare. Jag tror att det är ganska mycket. Om man nu kan hitta en bra förklaringsmodell för hur en AI fungerar. Så då är det klart att det är bra. Men den får ju liksom inte bli speciellt komplicerad heller. För då kan man inte ta till sig av den."* (Respondent 1).

Respondent 2 förklarar vikten av att en "blackbox" måste vara tydligt reglerad och regelstyrd för att den ska vara effektiv och säker. Respondent 2 fortsätter även med att modellen bör kunna utvecklas i takt med datapunkter utifrån den nya verkligheten, eftersom data alltid förändras. Algoritmerna måste kunna ändras utifrån för att det ska vara möjligt för en "black box" -modell att hålla sig uppdaterad och för att förtroendet ska finnas där.

"Jag tror att black box-tillämpningen, den måste vara väldigt hårt regelstyrd. Den får inte själv utveckla sina regler utan algoritmerna ska vara statiska. De måste vara, inte förutsägbara, men de måste egentligen bygga på att vi har tänkt någonting. Fånga in det här mönstret, skapa index i bakgrunden eller egentligen gör vad som behövs i bakgrunden för att hitta anomalier, hitta saker och ting som dyker ut, hitta mönster, kunna peka på mönstren eller avvikelse i mönstren." (Respondent 2).

Respondent 1 diskuterar även att seniora utvecklare kan känna sig hotade av AI-lösningar eftersom de gör juniora utvecklare mycket effektivare. En senior utvecklare tenderar att vara mer effektiv eftersom den personen har utvecklat unika kompetenser och erfarenheter under sitt arbetsliv, men de blir mer eller mindre oväsentliga eftersom AI-lösningar har bevisat att de kan förse utvecklare med nödvändig information snabbt.

"Så till exempel om du är ninja på ett specifikt frågespråk för en viss plattform. Och det helt plötsligt kan göras via en AI. Så får du ju kanske ganska mycket social konkurrens på dina kollegor som helt plötsligt kommer upp till samma nivå som dig. Och det kanske kan vara stressigt för vissa och kanske jättepositivt för andra. Du kanske får en person om man är rädd för att bli konkurrensutsatt. Då kan ju den konkurrensutsattheten öka med takt att ditt jobb är enklare att utföra. Och kanske också att du känner att om det kan utföras av någon som är nyexad 23-åring. Jag sitter här och är 40. De är förmodligen mycket billigare än vad jag är. De har inte haft den löneökningscirkus som har pågått i nästan 20 år. Vad händer med mitt jobb nu då?" (Respondent 1).

Respondent 1 påpekar även att det kan uppstå frustration eller social osäkerhet när tidigare omfattande investeringar i vissa arbetsmoment ersätts av automatiserade lösningar från AI. Respondenten fortsätter med att oförutsägbara och obekväma situationer kan uppstå när en persons arbete ersätts av en AI-lösning.

"Det är klart att man kan bli socialt hotad i den rollen. Jag har ju lagt ner all den här jävla tiden på de här playbooks. Kommer du säga att de inte behövs längre? Så det är klart att det finns en socialt utmanande faktor där också." (Respondent 1).

Respondent 2 förklarar att det råder en olik fördelning i förhållande till inställningen av AI-verktyg. Avdelningar som HR, kommunikation och IT har en god inställning till verktygen. Anledningen till detta är att de kreativa delarna av företaget gärna nyttjar AI-verktyg för att maximera sin kreativa förmåga, menar Respondent 2.

"När det är kreativitet inblandat, då finns hela tiden den här viljan att hitta andra ingångar i en kreativitet. Finns det något mer briljant egentligen än en AI där som kan tillföra kreativitet?" (Respondent 2).

Att andra delar av företaget inte är lika framåtlutade gentemot AI-verktyg tror Respondent 2 har att göra med de olika arbetsuppgifterna.

"Men sen om vi tittar på backoffice-relaterade funktioner, där är ju intresset nästan obefintligt för att man jobbar ganska statiskt i processerna. I andra hänseenden så, och det förstår jag genom HR och kommunikation, det är ett helt annat spelfält." (Respondent 2).

Båda respondenterna är överens om att det finns ett bristande förtroende för nya AI-lösningar. Respondent 2 påpekar att förtroendet på dennes företag gentemot en global aktör som Microsoft och deras tjänster är större kontra mot exempelvis OpenAI som fortfarande är en ny etablerad organisation. Respondent 2 fortsätter även med att visa en viss skepticism mot AI-verktyg generellt eftersom det är enkelt att få verktygen att generera felaktig eller fabricerad information.

Jämfört med LLM som litegrann till sakens natur bygger på att börja svara på frågor så kan vi utveckla svaret hur långt som helst beroende på hur mycket vi fortsätter att pusha i promptern. Men där kommer också den här risken som vi har sett ganska så mycket nu i LLM, att det inte är svårt att få en LLM att börja hallucinera och svara fel. Skitenkelt. (Respondent 2)

Respondent 1 fortsätter förklara att förtroendet från kunder gentemot AI-verktyg varierar. Vissa kunder har krävt att data som ertappas från dem aldrig får befinna sig i en AI-lösning. Denna generella skepticism till AI-baserade lösningar finns och Respondent 1 menar att det inte går att lita fullt ut på nya aktörer inom branschen.

