Tekoälyn rooli kyberturvallisuudessa

Mikaela Hirvonen

Haaga-Helia ammattikorkeakoulu

Tradenomi

Tutkimusraportti

2025

Tiivistelmä

Tekijä(t)
Mikaela Hirvonen

Tutkinto
Tradenomi, tietojenkäsittelyn koulutusohjelma

Raportin/Opinnäytetyön nimi
Tekoälyn rooli kyberturvallisuudessa

Sivu- ja liitesivumäärä
27

Opinnäytetyön tarkoituksena on tutkia tekoälyn hyötyjä kyberturvallisuudessa. Työn tutkimus toteutettiin kolmen tutkimuskysymyksen avulla, joista ensimmäinen tutkimuskysymys oli, että miten tekoälyä voi hyödyntää kyberturvallisuudessa? Toisena tutkimuskysymyksenä oli, että mitä riskejä tai haasteita tekoäly tuo kyberturvallisuuteen? Viimeisenä kysymyksenä oli, että mitä eettisiä puolia pitää ottaa huomioon tekoälyn käytössä kyberturvallisuudessa?

Opinnäytetyön teoriaosuus on jaettu kahteen osaan; tekoäly ja kyberturvallisuus. Tekoälyn osiossa käydään läpi sen peruskäsitteet ja historia. Sen lisäksi käydään läpi tekoälyn eri osa-alueita, kuten koneoppiminen, syväoppiminen, generatiivinen tekoäly ja luonnollisten kielten käsittely. Tämän jälkeen esitetään kyberturvallisuuden peruskäsitteet, sekä sen lisäksi keskitytään verkkoturvallisuuteen ja tietoturvallisuuteen. Näistä kummastakin aiheesta käydään läpi niiden peruskäsitteet sekä niihin liittyvät uhat.

Tutkimustuloksista tuli ilmi, että tekoälyn avulla voidaan automatisoida käytäntöjä, kuten uhan havainnointi ja voidaan analysoida isoja määriä tietoja. Tekoälyn riskejä on tekoälyjärjestelmän tietoturva sekä tekoälyn avulla ehostetut kyberhyökkäykset. Tekoälyn eettiset puolet olivat tekoälyn läpinäkyvyys, järjestelmän ennakkoluulot ja tietojen suojaus.

Työn tutkimus toteutettiin laadullisella kirjallisuuskatsauksella. Työssä lähteinä käytettiin Google Scholarista löytyneitä artikkeleita ja googlesta löytyneitä raportteja. Sekä HH finnan tietokannasta saatuja E-kirjoja ja tieteellisiä artikkeleita.

Asiasanat
Kyberturvallisuus, tekoäly

Sisällys

1 Johdanto

Maailma kulkee vauhdilla kohti yhä digitaalista tulevaisuutta, jossa teknologia kietoutuu yhä tiiviimmin osaksi ihmisten arkea ja työympäristöjä. Esimerkiksi digitaalisten lompakoiden ja henkilökorttien yleistyminen tuo mukanaan merkittäviä etuja, kuten helpompaa asiointia, mutta samalla myös uudenlaisia tietoturvariskejä. Digitalisoitumisen lisäksi tekoäly kehittyy jatkuvasti ja tuo mukanaan uudenlaisia riskejä niin arkeen kuin yritysmaailmaan. Kehittyneet tekoälypohjaiset hyökkäykset tekevät kyberuhista entistä monimutkaisempia ja vaikeammin havaittavia. (Zeadally, Adi, Baig & Khan 2020, luku 1).

Tämä opinnäytetyö on tutkimuksellinen kirjallisuuskatsaus, jonka tarkoituksena on tarkastella kyberturvallisuuden maailmaa ja selvittää, miten nykypäivän tekoäly pystyy edesauttamaan kyberturvallisuutta.

Molemmat aiheet tekoäly ja kybertuvallisuus ovat hyvin laajoja alueita, joten työssäni olen rajannut aiheita. Tekoälyn rajaus tehtiin sen osa-alueisiin, jotka ovat konkreettisia kyberturvallisuuden kannalta. Siinä keskitytään koneoppiin, syväoppiin, generatiiviseen tekoälyyn sekä luonnollisen kielen käsittelyyn. Kyberturvallisuudessa taas keskitytään erityisesti tietoturvaan ja verkkoturvallisuuteen, sekä niiden uhkiin. Rajauksen tavoitteena on varmistaa tutkimuksen selkeä keskittyminen olennaisiin aiheisiin sekä välttää liiallista teknistä syvyyttä.

Tutkimuksen tavoitteena on selvittää, millä tavoin tekoälyä voi hyödyntää kybertuvallisuudessa ja mitä tekoälyn eri osa-alueita käytetään hyödyksi. Kirjallisuuskatsauksessa selvitetään, mitä haasteita ja riskejä tekoäly tuo kyberturvallisuuteen ja millä tavoin riskejä esiintyi. Kirjallisuuskatsauksen tulososiossa pohditaan tekoälyn ja kyberturvallisuuden eettisiä puolia, niin hyvässä kuin huonossa.

Valitsin tämän aiheeksi, koska kyberturvallisuus ja siinä erityisesti, tieto- ja verkkoturvallisuus, ovat isoimpia kiinnostuksen kohteitani. Kirjallisuuskatsauksen yhtenä aiheena tekoäly sisällytettiin sen ajankohtaisuuden ja nopean kehittymisen vuoksi. Lisäksi halusin oppia ymmärtämään tekoälyä enemmän.

Peittomatriisi

Taulukko 1 Peittomatriisi

Aihe / kysymys	**Tietoperusta**	**Tutkimus**	**Tutkimuksen tulokset**	**Pohdintaa**
Mitä tekoäly on?	**2,2.1, 2.2, 2.3**			
Mitä kyberturvallisuus on?	**3,3.1, 3.2, 3.3, 3.3.1, 3.3.2, 3.3.3, 3.3.4**			
Tutkimuksen taustaa.		**4, 4.1, 4.2, 4.3, 4.4**		
Tekoälyn hyöty.			**5.1**	**6.4**
Tekoälyn riskejä.			**5.2**	**6.4**
Tekoälyn eettiset puolet			**5.3**	**6.4**

2 Tekoäly

Tekoäly on teknologiaa, jolla pyritään luomaan järjestelmiä ja laitteita, jotka matkivat ihmisten älykkyyttä ja ongelmanratkaisukykyä. Tekoälypohjaiset laitteet pystyvät havaitsemaan ja tunnistamaan ympäröiviä asioita kuten esineitä, liikkeitä ja ääniä sekä ymmärtämään ja tuottamaan ihmisten kieltä. (Stryker & Kavlakoglu 9.8.2024.)

Tekoäly on hyvin laaja osa-alue ja sen voi jakaa kahteen eri päätyyppiin: vahvaan tekoälyyn tai heikkoon tekoälyyn. Erottelukriteerinä toimii teknologian älykkyys ja toiminnot. (Geeks for geeks 28.6.2024).

Heikko tekoäly ei yleensä omaa ymmärrystä ja tietosuutta, vaan se keskittyy suorittamaan yhden tehtävän tai pienen paketin erilaisia tehtäviä. Sen sijaan vahva tekoäly ei ole rajoitettu yhteen tehtävään, vaan se omaa yleisen älykkyyden, joka kykenee oppimaan, ymmärtämään ja soveltamaan tietoa. Periaatteessa pyritään tekemään järjestelmä, joka pystyy suorittamaan minkä tahansa älyllisen tehtävän, samalla tavalla kuin ihminen pystyisi. (Geeks for geeks 28.6.2024.)