"Jag tror än så länge att det kommer nog dröja innan det är ett problem. Men... om det kommer att bli spelare på marknaden så kan det nog bli ett problem. Veta om man kan lita på det leverantören säger att den här AI-modellen kommer att göra." (Respondent 1)

Respondent 2 berättar även om att införandet av AI-lösningar har gett utrymme för så kallat Skugg-IT inom företaget som kan innebära problem. Dessa problem uppstår när de AI-lösningar som förespråkas på företaget inte används, utan att de anställda använder egna. Detta kan bli en säkerhetsrisk ifall konfidentiella data hamnar i öppna AI-plattformar där data inte regleras på samma sätt som i detta fall; deras Microsoft miljö.

"Det här är ju generellt lite bekymmersamt för att om inte vi som arbetsgivare levererar AI som verktyg för de som efterfrågar och vill ha det, oavsett vad drivkraften är, så kommer AI att komma in någonstans ifrån. Bring your own AI. Så det är skugg-IT när det är som bäst. Vi gör inte, som jag var inne på, vi har inte tecknat något avtal med OpenAI för att använda Chat GPT. Däremot så kan vem som helst naturligtvis teckna ett pro-konto med sitt kreditkort

och få tillbaka utlägget från det. Och så använder man kanske det för att skicka upp organisationens information." (Respondent 2).

Utbildning
Respondent 1 förklarar att utbildningsbehovet varierar mycket mellan olika roller på företaget. Exempelvis utvecklare som arbetar med kodgenererande AI-verktyg kräver mer djupgående utbildning för att undvika säkerhetsrisker som kan introduceras genom felanvändning. Däremot finns inte samma behov för anställda som använder sig av lite enklare AI-verktyg, där fokus ligger på att försöka förespråka funktionaliteten och effektiviteten av dem.

"Just nu om vi går lite vid sidan om säkerhets spåret. För just det med att Microsoft Copilot är det största användarfallet vi har. Där är vi ute med en pilotutrundning. Så användarbasen har framförallt IT och seniora chefer som jag tror förstod de här verktygen ganska bra. Så där kände vi inte att vi behövde utbilda särskilt mycket. Men om vi nu ska ta det lite på bredden så måste vi kolla på helt andra utbildningspaket. Kanske specifikt för varje AI-produkt som du får tillgång till." (Respondent 1).

Respondent 1 berättar även om deras "Cyber security awareness-program", som de anställda ska gå på. Detta har resulterat i att de upptäckt bland annat "robocalls", som skulle varit samtal från vd:n men visade sig vara en AI-generad röst som skulle imitera vd:n och lura personer på pengar. Utbildningen fick personer att misstänka detta och det gick fort att identifiera att det var en bluff och kampanj mot företaget.

"Vi har ju faktiskt ett Cyber Security Awareness-program där alla anställda måste gå en nästan två timmar lång utbildning. Där vi informerar om de här grejerna. Så det första som hände var att folk ringde in till vår servicedesk och sa att jag inte tror att det här var Helena liksom. Varför skulle vd:n vara med och ringa mig och be om pengar? Och så kom vi fram till att det pågick en aktiv kampanj mot oss." (Respondent 1).

Respondent 2 har en annan inställning till utbildning i detta sammanhang. På det företaget har de målet att deras riktlinjer ska vara enkla att följa istället för att utbildningen ska vara lång och krånglig. De har utbildat lite grann men deras grundpelare i detta är att "det ska vara lätt att göra rätt".

"Det är nog en av mina käpphästar att det måste vara lätt att göra rätt. Man kan inte sätta upp regelverk som att ni får göra si men inte det och personuppgifter får ni inte skicka om de inte är anonymiserade och så vidare. Jag har sett jättemycket arter på det där men då blir det svårt." (Respondent 2)

4.2.2 Tekniska aspekter

Balans
Respondent 1 förklarar att AI-lösningar har börjat användas på företaget i enklare former och med lätta uppgifter. Det kan vara exempelvis chatt-botar som tar emot frågor och levererar information istället för att man som anställd ska behöva leta upp det som behövs. Respondenten diskuterar även framtiden och AI:s potential i former av automatiserade processer, analys av stora mängder data och stödjande av beslutsfattning inom områden som cybersäkerhet.

"De första funktionerna som började komma var, vi har 3000 sidor dokumentation av någon produkt. Nu har vi en chattbot som du kan ställa frågan till så letar den upp vart någonstans dokumentationen, svaret på din fråga finns." (Respondent 1).

Respondent 2 anser också att deras företag har tillämpat enklare funktioner med stöd av AI. Ett av de viktigaste områdena är textbearbetning, där en AI-lösning har gjort arbetet väldigt effektivt eftersom de producerar enorma mängder text. Ett annat område är faktainsamling, men Respondent 2 är lite osäker på i den aspekten. Anledningen är att en AI kan börja generera fabricerad information, menar Respondent 2. Där finns det anledning till en balans av mänsklig övervakning eller bekräftelse och automatisering. Ett annat område där de arbetar mycket med AI är kodgenerering. Där är det förespråkat att det kan användas men det finns en viktig del att komma ihåg. Det är att all kod ansvaras av personen som använder verktyget, inte själva AI-modulen. Detta är viktigt för att all kod som genereras ska ha en ansvarig och för att den kod som genereras ska vara granskad och godkänd.

"GitHub Copilot använder vi... där gäller det ju då för oss att vara jättetydliga med vem som är ansvarig. Ja, absolut, använd det här... Men kom ihåg, det som ska upp i en commit, det är du som är ansvarig." (Respondent 2)

Det finns även ytterligare en viktig funktion för AI som Respondent 2 nämner, att hitta anomalier i loggar. Den är mycket bättre än det mänskliga ögat på att hitta de minimala förändringarna i data och kan effektivt larma om dem.