2.1 Historiaa

”Voiko koneet ajatella?” ja ”voiko kone käyttäytyä älykkäästi?” olivat kysymyksiä, jotka heräsivät jo vuonna 1950. Brittiläinen matemaatikko Alan Turing oli henkilö, joka pohti tätä kysymystä ja sen pohjalta hän julkaisi artikkelin” Computing Machinery and Intelligence”, jossa hän esitteli idean koneiden älykkyyden mittaamisesta. Siitä syntyi perusta, kuinka koneitten älykkyyttä voitaisiin testata ”imitaatiopelin” avulla, nykyään tunnettu Turingin testinä. (Stryker & Kavlakoglu 9.8.2024; Mucci 21.10.2024.)

Turingin testi toteutettiin siten, että testiin osallistuva ihminen on testiympäristössä, jossa osallistuu kirjalliseen keskusteluun toisen ihmisen sekä koneen kanssa. Tehtävänä on keskustelun perusteella tunnistaa kumman tuottaman teksti on ihmisen ja kumpi koneen. Tavoitteena on katsoa tunnistaako testiin osallistuva ihminen koneen. Mikäli konetta ei voida erottaa ihmisestä, tarkoittaa se sitä, että kone on läpäissyt testin. (Numminen 19.10.2023.) Kuvassa 1 kuvaillaan miten Turingin testi toteutetaan.

Kuva 1 Turingin testi (mukailen Numminen 19.10.2023)

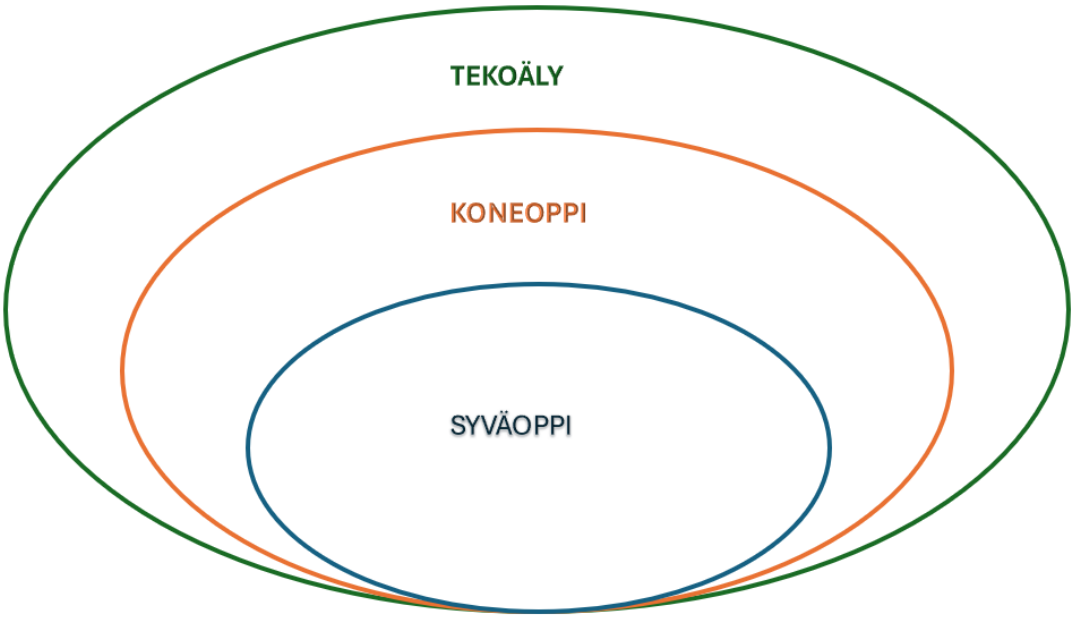
2.2 Koneoppi & syväoppiminen

Koneoppiminen on yksi tekoälyn keskeisistä osa-alueista. Koneoppi on järjestelmä, joka voi oppia ja mukautua saamansa datan perusteella. Perinteisesti ohjelmistot toimivat ennalta määriteltyjen sääntöjen mukaisesti, mutta koneoppimisjärjestelmät kykenevät analysoimaan tietoa, tunnistaa kuviota ja optimoimaan suorituskykyään itsenäisesti. (Kanade 2022; Kufel ym. 2023 luku 4.)

Koneoppimisen mekanismit voidaan jakaa kahteen tyyliin, ohjattu oppiminen (supervised learning) ja ohjaamaton oppiminen (unsupervised learning). Näiden menetelmien perimmäinen ero on datan käsittelytapa sekä se, millä tavoin algoritmi oppii tunnistamaan ja luokittelemaan tietoa. (Zeadally, ym. 2020, luku 2.)

Ohjatussa oppimisessa koulutusdata on jo etukäteen luokiteltu antamalla esimerkkejä, mikä data on haitallista ja mikä on sallittua. Malli oppii näiden etukäteen annettujen vastauksien pohjalta ja pystyy sen jälkeen itsenäisesti luokittelemaan uutta dataa vastaavien ominaisuuksien perusteella. Ohjaamattomassa oppimisessa taas koulutusdataa ei ole merkitty etukäteen, eikä mallille anneta valmiita vastauksia, esimerkiksi mikä on sallittua tai mikä on kiellettyä. Sen sijaan algoritmi saa erilaisia datapaloja, joiden välillä se etsii yhtenäisyyksiä ja muodostaa niiden perusteella ryhmiä. Näin malli löytää itse datan sisäisiä rakenteita ilman ihmisen antamia ohjeita. (Zeadally ym. 2020, luku 2.)

Koneopin lisäksi tekoälyn joukkoon kuuluu myös syväoppiminen. Syväoppiminen on laite tai kone, joka käsittelee ja analysoi tietoa loogisesti samalla tavalla, miten ihminen käsittelisi. *Se* hyödyntää kerrostettua algoritmirakennetta, jota kutsutaan neuraaliverkoksi. Nämä neuraaliverkot ovat malli tai menetelmä, jossa mahdollistetaan tekoälyjärjestelmien kyvyn matkia ihmisten päättelykykyä. (IBM 6.10.2021; Holdsworth & Scapicchio 17.6.2024.) Kuvassa 2 kuvataan tekoälyn rakennetta osajoukoittain.



Kuva 2 Tekoälyn osajoukot (mukaillen Deep instinct s.a.)

2.3 Generatiivinen tekoäly & luonnollisen kielen käsittely

Nykypäivänä tunnettu ChatGPT on generatiivinen tekoäly. Se on tekoälyteknologiaa, joka pystyy luomaan alkuperäistä tekstiä, kuvia, videoita tai muuta sisältöä. Eli pystyy kommunikoimaan käyttäjän kanssa saman lailla, miten ihminen voisi. Generatiivinen tekoäly käyttää hyväkseen aikaisemmassa kappaleessa mainittua koneoppimista sekä syväoppimista. (Stryker & Scapicchio 22.3.2024.)

Tekoäly kuten ChatGPT, kykenee ymmärtämään kieltä kirjallisena tai puheena, hyödyntämällä tekoälyn luonnollisen kielen käsittely osa-aluetta (Natural Language Processing, NLP). Se toimii käyttäen koneoppia, jolla se analysoi ihmisen kieltä ja opettelee sen imitoimista. (Copeland 16.5.2025; Cloudflare s.a.)

Luonnollisen kielen käsittelystä kehittyneempi versio on suuret kielimallit (Large Language Models, LLM). Suurin ero näiden välillä on siinä, että NLP keskittyy tarkkuuteen ja noudattaa tiettyjä

sääntöjä sekä käytäntöjä, kun taas LLM hyödyntää koneoppia ja syväoppia, käsitelläkseen valtavia datamääriä. (Cloudflare s.a.; Elastic Platform Team 2024.)

3 Kyberturvallisuus

Kyberturvallisuus koostuu erilaisista teknologioista, prosesseista ja toimintatavoista, joiden tavoitteena on suojata laitteita sekä käyttöjärjestelmiä hyökkäyksiltä ja luvattomalta pääsyltä. Kyberturvallisuus ei ole vain järjestelmäpohjaista turvallisuutta, vaan se on myös hyvin riippuvainen ihmisistä. Ihmisten huolimattomuus, tietoisuuden puute ja virheet luovat riskejä. (Fortinet s.a.) Kyberturvallisuuden voi jakaa useisiin eri osa-alueisiin, joista jokainen keskittyy teknologian eri osiin. Näitä ovat esimerkiksi, verkkoturvallisuus, sovellusturvallisuus, tietoturvallisuus, pilviturvallisuus ja laiteturvallisuus.

Kyberturvallisuuden ja tietosuojan laiminlyönti voi johtaa vakaviin seurauksiin, kuten tietovuotoihin. Tietovuoto voi syntyä kyberhyökkäyksen seurauksena, kun haitalliset toimijat onnistuvat murtautumaan suojattuihin järjestelmiin ja saamaan luvattoman pääsyn arkaluontoisiin tietoihin. Tällaiset tiedot voivat olla esimerkiksi pankkitilitiedot, henkilötunnukset tai yrityksien liikesalaisuudet. Arkaluontoisten tietojen päätyminen vääriin käsiin voi aiheuttaa vakavia seurauksia, kuten identiteettivarkaus ja mainehaittoja sekä mahdollisesti yrityksen taloudellisia menetyksiä. (Fortinet s.a.)

Hakkerilla tarkoitetaan yleisimmin henkilöä, joka murtautuu tietokoneelle tarkoituksenaan tehdä jotain haitallista kuten, varastaa tietoja tai häiritä palveluita. Hakkerit voivat myös toimia positiivisessa mielessä sekä laillisesti ja näitä kutsutaan valkohattuhakkereiksi. Tällöin organisaatiot kutsuvat ulkopuolisen henkilön hyökkäämään heidän järjestelmäänsä, josta he laativat raportin ja organisaatio voi sen pohjalta parantaa järjestelmiään. (Cisco s.a.) Välimaastona on harmaahattuhakkeri. Harmaahattuhakkeri skannaa verkkoa ja järjestelmän haavoittuvuuksia ilman lupaa. Löydettyään haavoittuvuuden, harmaahattuhakkeri joko ilmoittaa haavoittuvuudesta yritykselle, hyödyntää haavoittuvuutta itse tai on tekemättä mitään asialle.

3.1 Verkkoturvallisuus

Verkkoturvallisuutta ymmärtääkseen, on hyvä tietää, mistä verkko koostuu. OSI (Open Systems Interconnection) malli jakaa verkon tiedonsiirron seitsemään eri kerrokseen. Kerrokset tarkastellaan ylhäältä alaspäin ja ne voidaan jakaa kolmeen eri ryhmään tehtävien perusteella. (China & Goodwin 11.6.2024).

Ensimmäiseksi tulee sovelluskerros, jossa tapahtuu kommunikaatiot ja datan vaihto sovellusten ja järjestelmien välillä. Tähän sisältyy sovelluskerros, esityskerros ja istuntokerros. Toinen kerros on OSI mallin ydin eli kuljetuskerros, jossa se käsittelee kaiken dataliikenteen, joka kulkee laitteen järjestelmän ja verkon välillä. Viimeinen kerros on laitekerros, joka koostuu verkkokerroksesta,

datakerroksesta ja fyysisestä kerroksesta. (China & Goodwin 11.6.2024.) Alla oleva kuva 3 hahmottaa OSI mallin kerrokset.

Kuva 3 OSI Verkkomalli (mukailen Cloudflare s.a.)

Verkkoturvallisuus on yksi kyberturvallisuuden osa-alueista. Se keskittyy suojaamaan verkon infrastruktuuria luvattomalta pääsyltä ja sen väärinkäytöltä. Jokainen verkon kerros toteuttaa tiettyjä käytäntöjä ja valvontamekanismeja. Tällöin jos yksi kerros on uhattuna jatkaa toinen kerros suojausta. (Cisco s.a.; Darktrace s.a.)

Verkkoturvallisuus on olennainen osa kyberturvallisuutta, sillä se suojaa arkaluontoista tietoa ja varmistaa verkon moitteettoman toiminnan. (Institute of data 2024.) Verkkoa voidaan turvata monella eri tapaa, mutta suosituimmat ja tunnetuimat tavat ovat palomuuri, virustentorjuntaohjelma, tunkeutumisen havaitsemisjärjestelmä ja VPN.

Palomuuri toimii niin sanotusti verkon portinvartijana, tehtävänä on säätää saapuva ja ulos menevä liikenne asetettujen turvakäytäntöjen mukaan. Käytännöt toimivat siten, että silloin, kun dataa kulkee verkossa, palomuuri tarkastaa saadun paketin ylätunnisteen ja vertaa sitä määritettyihin sääntöihin. Kun tiedot vastaavat sääntöä, pääsee paketti läpi ja jos tiedot eivät vastaa sääntöjä, paketti estetään. Tällöin varmistetaan, että mitään hyökkäyksiä ei pääsisi tapahtumaan. (Cisco s.a.)

Virustentorjuntaohjelma taas havaitsee ja poistaa mahdollisia viruksia koneesta. Tämän lisäksi se skannaa käyttäjän verkkoa ja etsii sieltä mahdolliset poikkeavuudet. (National Cyber Security Centre 2019; Santos, Salam & Dahir 2024, luku 2.)

Intrusion detection system (IDS) on tunkeutumisen havaitsemisjärjestelmän työkalu, joka valvoo verkon liikennettä ja laitteita, etsien epäilyttävää käytöstä. IDS toimii kahdella eri metodilla; poikkeavuuspohjainen havaitseminen tai allekirjoituspohjainen havaitseminen. (IBM 19.4.2023.)

VPN on Virtual Private Network ja sen tarkoituksena on suojata käyttäjän yhteyttä esimerkiksi julkisessa verkossa. VPN salaa yhteyden piilottamalla käyttäjän IP-osoitteen ja uudelleen ohjaamalla liikenteen tiettyyn palvelimeen. (Kaspersky s.a.) IP-osoitteella tarkoitetaan laitteelle annettua yksilöllistä numerosarjaa, joka mahdollistaa laitteen yhteyden ja kommunikaation verkon kanssa. (Fortinet s.a.)

3.2 Palvelunestohyökkäys

Englanniksi tunnettu denial-of-service (DoS) eli palvelunestohyökkäys on hyökkäys, jonka tarkoituksena on ylikuormittaa palvelimen tai verkkosivun liikenne niin pahasti, että kyseinen palvelu on käyttökelvoton. (Cloudflare s.a.)