"Det är där som jag tänker att SIEM eller loggövervakning med machine learning, alltså AI-stödd hantering, verkligen kan hjälpa till. För där kan system fort finna anomalier i det här mönstret." (Respondent 2)

Dock menar Respondent 2 att det fortfarande finns behov av en HL för att göra den slutliga bedömningen av det identifierade hotet, eftersom detta är AI oförmögen att göra självständigt idag.

Det finns även viktiga AI-funktioner som används för att stoppa cyberattacker på företaget. Respondent 1 berättar om hur de har implementerat ett AI-drivet system som stoppar runt 70 000 phishing-mail per dag. Detta görs helt utan mänsklig övervakning idag och fungerar väldigt effektivt.

"Om vi ser till exempel om vår phishing-hantering, så stoppar vi uppemot 70 000 phishing-anfall per dag automatiskt. Där vill vi inte ha en human-in-the-loop. Då hade det varit helt meningslöst. Så i vissa fall vill vi absolut inte ha något annat än ren automatisering. Men det beror på hur känsligt eller hur stor konfidens man har i att de här skydden faktiskt fungerar." (Respondent 1)

Dock menar Respondent 1 att det finns mer avancerade cyberattacker som fortfarande kräver mänsklig övervakning, alltså en HL. Att ha mänsklig övervakning som balans kan vara avgörande för att säkerställa kvalitet och precision.

Transparens

Båda respondenterna är överens om att transparens är en viktig aspekt i en effektiv användning av AI-verktyg. Denna transparens är inte något som tillgängliggörs i blackboxmodellerna eftersom de inte förklarar deras beslut. Respondent 2 tycker däremot att om blackbox modeller ska tillämpas bör de vara hårt regelstyrda och ha statiska algoritmer

för att göra deras beslut förståeliga och förutsägbara. De behöver även vara oförmögna att kunna ändra sina algoritmer eftersom det kan utgöra en säkerhetsrisk ifall den förändrar sin förutsägbarhet. Respondent 1 tar som exempel de tidiga versionerna av Chat GPT, där verktyget inte specificerade vilka källor informationen hämtades från. Detta gjorde det omöjligt att verifiera källor och tilliten till svaren var låg.

"Ja men om vi kollar på första släppet av ChatGPT. Så hade ju inte den källreferenser. Det har den ju nu. Det är ju ett perfekt exempel på, nu kan jag i alla fall dubbelkolla vilka källor som den plockade data ifrån. Och ibland är det ju ren jävla dynga liksom." (Respondent 1)

Båda respondenterna belyser även vikten av att LLM:s som används på företagen ska ha en egen "container" där data hanteras internt. Detta är viktigt för att hålla data konfidentiell och att det råder en transparens på hur och var den får användas. Respondent 2 förklarar att det finns en tillit till deras LLM och detta medför en god transparens på företaget.

"Jag känner god tillit till att vi har gjort inställningen att vi skickar upp det, det stannar hos oss, det kommer inte läras i det vidare perspektivet." (Respondent 2)

Respondent 1 menar också att använda en privat container för LLM innebär att data som behandlas inte blir tillgänglig för andra. Detta säkerställer att användarnas information inte delas eller används i otillåtna syften, något som är avgörande för att skapa förtroende och transparens

Implementation

Att integrera och implementera AI i äldre system är något som beskrivits som komplicerat av båda respondenterna. Det finns flera olika orsaker till detta och Respondent 2 menar att en svårighet med detta är att det finns press från företaget att implementera AI men det är fortfarande oklart om det ger nytta eller inte. Om man ska implementera AI i äldre system finns en risk att det blir komplicerat och ineffektivt, vilket är något man vill undvika. Respondent 1 berättar om implementering och belyser vissa tekniska utmaningar de stött på. Exempelvis att en säkerhetsanalytiker kan behöva hantera flera olika plattformar med olika frågespråk, vilket gör det komplext och ineffektivt att arbeta. Ett exempel är att varje leverantör tenderar att utveckla sina egna unika system för att hantera säkerhetsdata, vilket kan skapa "tekniska silos" som försvårar samarbete och integration.

"Alla IT-säkerhetsleverantörer tenderar att ha olika smaker av vilket frågespråk man använder. Vilket gör det ganska jobbigt för en säkerhetsanalytiker att ha sex olika plattformar, sex olika språk och sex olika sätt att ställa frågor på." (Respondent 1)

Det framgår att maskininlärningsmodeller är enklare att implementera än LLM:s menar Respondent 1. Anledningen till detta är att på företaget finns det duktiga ingenjörer som jobbat med maskininlärning länge och som snabbt kunde anpassa sig till de nya arbetssätten. De är även billigare att utveckla och enklare att arbeta med enligt Respondent 1.

"Det är ganska snabbt att komma igång med machine learning. Även om Googles AI-modell är open source så är den ju inte granskningsbar. Det är liksom för mycket kod för att någon ska sitta och läsa igenom allt det här och förstå hur det fungerar. Jag tror att införandet sker på helt olika nivåer. Och det blir ganska ofta att man bara vill testa." (Respondent 1)

Att de implementerade AI-lösningarna uppfyller behov och förväntningar, menar båda respondenterna. Det är dock svårt att dra slutsatser kring vilken nytta som ges menar Respondent 2. Dess fulla potential är ännu outforskad och det är viktigt att fundera på vilka användningsområden som har nytta av AI i arbetet med cybersäkerhet.