Distributed denial-of-service attack (DDoS) on alaluokka palvelunestohyökkäyksestä. Siinä hyökkäys toteutetaan laajasti ja siihen käytetään mahdollisesti tuhansia eri laitteita, jotka yhdessä muodostavat bottiverkon. Kun taas DoS toteutetaan yhdellä koneella, joka tekee hyökkäyksestä yksinkertaisemman ja jopa helpommin estettäväksi. (Fortinet s.a.)

Bottiverkot koostuvat monista eri saastuneista tietokoneista ja laitteista. Hakkeri saastuttaa laitteen haittaohjelmalla. Tällöin hyökkääjä pystyy hallitsemaan saastunutta konetta etänä. Lisäksi se mahdollistaa haittaohjelman leviämisen muille samassa verkossa oleville laitteille. (Fortinet s.a.; F-Secure 21.7.2022.) Kun bottiverkko on saavutettu, hyökkääjä antaa käskyn jokaiselle bottikoneelle lähettämään pyynnon kohteen IP-osoitteeseen, aiheuttaen laitteen ylikuormittumisen. Joissain tapauksissa hyökkäykset voivat olla niin vakavia, että ne aiheuttavat vahinkoa laitteen fyysisiin osiin. (Cloudflare s.a.; F-Secure 21.7.2022.)

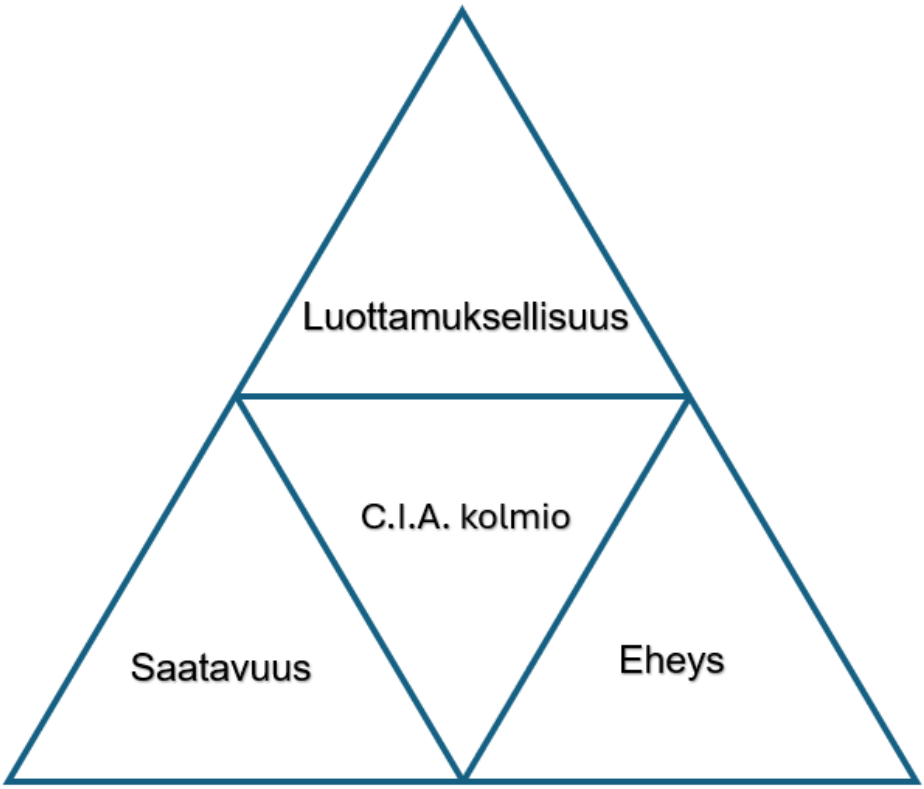
3.3 Tietoturvallisuus

Tietoturvallisuus pyrkii estämään asiattomien henkilöiden pääsyn arkaluontoisiin tietoihin. Tietomurrot ja muut hyökkäykset kohdistuvat yleensä henkilökohtaisten tietojen, kuten pankkitietojen tai yrityssalaisuuksien varastamiseen. Tietoturvallisuuden käytännöt auttavat suojaamaan tietoja.

(Holdsworth & Kosinski 2024.) Perusvahvalla salasanalla ja järjestelmien päivittämisellä pääsee jo pitkälle, eikä hyökkäyksiltä suojautuminen tarvitse olla aina sen monimutkaisempaa.

3.3.1 CIA kolmio

CIA kolmio koostuu kolmesta keskeisestä osa-alueesta, luottamuksellisuus (Confidentiality), eheys (Integrity) ja saatavuus (Availability). Kyseinen malli toimii tietoturvan perustana ja auttaa ohjaamaan tietoturvakäytäntöjä ja -strategioita, sekä auttaa organisaatioita hahmottamaan helpommin mahdollisia heikkouksia järjestelmässään ja havainnollistamaan mitä on mennyt vikaan. (Fortinet s.a.; Fasulo 1.9.2021.) Alla oleva kuva 4 kuvaa CIA kolmion rakennetta.



Kuva 4 C.I.A. kolmio (mukailen Lee 7.4.2025.)

Luottamuksellisuus viittaa tiedon suojelemiseen asiattomalta pääsyltä, tahalliselta tai vahingolta. Keskeisin osa luottamuksellisuuden ylläpitämisellä on varmistaa, että oikeilla henkilöillä on oikeat oikeudet tiettyihin datoihin. (Lee 7.4.2025; Fortinet s.a.)

Yrityksessä tietojen käyttöoikeudet ovat määritelty roolien mukaan esimerkiksi yrityksessä johtohenkilöillä on pääsy työntekijöiden tietoihin. Esihenkilöillä puolestaan on oikeus nähdä vain oman tiiminsä työntekijöiden tiedot. Henkilöstön jäsenillä ei ole pääsyä kenenkään muun yksityisiin

henkilötietoihin. Tällöin pidetään huolta, että kukin saa käyttöönsä vain ne tiedot, jotka ovat tehtäviensä kannalta tarpeellisia.

Eheys osio tarkoittaa tiedon suojelemista luvattomilta tai vahingollisilta muutoksilta, väärentämiseltä tai poistamiselta. Se varmistaa, että tieto pysyy luotettavana ja alkuperäisenä koko sen elinkaaren aikana. (Fortinet s.a; Fasulo 1.9.2021.)

Tällä tarkoitetaan, että tiedot säilytetään paikassa, jossa asiaan kuulumattomilla ei ole pääsyä muokkaamaan tietoja. Esimerkiksi tiedoston jakamisessa yhdellä henkilöllä voi olla pääsy vain katsomaan tiedostoa, kun taas toisella saattaa olla pääsy katsomaan sekä muokkaamaan tiedostoa.

Lopuksi **saatavuudella** tarkoitetaan sitä, että tiedot, järjestelmät ja verkot ovat tarvittaessa saatavilla valtuutetuille henkilöille silloin, kun he haluavat. (Fortinet s.a.) Esimerkiksi yritysten tuottamat palvelut ovat asiakkailleen saatavilla, ajasta ja paikasta riippumatta.

3.3.2 Haittaohjelma

Haittaohjelman tarkoituksen on yleensä vahingoittaa, varastaa tietoa tai keskeyttää laitteen sekä järjestelmän toiminnan. Haittaohjelman hyökkäyksen kohteena voi olla kuka vain aina jokapäiväisestä käyttäjästä hallitukseen asti. (F-Secure 18.2.2022.) Haittaohjelmia on monenlaisia eri tarkoituksiin. Osa on suunniteltu vakoilemaan käyttäjää, huijaamaan tai yleisesti aiheuttamaan vahinkoa.

Vakoiluohjelman tarkoituksena on seurata käyttäjän toimintaa laitteella. Vakoiluohjelman tavoitteena on kerätä arkaluontoista tietoa, sekä sen kanssa käytetään yleensä toista haittaohjelmaa nimeltään näppäinloki. Se tallettaa käyttäjän näppäinpainallukset ja lähettää tiedon hakkerille, jolloin hakkeri saa tietoon esimerkiksi käyttäjän salasanat. (Fortinet s.a.)

Virukset ja madot ovat tunnetuimpia haittaohjelmia ja niiden kyky levitä järjestelmästä toiseen tekee niistä erityisen vaarallisia. Virukset ovat riippuvaisia käyttäjän toiminnasta. Ohjelma leviää esimerkiksi silloin, kun käyttäjä avaa tiedoston tai linkin, johon on upotettu virus. Kun tiedosto tai linkki on avattu, virus aktivoituu ja suorittaa toimintansa. Tämän jälkeen virus voi jatkaa leviämistään muihin laitteella oleviin tiedostoihin. Virus ei kuitenkaan pysty itsestään leviämään ulkopuolisiin laitteisiin ilman käyttäjän apua. (Fortinet s.a.)

Mato taas toimii itsenäisesti. Se käyttää hyväkseen haavoittuvuuksia tai päivittämättömiä ohjelmia levittääkseen itseään eteenpäin. Levitessään mato pystyy asentamaan takaovia, joita hakkeri voi myöhemmin käyttää hyväkseen esimerkiksi ottamalla etäyhteyden laitteeseen. Toisin kuin virus, mato pystyy leviämään itse muihin ulkopuolisiin laitteisiin, jotka käyttävät samaa verkkoyhteyttä, kun saastunut kone. (Fortinet s.a.)

Aina välttämättä haittaohjelman tarkoituksena ei ole vahingoittaa konetta tai varastaa tietoja, joskus tavoitteena on huijata ihmisiä tai aiheuttaa vaivaa. Ensimmäisenä on mainosohjelma, jossa käyttäjän tietokoneelle ilmestyy paljon ei-toivottuja mainoksia. Toinen on pelotteluohjelma, jossa nimensä mukaan yritetään pelotella tai uskotella käyttäjälle, että laite on haavoittuvainen hyökkäyksiä kohtaan, vaikka näin ei ole. (Fortinet s.a.) Tietysti kyseisiä haittaohjelmia voidaan käyttää vahingollisiin tarkoituksiin, kuten pelotteluohjelma kehottaa käyttäjää lataamaan ilmaisen tai maksullisen haittaohjelman poistotyökalun, joka onkin itse haitallinen sovellus. (Fortinet s.a.)

Kiristyshaittaohjelmat ovat yksi suurimmista kyberuhkista sen vaarallisuuden takia. Ne lukitsevat käyttäjän kokonaan ulos laitteeltaan, jolloin laite on käyttökelvoton ja pakottavat maksamaan lunnaita (Kosinski 4.6.2024). Nämä haittaohjelmat yleensä kohdistuvat suuriin organisaatioihin, kuten pankkeihin, sairaaloihin tai muihin kriittisiin infrastruktuureihin. Tällöin hyökkääjät voivat päästä käsiksi elintärkeisiin laitteisiin, esimerkiksi sairaaloiden hengityslaitteisiin. Laitteiden lisäksi vaarana on myös potilastiedot, jolloin hyökkääjät voivat hallita kaikkea ja voivat uhata henkilökuntaa maksamaan lunnaat tai esimerkiksi vuotavat potilastietoja julkisuuteen. (Reed 30.1.2025.)

3.3.3 Social engineering

Social engineering on suomeksi sosiaalinen manipulointi, tällä yritetään manipuloida henkilöä luottamaan lähteeseen ja mahdollisesti jakamaan arkaluontoista tietoa mm. henkilötietoja, salasanoja tai muuta tietoa, jota hyökkääjä voisi katsoa edukseen. Sosiaalinen manipulointi käyttää hyväkseen inhimillisiä virheitä ja heikkouksia, teknisten tai digitaalisten järjestelmien sijasta. (IBM 14.6.2022.)

Yleisin sosiaalinen manipulointi hyökkäys tapa on kalastelu eli Phising. Tässä hyökkääjä lähettää harhaanjohtavan viestin, joka saattaa vaikuttaa aidolta käyttäjälle. Tarkoituksena on saada käyttäjä klikkaamaan viestissä olevaa linkkiä, joka edellyttää kirjautumistunnuksia tai avaamaan jotain liitettä, jonka seurauksena latautuu haittaohjelma. (IBM 14.6.2022.)

3.3.4 Tietomurto

Tietomurrolla tarkoitetaan kyberhyökkäystä, jossa luvattomat henkilöt pääsevät käsiksi luottamuksellisiin tai henkilökohtaisiin tietoihin. Palvelunestohyökkäys, jossa ylikuormitetaan järjestelmää ei ole tietomurto, koska tavoitteena ei ole saada tietoja haltuun. Kalasteluviesti on tietomurto, koska siinä pyritään pääsemään käsiksi arkaluontoisiin tietoihin. (Kosinski 24.5.2024.) Tietomurrot voivat johtua monesta eri syystä, kuten tahattomasta virheestä, järjestelmän heikkoudesta tai ihan jopa pienestä inhimillisestä virheestä (Fortinet s.a.).

4 Tutkimus

Tässä kappaleessa käsitellään opinnäytetyön tutkimusmenetelmiä sekä perustelut, miksi menetelmä valittiin. Lisäksi käsitellään työn aineistoa, tavoitteita ja sen luotettavuutta.

4.1 Tutkimuksenmenetelmät

Tutkimusmenetelmäksi valittiin kirjallisuuskatsaus, joka toteutettiin kvalitatiivisena eli laadullisena tutkimuksena. Kirjallisuuskatsaus valittiin tutkimusmenetelmäksi, koska se tarjoaa kattavan ja monipuolisen lähestymistavan aiheeseen, sekä auttaa hahmottamaan kokonaiskuvaa ilman yksittäisiin tapauksiin rajoittuvaa aineistoa.

Kirjallisuuskatsaus on keskeinen osa tutkimusprosessia, sillä se kartoittaa aiemmat tutkimukset ja analysoi niiden tuloksia. Katsauksessa voidaan hyödyntää monenlaisia lähteitä, kuten tieteellisiä artikkeleita, kirjoja, konferenssijulkaisuja ja muita relevantteja dokumentteja. (Editage 3.4.2024.)

Kvalitatiivinen tutkimus eli laadullinen tutkimus pyrkii syvälliseen ymmärrykseen tutkittavasta kohteesta, keskittyen kokemuksiin, mielipiteisiin ja merkityksiin numeeristen analyysien sijasta. Laadullisen tutkimustyypin vahvuus on sen joustavuus, koska se mahdollistaa aineiston keräämisen monipuolisilla menetelmillä, kuten haastatteluilla, havainnoilla ja tekstianalyysillä. (Bhandari 19.6. 2020.)