"Vi måste backa tillbaka till innan vi kan besvara precis just den frågan. Då måste vi till att börja med ha klart för oss vad vi är ute efter. Där är vi inte riktigt än. Där utforskar vi och väldigt många andra vad det är vi behöver. Vad är det vi vill? Det måste vi ha varit klara med. Vad är våra faktiska use case?" (Respondent 2)

Att implementera och integrera AI i befintliga system som exempelvis i hotdetektion tillförser både för- och nackdelar menar Respondent 1. En människa och en AI kan bli lurade på olika sätt, exempelvis baserar en AI sina beslut på tidigare erfarenheter eller data som enkelt kan exploateras. Respondent 1 påpekar även vikten av att använda datakällor med hög trovärdighet när man tränar sin AI, eftersom opålitliga och sårbara datakällor kan "förgifta" koden som den genererar.

"Ja plus att man skulle ju tänka sig beroende på vilka datakällor man tränar sin AI på. Det kan ju vara så att man förgiftar datakällan och att alla utvecklare plötsligt börjar generera kod som innehåller samma sårbarhet. Hur hanterar man det när man väljer att outsourca de här modellerna? Alltså har man förtroende för om det är OpenAI om ni får GPT-versionen där?" (Respondent 1)

4.2.3 Miljömässiga aspekter

Hotbild

Cyberkriminella har arbetat effektivt för att skapa nya typer av hot genom AI-teknik. Ett exempel på hot som förvärrats genom AI är phishing-mail som med hjälp av LLM:s är skräddarsydda till mottagaren och skickas i en tidigare omöjlig omfattning. Respondent 1 beskriver förändringen och uppskattar att organisationen tar emot och förhindrar 70 000 sådana typer av mail per dag med hjälp av AI.

"Så jag tror att hotbilden har ändrats dramatiskt för oss. Förut fick man felstavade phishing-mail och folk tänkte hej kom och hjälp mig, vem går på det här? Till att de faktiskt ser jävligt bra ut. Och vi hade för två-tre månader sedan en robotkampanj. Det var en vd ringde upp, högre chefer, och bad dem hjälpa henne med olika typer av problem. Och det var ju inte hon naturligtvis, det var ju en AI-genererad röst i andra änden." (Respondent 1)

Utöver att försöka lura mänskliga användare och automatiska system så kan AI-tekniken som är ämnad till att stärka cybersäkerheten påverkas av cyberkriminella. Respondent 1 beskriver risker med att en LLM tränas på en så kallad förgiftad datakälla, vilket medför att den har inbyggda svagheter som kan leda till säkerhetsrisker. Respondent 2 beskriver att man kan lura en AI-modell att bortse från anomalier om man lyckas interagera med den utifrån och har dåliga intentioner.

"En människa kan ju bli lurad. En AI kan ju också bli lurad men kanske inte på samma sätt." (Respondent 2)

Konsensus bland respondenterna är att dessa nya typer av AI-hot kommer att behöva bekämpas med AI. Inom cybersäkerhetsfären har det konstant varit en dynamik av

konkurrens mellan attackerade och försvarare gällande möjligheter med ny teknik. AI har fått omfattande effekter genom dess förmåga att analysera och skapa stora mängder data med hög grad av precision.

"Det kommer inte finnas något alternativ än att bekämpa AI med hjälp av AI. Det är min absoluta övertygelse." (Respondent 2)

"Jag tror att om man ska kolla på teknikadaption så är det inte de som försvarar som adapterar sig snabbast utan de som attackerar. Så de cyberkriminella stora organen som Black Act och så vidare eller de större APT-grupper som har flyttat till Cozy Bear och så vidare i Ryssland. De var ju extremt snabba på att använda AI för att till exempel skapa phishing-mail. Det var ju ganska tafatt i början för då kunde inte AI göra det här så jävla bra. Men nu kan ju en AI skriva phishing-mail som är helt suveräna. På en skala som en människa aldrig skulle kunna gå med." (Respondent 1).

Marknad

Samtliga respondenter menar att licensiering är det naturliga valet för AI-modeller idag eftersom utvecklingen av en LLM är kostsam och riskfylld.

"Just nu tror jag att den ekonomiska förklaringsmodellen är så pass enkel att det blir så mycket enklare eller billigare och bättre att licensiera. Då får man alltid välja att licensiera" (Respondent 1)

Konkurrensen på marknaden bidrar till att företagen implementerar AI i syfte att inte halka efter i utvecklingen. Enligt respondent 2 så finns det dock ett stort utbud av AI implementationer som inte uppfyller faktiska behov som kunder har efterfrågat. Det beskrivs som ett tillägg på produkter i marknadsföringssyfte.

"AI har haft en tendens att vara lite grann, det här gnistrande, blänkande strösslet uppe på snart varenda produkt. Alla ska ha någon form av AI-implementation och visa upp att vi har AI-implementation på vår produkt. Och alldeles för många gånger så har AI-implementationen egentligen bara byggt på en LLM-chatt nere i högra hörnet av fönstret." (Respondent 2)

Samtliga respondenter menar att utbudet inte korrelerar med efterfrågan på AI-lösningar och företagen själva behöver tid att identifiera sina behov av AI inom cybersäkerhet.

"Men jag skulle säga rent övergripande om det här med om nuvarande AI-modeller uppfyller krav eller inte. Det är nog mest bara för att vi inte vet kraven riktigt. Det är inte vi som driver den här teknikutvecklingen, det är någon annan. Och då är det svårt förklara riktigt vad man vill med AI och sen faktiskt få gehör för det. Man vill att någon annan listar ut mina behov åt mig. Och sen gärna kommer till mig och försöker följa dem till mig." (Respondent 1)

Respondent 1 gav även exempel på när en specifik upphandling av AI-driven teknik uppfyller ett faktiskt behov och gav företaget en fördel inom sin bransch. Respondent 1 beskriver hur AI-teknik har blivit en del i kapplöpning om att vara den bästa inom sin bransch.