4.2 Tutkimuksen aineisto

Työssä on hyödynnetty internetistä löydettyjä artikkeleita, blogipostauksia ja raportteja. Hakukoneena on pääsääntöisesti käytetty Google ja Google Scholar. Olen hyödyntänyt myös koulun kirjastopalvelua (HH finna) ja pääsyä O'reilly e-kirjoihin. Hakukysymyksinä toimi mm. How ai enhances cybersecurity? How ai is used in cybersecurity? What are the risks of ai in cybersecurity? Ethics of ai in cybersecurity?

4.3 Lähteiden luotettavuus

Taatakseni lähteiden luotettavuuden, keskityin lähteisiin, jotka ovat julkaistu viimeisen viiden vuoden sisällä. Työssä käytetyistä lähteistä on peräisin tunnetuista ja asiantuntevista yrityksistä kuten IBM tai Fortinet. Lisäksi lähteiden valinnassa käytin omaa kriittistä ajattelukykyä, joka yhdistyy aiemmin opittuihin asioihin ja tietotaitoon. Vertasin myös sisällön tietoa muihin tarjottuihin lähteisiin varmistaakseni paikkansapitävyyden ja luotettavuuden.

4.4 Tutkimuksen tavoitteet

Tämän tutkimuksen tavoitteena on tarkastella, kuinka tekoäly voi parantaa kyberturvallisuutta uhkien havaitsemisessa ja ennaltaehkäisemisessä. Lisäksi tarkastellaan muutamia keskeisiä kyberuhkia ja sitä, miten tekoäly voi auttaa niiden torjunnassa. Lopuksi käsitellään tekoälyn käyttöön liittyviä eettisiä kysymyksiä.

Tutkimuskysymykset

Lähdin tutkimaan aihetta etsien vastausta kysymykseen:

”Miten tekoälyä voi hyödyntää kyberturvallisuudessa?”.

Vastauksien löytämisen apuna oli muutama alakysymys:

”*Mitä riskejä tai haasteita tekoäly tuo kyberturvallisuuteen?”*

”Mitä eettisiä puolia pitää ottaa huomioon tekoälyn käytössä kyberturvallisuudessa?”

5 Tutkimuksen tulokset

Tutkimuksen tavoitteena oli selvittää tekoälyn rooli kyberturvallisuudessa sekä arvioida siihen liittyviä mahdollisia haittoja sekä riskejä. Tutkimuksen aineiston perusteella kävi ilmi, että tekoälyllä on merkittävä rooli kyberuhkien ennaltaehkäisyssä ja torjunnassa. Seuraavissa kappaleissa käsitellään tarkemmin tutkimuksen keskeisiä löytöjä.

5.1 Tekoälyn hyödyt kyberturvallisuudessa

Valittujen lähteiden perusteella kävi ilmi useita merkittäviä tapoja, joilla tekoälyä voidaan hyödyntää kyberturvallisuuden parantamisessa. Keskeisiä havaintoja olivat: uhan havaitseminen, ison datamäärän analysointi ja käytäntöjen automaatio.

Uhan tiedustelu on yksi kyberturvallisuuden käytännöistä, jolla tietoturva-asiantuntijat pysyvät ajan tasalla uhkista. Tekoälypohjaisella järjestelmällä pystytään tehostamaan uhkien tiedustelua. Tällöin on mahdollista analysoida suuria määriä dataa samanaikaisesti eri lähteistä, kuten keskustelupalstoilta, pimeistä verkoista ja avoimista lähteistä. (Kolosnjaji, Xiao, Xu & Zarras 2024, luku 1.)

Näiden niin sanotusti perinteisten lähteiden lisäksi, erityisesti syväoppimisen mekanismin avulla on mahdollista laajentaa lähteitä sosiaaliseen mediaan sekä jopa ulkomaalaisille keskustelupalstoille, joissa yleisesti on ollut esteenä kielimuuri. (Marchal, Nawrotek & WithSecure 2024, 21-22.) Tämä tarkoittaa enemmän monipuolisia lähteitä, joka laajentaa tietoisuutta uhista ja haavoittuvuuksista, jolloin organisaatiot pystyvät olemaan proaktiivisempia mahdollisia hyökkäyksiä kohtaan.

Toinen tapa, jossa hyödynnetään tekoälyn kykyä analysoida isoja määriä tietoa, on yrityksen verkoissa. Tässä koneoppi malli analysoi verkossa liikkuvia tiedostoja ja käyttäytymisiä, etsien mahdollisia poikkeavuuksia. Näistä asioista tekoäly pystyy tekemään asiantuntijalle kootun raportin verkossa olevasta liikenteestä. (Marchal ym. 2024, 11-14.)

Verkossa käyttäytymisellä tarkoitetaan työntekijöiden käyttäytymistä yrityksen verkossa. Koneoppimallin algoritmille on opetettu käyttäjien käyttäytymisen historiaa organisaatiossa, mm. kirjautumismalleista, tiedostoista, verkkoliikenteen käytöstä ja sovelluskäytöstä. Tällöin järjestelmä saa kuvan käyttäjän, niin sanotusta normaalista käyttäytymisestä, jolloin järjestelmä havaitsee tämän ja ilmoittaa asiantuntijalle käyttäytymisen muutoksesta ja poikkeavaisuuksista. Tämä mahdollistaa yritysten ajan tasalla pysymisen mahdollisista sisäisistä uhista, olivatpa ne tahallisesti tai vahingossa syntyneitä. (Mohamed 2025, luku 4.2.2.) Esimerkiksi jos työntekijän käyttäjätili on päätynyt jonkun ulkopuolisen tietoon ja käyttäjä yrittää kirjautua jostain poikkeuksellisesta paikasta, tekoäly voi estää käyttäjän tai hälyttää asiasta asiantuntijalle.

Tästä päästään uhan reagoinnin automatisointiin. Uhan havaitessa tekoälyjärjestelmä voi itse reagoida hyökkäykseen tai uhkaan, yrityksen ennalta määritettyjen toimintamallien mukaan. Tekoälyjärjestelmä voi reagoida esimerkiksi tilien estämisellä tai saastuneen laitteen poistamisella verkosta. (Mohamed 2025, luku 8.2.)

Tietoturvaraportit ovat asiakirjoja, jotka kokoavat yhteen koko organisaation tietoturvapoikkeamia eli tapahtumia, jossa organisaation tietoturva on ollut uhattuna. Tekoälyn avulla näiden pitkien raporttien analysointi voidaan automatisoida, jolloin järjestelmä pystyy tarjoamaan asiantutijalle tiiviin raportin. Raportti sisältää yrityksen järjestelmän haavoittuvuuksia ja hyökkäystekniikoita, joita hakkerit voisivat mahdollisesti hyödyntää. Lisäksi raportti sisältää analyysin suojautumiskeinoista ja torjuntastrategioista. (Mohamed 2025, luku 4.4.1.)

Automatisointia voidaan myös hyödyntää käyttäjien tilien ylläpidossa. Uuden työntekijän aloittaessa yrityksessä, järjestelmä pystyy heti luomaan työntekijälle sopivat käyttöoikeudet. Työntekijän lopettaessa yrityksessä järjestelmä poistaa käyttöoikeudet ja käyttäjän. (Santos ym. 2024, luku 3.) Tämä auttaa ylläpitämään tietoturvallisuuden CIA kolmiota, erityisesti luottamuksellisuutta. Tällöin voidaan taata, että oikeutetuilla ihmisillä on pääsy tietoihin.