"De ligger långt efter. Så det är klart att man hela tiden försöker bräcka varandra. Och AI är ett jättebra exempel på det." (Respondent 1)

Respondenterna har inte upplevt påtryckningar från kunderna att implementera AI-lösningar i deras verksamhet. Respondent 1 har snarare upplevt skepticism mot nyttjande av kunddata i AI-modeller.

"Ja, vi har haft kunder som säger att data som ni får av oss får ni aldrig under några omständigheter hantera i en AI-modell. Sådana krav finns." (Respondent 1)

Regelverk
Implementeringen av AI inom cybersäkerhet styrs av gällande regelverk och standarder för samtliga deltagande i studien. Respondent 1 har gjort ett aktivt val att utforma sina riktlinjer efter NIST för att säkerställa riktlinjer som främjar cybersäkerheten inom företaget. De har valt att följa NIST AI Risk Management Framework och använder AI Act som en mer generell vägledning. Detta val grundas i att EU:s AI Act är jämförelsevist simpelt och Respondent 1 menar att de ligger något efter i förhållande till utvecklingen av AI.

"Problemet med EU är att man inte har ett speciellt vettigt cybersäkerhetsorgan. AI Act *är relativt ny, men den är inte tekniskt detaljerad nog för att möta de verkliga utmaningarna. Vi följer istället NIST-ramverket eftersom det är mer krispigt och tekniskt användbart."* (Respondent 1)

Respondent 2 har valt ett liknande tillvägagångssätt för att implementera AI Act. Genom att proaktivt implementera relevanta standarder och ramverk i sina interna processer så säkerställer de efterlevnad automatiskt istället för att ständigt försöka tolka lagtexten.

"Vi jobbar inifrån och ut – om vi gör rätt internt följer vi automatiskt de flesta regelverk. Det skall vara lätt att göra rätt" (Respondent 2)

Ytterligare relevanta regelverk är Cyber Resilience Act som innebär att företag som levererar lösningar med källsårbarheter kan bötfällas uppemot 15 miljoner Euro. Detta innebär förhoppningsvis att det som tidigare prioriterats, att vara först på marknaden delvis kommer skiftas till att säkerställa en högre säkerhetsnivå på dessa produkter, då vinsten att vara först inte överväger potentiella böter, menar Respondent 1.

"Utan att man kanske får ta ett steg tillbaka och fundera på att följer vi verkligen kraven? För vi vet att det kan bli jävligt kostsamt om man inte gör det. Så det är väl en intressant faktor kan jag tycka" (Respondent 1)

5. Diskussion

I följande kapitel analyseras och diskuteras resultatet av empirin gentemot teorin och tidigare forskning.

Rahman et al. (2022) skriver om hur AI kan minska mänskliga misstag genom automation men att behovet av mänsklig ökning kommer att kvarstå. Detta överensstämmer med resultatet där AI kan används som skydd mot phishing-mail men behovet av mänsklig övervakning för mer avancerade hot kvarstår. AI används även för automatisk övervakning av loggar för att upptäcka anomalier men även här betonas behovet av en slutgiltig mänsklig bedömning, både tidigare forskning och empirin efterfrågar en HL för att säkerställa kvalité och säkerhet.

Att ökad automation kan leda till ökad arbetsbörda är ett fenomen Michael et al. (2023) benämner "Automation paradox". Detta fenomen återfinns till viss del i resultatet där automatiseringen genom AI effektiviserat arbetsuppgifter men även ställt nya krav på de anställda gällande förståelse och de beslutsunderlag som den nya tekniken medför. Även en oro för att tekniken kan konkurrera med en anställds spetskompetens uttrycktes, men då behovet att verifiera de beslut som en AI tar finns behovet av en ny arbetsuppgift som bygger på spetskompetensen.

Rahman et al. (2022) understryker vikten av kontinuerlig utbildning för att hantera de sociala och tekniska utmaningarna som uppstår vid en AI-implementation. Resultatet visar att bristande utbildning negativt påverkar arbetsbördan, eftersom anställda kan undvika att använda tekniken och istället välja alternativ de upplever som enklare men medför en säkerhetsrisk.

Michael et al. (2024) och Capuano et al. (2022) menar att valet av AI-tekniken är viktig för en lyckad implementation och de förespråkar användningen av XAI. Resultatet visar behovet av att en AI har förmågan att förklara sina beslut vilket ökar både förståelsen och tilltron till tekniken men även säkerheten då möjligheten att kontrollera dessa beslut ökar.

Zhang et al. (2022b) och Michael et al. (2023) skriver båda om riskerna med AI:s dubbelsidiga natur. De menar att AI används både för att förbättra cybersäkerheten men används även av aktörer med illvilliga intentioner för att skapa nya typer av hot. Resultatet visar att samtliga respondenter upplevt dessa nya typer av hot och identifierat behovet av att bekämpa dagens cyberattacker med ett säkerhetssystem som implementerar AI.