Sosiaalisen manipuloinnin estäminen on enemmän käytännön läheisempää esimerkiksi organisaation ja työntekijöiden välillä. Isossa terveydenhuollon organisaatiossa työskennellessäni sain kokemusta yrityksen lähettämistä tietojen kalastelua simuloivista sähköpostiviesteistä työntekijöille, jotka heidän tuli tunnistaa ja ilmoittaa tietojenkalasteluksi. Tällä tavoitellaan henkilöstön tietoisuuden ja osaamisen edistämistä tietoturva asioista. Apuna käytetään generatiivista tekoälyä, joka pystyy tekemään realistisia kalasteluviestejä työntekijöille tunnistettavaksi (Islam 2024, luku 2.6).

Lisäksi generatiivinen tekoäly pystyy tekemään hyökkäysskenaarion palvelunestohyökkäyksestä. Tällöin asiantuntijat pystyvät testata jo olemassa olevia suojauksia ja tunnistaa heikkoudet järjestelmässä. (Islam 2024, luku 4.2.2.) Tämä auttaa yritystä rakentamaan vahvempia ja kestävämpiä järjestelmäsuojauksia, jotka mahdollisesti pystyvät estämään hyökkäyksen tai minimoimaan sen haittavaikutukset. Katsotaan tätä verkkoturvallisuuden lisäksi myös CIA kolmion kautta, jossa tämä auttaa ylläpitämään saatavuus osiota CIA kolmiosta.

Tekoäly on keskeinen osa kyberturvallisuuden kehitystä, erityisesti kiristyshaittaohjelmien hyökkäyksien torjunnassa. Tekoälyn kyky analysoida aiempia hyökkäysmalleja eri lähteistä ja oppia niistä mahdollistaa ennakoivan suojautumisen ja hyökkäyksien tehokkaan estämisen. Ennakoivan analyysin ja uhan tunnistamisen avulla tekoäly voi tunnistaa varhaiset merkit kiristyshaittaohjelma hyökkäyksestä jo ennen sen toteutumista. (Ferdous, Islam, Mahboubi & Islam 2024, luku 3; Mohamed 2025, luku 16.4.2.)

5.2 Tekoälyn riskit ja haasteet

Aineistoni perusteella muodostui myös käsitys, että tekoälyyn liittyvät riskit ja haasteet voidaan jakaa kahteen osaan; hyökkäykset ja tekoälyjärjestelmän haasteet, kuten esimerkiksi tekoälyn oma turvallisuus ja tietojen yksityisyys sekä tekoälypohjaiset hyökkäykset.

Tekoälyn tulee saada suuria määriä erilaista koulutusdataa, jotta tekoälyjärjestelmä pystyisi tuottamaan mahdollisimman tarkkoja havaintoja ja analyysejä. Näitä ovat esimerkiksi hyökkäystapoja ja yrityksen lokitiedostoja, joka luokitellaan korkeanlaatuiseksi tiedoksi. Tämä saattaa vaikuttaa yritysten tietojen luovutukseen. (Mohamed 2025, luku 9.1.) Nämä saattavat lisätä entisestään yritysten epäluottamusta tietojen luovuttamista kohtaa.

Datamyrkytys ja adversarial hyökkäys ovat yleisimpiä tekoälyn kohtaamia hyökkäystapoja. Hyökkääjä muokkaa järjestelmän syötetietoja, jolloin seurauksena voi olla tekoälyn virheellisiä tuloksia. Hyökkääjä voi manipuloida tekoälyjärjestelmää sivuuttamaan esimerkiksi haittaohjelmat. Tekoälyn datan muutoksen ei aina tarvitse olla iso, pienikin muutos esimerkiksi koulutusdatassa voi muokata kokonaan tekoälyn käytöstä. (Kolosnjaji ym. 2024, luku 18; Mohamed 2025, luku 9.3.)

Tekoälyn joutuessa hyökkäyksen tai datan muutoksen kohteeksi, voi seurauksena olla jatkuvien väärien ilmoitusten tekeminen, joka voi johtaa siihen, että asiantuntijat eivät enää usko järjestelmään ja saattavat jättää huomioimatta tulevaisuudessa tulevat ilmoitukset. (Mohamed 2025, luku 9.2). Hyökkäykset eivät ole ainoa asia miksi tekoälyjärjestelmä tuottaisi virheellisiä tuloksia, myös huonon koulutusdatan takia tekoäly voi antaa virheellisiä tai harhaanjohtavia tuloksia.

On hyvä huomioida myös asiantuntijoiden mahdollinen yliluottamus tekoälyyn. Riskinä yliluottamuksessa on asiantuntijoiden liiallinen luottaminen tekoälyjärjestelmään, joka voi heikentää ihmisten ja tekoälyn välistä vuorovaikutusta. Vaarana on, että ihmiset hyväksyvät tekoälyn tuottamat tulokset ilman kyseenalaistamista, jolloin järjestelmän mahdolliset virheet voivat jäädä huomaamatta. (Dinu, Vasile & Georgescu 2024, 44.) Tämän takia työntekijöiden kunnollinen koulutus ja riskienhallinta ovat olennaisia asioita tekoälyn käytössä.

Tekoälyjärjestelmään kohdistunut hyökkäys ei välttämättä ole itse tekoälyjärjestelmän vika, vaan ulkopuolisen järjestelmän, kuten yrityksen verkkojen tai sovellusten aiheuttamia. Näiden järjestelmien kautta hakkeri on voinut päästä sisälle yritykseen ja tällöin hyödyntää olemassa olevia haavoittuvuuksia, sekä mahdollisesti toteuttaa hyökkäyksen tekoälyyn kuten datan myrkyttämisen. (Hamon, Junklewitz, Soler Garrido ja Sanchez 2024, luku 4.)

Tekoälyjärjestelmän käyttöön otettaessa yrityksessä voi tapahtua inhimillinen virhe ja yritys saattaa keskittyä tekoälyyn ja sen turvaamiseen enemmän, kuin muihin osa-alueisiin. Tästä syystä on hyvä

muistaa muiden järjestelmien turvallisuuden ylläpito, koska tästä saattaa seurata perhosefekti eli pieni haavoittuvuus yhdessä osassa, voi johtaa laajempaan kyberhyökkäykseen.

Ihmisen rooli pysyy kriittisenä, vaikka tekoäly voi merkittävästi parantaa kyberturvallisuutta. Virheet, väärät päätökset ja huolimattomuus voivat johtaa suuriin ongelmiin riippumatta siitä, kuinka edistynyt tekoälyjärjestelmä on. Järjestelmäriskien lisäksi tekoäly aiheuttaa kyberturvallisuudelle haasteita myös siksi, koska sen avulla pystytään toteuttamaan kehittyneempiä kyberhyökkäyksiä, kuten kalasteluviestejä tai haittaohjelmia.

Tekoälyä voidaan hyödyntää yhtä lailla sekä kyberturvallisuudessa että hakkeroinnissa, esimerkiksi automatisoimalla tiedon keruuta mm. salasanat ja järjestelmän heikkoudet. Automatisoinnin avulla hakkerit voivat toteuttaa monia hyökkäyksiä samanaikaisesti sekä löytää uusi tapoja hyökätä. Kiinnijäämisen riskiä pienentää tekoälyn kyky tehdä nopeaa tiedonkeruuta, analysointia ja skannaamista, jolloin itse hakkeri viettää vähemmän aikaa kohteen järjestelmässä tai verkossa. (Aksela, Marchal, Patel, Rosenstedt & WithSecure 2022, 9-10.)