Både Zhang et al. (2022b) och Capuano et al. (2022) är övertygade om att AI inom cybersäkerhet är en marknad som kommer att öka markant inom kort. Den ökade tekniska sammankopplingen i samhället och komplexiteten har ökat behovet av säkerhetslösningar som AI kan bidra med. Kombinationen av AI och cybersäkerhet ger även utrymme till innovation och ett förbättrat skydd mot avancerade cyberattacker. Konkurrens nämns även i båda artiklarna och de tror att AI kommer bidra med flera konkurrensfördelar. Zhang et al. (2022) menar att konkurrensen driver cybersäkerhetsföretag att utveckla sina produkter så att de är i framkant med prestanda och skalbarhet. Capuano et al. (2022) diskuterar hur XAI och förklarliga AI-beslut kommer att vinna kunders och användares förtroende för deras lösningar inom cybersäkerhet. Resultatet visar att samtliga respondenter är medvetna om "tävlingen" på marknaden. Många vill vara ledande och visa sina nya lösningar, men respondenterna är osäkra på om implementeringen alltid effektiviserar lösningarna. Många gånger får de reflektera om det bidrar och uppfyller ett faktiskt behov för dem.

Capuano et al. (2022) och Zhang et al. (2022a) undersöker vilken effekt som lagstiftningen har på AI implementation. GDPR har anpassats för AI och det är ett krav att automatiserade beslut för individer kan förklara grunden för sina beslut. De anser att någon form av XAI kommer att krävas för automatiserade system som tar beslut på individnivå. Utöver GDPR så visar resultatet behovet av att förhålla sig till AI Act för företag i EU, denna lagstiftning skärper kraven runt dataskydd och transparens ytterligare. Detta medför att företag i Sverige som implementerat en AI-lösning i sitt cybersäkerhetssystem har höga krav på transparens och dataskydd. Detta påverkar inte bara teknikvalet utan företagets arbetsflöden och behov av utbildning för att säkerställa att regelverken efterföljs. Resultatet visar att företagen ibland väljer att ta höjd för dessa regler genom att utforma sin cybersäkerhetspolicy efter NIST som är striktare än tidigare nämnda regelverk. Detta kan ses som en konkurrensfördel då kunder upplever en trygghet när gällande regelverk efterföljs.

Michael et al. (2024) presenterar flera identifierade kunskapsgap i deras studie om hur AI-teknik bör implementeras med ett socioteknisk ramverk. Denna studie adresserar två av dessa kunskapsgap. Det första gapet belyser hur mänskliga faktorer är underrepresenterade i cybersäkerhetsforskning. Studiens empiri, med stöd från tidigare forskning, betonar att riktade åtgärder för användare av systemen, såsom utbildning och ökad transparens, kan stärka cybersäkerheten genom att skapa förtroende och engagemang. Det andra gapet handlar om bristande regelverk och policy för att bemöta cybersäkerhetsproblem. Studien understryker vikten av att integrera lagstadgade regelverk som EU:s AI Act och GDPR i cybersäkerhetspolicys samt betonar möjligheten att komplettera med riktlinjer från NIST för att ytterligare stärka organisationens resiliens mot cyberattacker.

6. Slutsats

Denna studie syftar till att undersöka hur stora svenska företag har implementerat och använder AI för att förbättra deras cybersäkerhet. Studiens syfte är även att ge en överblick över de potentiella fördelarna som ett socioteknisk förhållningssätt kan föra med sig. Studiens två intervjuer har möjliggjort en unik inblick i hur AI-drivna lösningar används och hur detta påverkat samspelet mellan tekniska och sociala aspekter inom företagen. Syftet presenteras och besvaras genom en tvådelad frågeställning.

1. Hur implementerar stora svenska företag AI för att förbättra cybersäkerheten?

Studien visar att stora svenska företag använder AI främst för automatiserad hotdetektion, anomaliövervakning och snabbare respons vid cybersäkerhetsincidenter. AI effektiviserar säkerhetsarbetet genom att ta över repetitiva uppgifter som loggövervakning. Samtidigt kräver AI-systemen mänsklig övervakning för att säkerställa beslutens tillförlitlighet vilket medför nya arbetsroller och uppgifter.

2. Kan implementationen av AI inom cybersäkerhet förbättras med hjälp av ett socioteknisk perspektiv för stora svenska företag?

Denna studie bygger vidare på tidigare forskning om möjligheterna med att utforma ett holistiskt system genom med den sociotekniska teorin som ledstång. Vilket genomförs genom att beakta balansen mellan sociala, tekniska och miljömässiga faktorer. Sociala faktorer som förståelse och förtroende för tekniken samt hänsyn för förändrad arbetsbörda och arbetsuppgifter har enligt tidigare forskning och denna studie visats sig vara väsentligt för en lyckad implementering av tekniken.

Viktiga tekniska faktorer vid implementeringen har visat sig vara hur väl AI kan integreras mot tidigare system, samt AI:s förmåga att förklara sina beslut och på så vis vara transparent. XAI är även nödvändigt för ett HL upplägg vilket kan öka systemets precision och säkerhet.

Miljömässiga aspekter som krävs är en gedigen uppfattning av den existerande hotbilden så att tekniken driftsätts där den ger maximalt med skydd. Även att utforma policys som stöds av gällande regelverk såsom AI Act och GDPR kommer att krävas. Dessa policys kan utvecklas ytterligare genom NIST direktiv för en lyckad implementation.

Resultatet visar på konsensus kring att framtidens system kräver AI-teknik för att försvara sig mot de komplexa attacker och hot som AI även möjliggör. Det råder även enighet om att tekniken idag inte är sofistikerad nog att självständigt ansvara för säkerheten, det behövs en människa som kontrollerar och godkänner besluten som systemet fattar.

Avslutningsvis betonar denna studie vikten av en holistisk syn vid AI-implementation för en förbättrad cybersäkerhet. Genom att integrera sociotekniska principer stärker företag sin säkerhet och har bra förutsättningar att möta kommande krav och möjligheter i Sveriges resa mot att bli en global ledare inom AI-teknik.

6.1 Förslag på framtida forskning

Den aktuella studien är avgränsad till svenska företags implementering av AI i cybersäkerheten. En intressant riktning för vidare forskning skulle kunna vara att utföra en komparativ studie med andra länder. Det skulle ge en intressant uppfattning om hur

organisatoriska skillnader som innovationskultur och regleringar kan skilja sig från situationen i Sverige till andra länder i världen. Dessutom ger detta en möjlighet att undersöka vilka globala trender som finns inom AI-användningen världen över.

Båda respondenterna anser att AI inom cybersäkerhet är ett fenomen som är under en utvecklingsfas. Det är fortfarande tidigt i fasen och resultaten som denna studie har genererat behöver inte nödvändigtvis vara aktuella inom en snar framtid. En av respondenterna förutspådde att om denna studie skulle genomföras om ett halvår eller ett år, så skulle den se väldigt annorlunda ut. Här finns det en möjlighet för vidare forskning att jämföra framtida studier med denna och exempelvis se hur utvecklingen har sett ut eller om de utmaningar man står inför har förändrats.

I takt med att AI utvecklas i förebyggande syfte för cybersäkerhetsarbetet växer det kriminella utnyttjandet av den. En mer tekniskt inriktad studie är av vikt för att förstå hur AI kan användas så effektivt som möjligt mot de nya, avancerade och sofistikerade hoten. Undersökningen bör särskilt fokusera på hur AI-baserade lösningar kan förstärka organisationers motståndskraft genom att upptäcka och förutse hot med hög precision. Samtidigt är det av central betydelse att minimera de risker som AI:s dubbelsidiga natur medför, där samma teknik som skyddar mot hot också kan användas av angripare för att skapa nya och mer sofistikerade attacker.

7. Källförteckning

Bryman, A. (2018). *Samhällsvetenskapliga metoder* (3 uppl.). Liber.

Capuano, N., Fenza, G., Loia, V., & Stanzione, C. (2022). Explainable artificial intelligence in cybersecurity: A survey. *IEEE Access*, 10, 93575–93593. https://doi.org/10.1109/ACCESS.2022.3204171

Dehghantanha, A., Yazdinejad, A., & Parizi, R. M. (2024). Autonomous cybersecurity: Evolving challenges, emerging opportunities, and future research trajectories. *Proceedings of the Workshop on Autonomous Cybersecurity (AutonomousCyber '24)*, 10 pages. https://doi.org/10.1145/3689933.3690832

Ehsan, U., Passi, S., Liao, Q. V., Muller, M., & Riedl, M. O. (2023). Charting the sociotechnical gap in explainable AI: A framework to analyze and address the gap in research and practice. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), Article 35. https://doi.org/10.1145/3579467

Europaparlamentet. (29 november 2021). *Cybersäkerhet: vikten av att minska kostnaderna för cyberattacker.* https://www.europarl.europa.eu/topics/sv/article/20211008STO14521/cybersakerhet-vikten-av-att-minska-kostnaderna-for-cyberattacker

European Union Agency for Cybersecurity. (2024). ENISA Threat Landscape 2024: July 2023 to June 2024. https://www.enisa.europa.eu

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy. *IEEE Access*, 11, 80218–80245. https://doi.org/10.1109/ACCESS.2023.3300381

Jada, I., & Mayayise, T. O. (2024). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data and Information Management*, 8(2). https://doi.org/10.1016/j.dim.2023.100063

Kolevski, D., & Michael, K. (2015). Cloud computing data breaches: A socio-technical review of literature. In *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)* (pp. 1486–1495). IEEE. https://doi.org/10.1109/ICGCIoT.2015.7380702

Lyytinen, K., & Newman, M. (2008). Explaining information systems change: A punctuated socio-technical change model. *European Journal of Information Systems*, 17(6), 589–593. https://doi.org/10.1057/ejis.2008.50

Michael, K., Abbas, R., & Roussos, G. (2023). AI in cybersecurity: The paradox. *IEEE Transactions on Technology and Society*, 4(2), 104–115. https://doi.org/10.1109/TTS.2023.3280109

Michael, K., Vogel, K. M., Pitt, J., & Zafeirakopoulos, M. (2024). Artificial intelligence in cybersecurity: A socio-technical framing. *IEEE Transactions on Technology and Society*. https://doi.org/10.1109/TTS.2024.3460740

Mumford, E. (2000). Socio-technical design: An unfulfilled promise or a future opportunity? In R. Baskerville, J. Stage, & J. I. DeGross (Eds.), *Organizational and social perspectives on information technology* (pp. 33–46). Springer. https://doi.org/10.1007/978-0-387-35505-4_3

Mumford, E., & Weir, M. (1979). *Computer systems in work design: The ETHICS method*. Associated Business Press.

Myndigheten för samhällsskydd och beredskap. (2020). Cybersäkerhet i Sverige 2020 – Hot, metoder, brister och beroenden. https://www.msb.se

National Institute of standards and technology. (2025). *Cybresecurity*. https://csrc.nist.gov/glossary/term/cybersecurity

Nationellt cybersäkerhetscenter (NCSC). (2024). Säkra verksamheten vid en cyberattack. Hämtad från https://www.ncsc.se/sv/om-centret/

Rahman, T., Rohan, R., Pal, D., & Kanthamanon, P. (2021). Human factors in cybersecurity: A scoping review. In *The 12th International Conference on Advances in Information Technology (IAIT2021)* (pp. 1–11). ACM. https://doi.org/10.1145/3468784.3468789

Regeringen. (2024, september 18). Ny utredning för trygg och tillförlitlig användning av AI i Sverige. https://www.regeringen.se/pressmeddelanden/2024/09/ny-utredning-for-trygg-och-tillforlitlig-anvandning-av-ai-i-sverige/

Rojko, A. (2017). Industry 4.0 concept: Background and overview. *International Journal of Interactive Mobile Technologies (iJIM)*, 11(5), 77–90. https://doi.org/10.3991/ijim.v11i5.7072

Rudolph, J. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning & Teaching*, 6(1), 364–389. https://doi.org/10.37074/jalt.2023.6.1.23

Schoenherr, J. R., Abbas, R., Michael, K., Rivas, P., & Anderson, T. D. (2023). Designing AI using a human-centered approach: Explainability and accuracy toward trustworthiness. *IEEE Transactions on Technology and Society, 4*(1), 9–23. https://doi.org/10.1109/TTS.2023.3257627

Statistiska centralbyrån (SCB). (2024, januari 12). Användningen av AI ökar mest bland små företag. https://www.scb.se/hitta-statistik/artiklar/2024/anvandningen-av-ai-okar-mest-bland-sma-foretag/

Taddeo, M., Jones, P., Abbas, R., Vogel, K., & Michael, K. (2023). Socio-technical ecosystem considerations: An emergent research agenda for AI in cybersecurity. *IEEE Transactions on Technology and Society*, 4(2), 112–118. https://doi.org/10.1109/TTS.2023.3278908

Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting. *Human Relations*, 4(1), 3–38. https://doi.org/10.1177/001872675100400101

Information Systems Theories (2024). Socio-technical theory. https://is.theorizeit.org/wiki/Socio-technical_theory

Zeadally, S., Adi, E., Baig, Z., & Khan, I. A. (2020). Harnessing artificial intelligence capabilities to improve cybersecurity. *IEEE Access*, 8, 23817–23837. https://doi.org/10.1109/ACCESS.2020.2968045

Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022a). Explainable artificial intelligence applications in cyber security: State-of-the-art in research. *IEEE Access*, 10, 93104–93122. https://doi.org/10.1109/ACCESS.2022.3204051

Zhang, Z., Ning, H., Shi, F., Farha, F., Xu, Y., Xu, J., Zhang, F., & Choo, K.-K. R. (2022b). Artificial intelligence in cyber security: Research advances, challenges, and opportunities. *Artificial Intelligence Review*, 55(2), 1029–1053. https://doi.org/10.1007/s10462-021-09976-0

8. Bilagor

Bilaga 1 - Intervjuguide

Öppnande frågor:

- Kan ni beskriva hur långt ni har kommit i implementeringen av AI för cybersäkerhet och vilka steg som varit de mest avgörande i denna process?

- På vilket sätt har implementeringen av AI för cybersäkerhet påverkat organisationen positivt?

Sociala

Fråga 1: Hur har anställda påverkats av implementationen av AI, exempelvis i termer av arbetsbörda eller förtroende för tekniken?

Fråga 2: Hur säkerställer ni att anställda har den kunskap som krävs för att använda AI effektivt i sitt arbete med cybersäkerhet?

Fråga 3: Hur upplever ni att acceptans och förtroende för AI som ett verktyg inom cybersäkerhet tas emot?

Fråga 4: Finns det specifika hinder, motstånd eller skepticism som ni behövt hantera, och hur har ni arbetat för att stärka tilliten till dessa system?

Tekniska

Fråga 5: Har ni några reflektioner gällande implementeringen av AI?

Fråga 5.5: Har ni utvecklat en egen AI-modell eller använder ni idag en tredjepartslösning?

Fråga 6: Hur avgör ni vilka specifika områden som lämpar sig bäst för automatisering, och vilka uppgifter som kräver mänsklig övervakning?

Fråga 7: Hur har ni resonerat kring användandet av en förklarande AI-modell (XAI) som är transparent kontra ett stängt system så kallat "black box" och vilka möjligheter och risker ser ni med dessa typer av system?

Frågor om XAI:
- Hur säkerställer ni att en förklarande AI-modell är tillräckligt transparent utan att kompromissa med prestandan?
- Kan förklarande AI-modeller skapa säkerhetsrisker, exempelvis genom att avslöja för mycket om hur systemet fungerar för potentiella angripare?

Frågor om BB:
- Hur hanterar ni situationen där ett beslut från ett black box-system ifrågasätts och måste förklaras?
- Hur bedömer ni och minimerar risken för bias i black box-system, särskilt i känsliga områden som cybersäkerhet?

Fråga 8: Vilka tekniska möjligheter och begränsningar upplever ni att AI-tekniken medför?
Miljömässiga

Fråga 9: Hur har utomstående faktorer påverkat organisationens cybersäkerhetsstrategi med användandet av AI?
Gällande kunders förväntningar och konkurrensen på marknaden?

Fråga 10: Vilka typer av externa hot upplever ni som mest relevanta, och hur hjälper AI till att hantera dessa?

Fråga 11: Vilket regelverk använder ni som riktlinjer för era egna cybersäkerhetspolicys?

- EU:s nya AI-förordning?
- Mer heltäckande som till exempel NIS II?

Avslutande fråga: Om ni skulle få ändra något i hur ni använder AI för cybersäkerhet idag, vad skulle det vara och varför?