Tekoälyn avulla hakkerit voivat tehostaa aikaisempia hyökkäyksiä. Laajan tiedonkeruun avulla hakkerit pystyvät tekemään tarkempia kalasteluviestejä. Esimerkiksi tekoälypohjainen haittaohjelma pystyy mukautumaan ympäristöönsä, sulautua verkkoon ja naamioitua. (Aksela ym. 2022, 12-16.)

Tekoälyn myötä kalasteluhyökkäykset voidaan toteuttaa kahdella tavalla; viestillä, joka on yleisin menetelmä sekä puhelimitse, jossa hyödynnetään deepfake-teknologiaa. Perinteisesti nämä hyökkäykset ovat toteutettu viestillä, mutta tekoälyn avulla hakkerit voivat jäljitellä ihmisen puhetyyliä ja ääntä. (Golda, Mekonen, Pandey, Singh, Hassija, Chamola & Sikdar 2024, luku 5.) Tämä mahdollistaa huijaukset, jossa hyökkääjä esiintyy kyseisenä henkilönä ja voi esimerkiksi soittaa kohde henkilön pankkiin saadakseen lisätietoja tai soittaa kohde henkilön kavereille saadakseen henkilötietoja.

5.3 Eettiset puolet

Tuotaessa tekoälyä kyberturvallisuuteen pitää ottaa huomioon kolme tekijää; läpinäkyvyys, tietosuoja ja ennakkoluulot. Kuten aikaisemmissa luvuissa on käsitelty, tekoälyn koulutuksessa ja ylläpidossa käytetään isoja määriä dataa, jotka voivat sisältää myös henkilökohtaisia tietoja. Tähän samaan tietosuoja kysymykseen liittyvät yrityksessä tehdyt käyttäytymisen profiilit. Tästä herää kysymys, missä ja miten tietoja säilytetään.

Mohamedin artikkelissa korostetaan yritysten vastuuta tekoälyn käytössä erityisesti sen vaikutusta yksityisyyteen ja läpinäkyvyyteen. Yritysten tulee varmistaa, että tekoälyratkaisut tukevat

turvallisuutta ilman, että ne vaarantavat työntekijöiden ja asiakkaiden tietosuojaa. (Mohamed 2025, luku 9.5.)

Samassa artikkelissa on kiinnitetty huomiota tekoälyn läpinäkyvyyteen, koska tekoälyn järjestelmän päättelykyky perustuu mustan laatikon (black box) operointiin. Tässä mustassa laatikossa asiantuntijat eivät näe miksi tai miten päätös tehtiin, heillä on vain tiedossa syötetty lähtötieto ja lopputulos. Tämän takia asiantuntijoilla saattaa olla vaikeuksia luottaa järjestelmän päätöksiin. (Mohamed 2025, luku 9.5; Kosinski 29.10.2024.)

Tekoälyn tehdessä virheellisen ratkaisun, läpinäkyvyyden säilymisen vuoksi, tulee huomioida vastuukysymykset. Tekoälyn tehdessä virheellisen päätöksen, tulee selvittää päätökseen johtaneet perusteet sekä kuka vastaa viime kädessä tehdystä päätöksestä. Esimerkiksi jos tekoäly toimisi virheellisesti uhan estämisessä ja aiheuttaisi vakavimpia ongelmia, herää kysymys, kuka siitä olisi vastuussa. Vaikka tekoäly toimisi itsenäisesti, sen toiminta kuitenkin perustuu yrityksen opetettuihin toimintatapoihin. Virheen tapahtuessa, tulee ratkaista, onko vastuussa järjestelmän kehittäjä, asiantuntija tai joku muu, joka on kouluttanut tekoälyä. (Dinu ym. 2024, 44.)

Lopuksi on syytä tarkastella epäoikeudenmukaisuutta ja vinoumaa. Mikäli tekoälyn koulutusdatassa esiintyy ennakkoluuloja, esimerkiksi oletus siitä, että tietty ryhmä todennäköisemmin toteuttaa hyökkäyksiä, tällöin saattaa tekoälyjärjestelmä tarkkailla kyseistä ryhmää enemmän kuin muita ryhmiä. Seurauksena voi olla epäoikeudenmukaisia johtopäätöksiä, jotka voivat johtaa syrjiviin rakenteisiin. (Mohamed 2025, luku 9.5.) Ennakkoluulojen ja vinoutumien välttämiseksi tulee miettiä, miten tekoälyä koulutetaan ja minkälaista koulutusdataa käytetään. Koulutusdatan valinnassa pitää huomioida, ettei se edusta vain yhtä näkökulmaa tai virheellisiä painotuksia. Asiantuntijoiden on tärkeää muistaa päivittää koulutusdataa huomatessaan epäoikeudenmukaisia ja virheellisiä päätöksiä.

6 Pohdintaa

Tässä luvussa on tarkoituksena pohtia kirjallisuuskatsauksen kulkua, tuloksia, tutkimuksen luotettavuutta ja mahdollista jatkotutkimusaihetta. Lisäksi arvioidaan tavoitteiden saavuttamista, tutkimuksen ongelmia sekä oman oppimisen kehittymistä.

6.1 Tutkimuksen luotettavuus

Tutkimuksen luotettavuuden varmistamiseksi lähteet on valittu vuosilta 2020 – 2025, jotta tieto olisi mahdollisimman ajantasaista ja luotettavaa. Tieteelliset artikkelit ovat vertaisarvioituja, mikä vahvistaa niiden akateemista luotettavuutta. Sekundaariset lähteet puolestaan perustuvat tunnettujen yritysten julkaisemiin raportteihin ja artikkeleihin, jotka ovat vertailtu muiden lähteiden kanssa puolueettomuuden varmistamiseksi, eikä tutkimukseen ole sisälletty mainospuhetta, jos sellaista on esiintynyt.

6.2 Tavoitteiden saavuttaminen

Työn tavoitteena oli selvittää tekoälyn vaikutus kyberturvallisuudessa. Voitaisiin todeta, että tavoitteet saavutettiin ja saatiin kattava kuva, millä eri tavoin tekoälyä voidaan hyödyntää. Hyötyjen lisäksi työssä saatiin selville tekoälyn mahdolliset haittavaikutukset. Lisäksi käsiteltiin eettisiä puolia ja haasteita, mitä tekoälyn käyttö mahdollisesti tuo kyberturvallisuuteen.

6.3 Ongelmat

Ongelmia ilmeni aluksi aiheen laajuuden vuoksi, mikä teki työn aloittamisesta haastavaa. Tästä johtui työn viivästyminen. Lähteiden etsiminen tekoälyn eettisistä puolista kyberturvallisuudessa osoittautui aluksi haastavaksi. Haasteena on ollut lisäksi luki- ja kirjoitushäiriö, joka on vaikeuttanut tekstin tuottamista sekä sen ymmärtämistä.

6.4 Tulosten tarkastelu

Tuloksia on tarkasteltu suhteessa tutkimuskysymyksiin ja niihin saatuihin vastauksiin.

Miten tekoälyä voi hyödyntää kyberturvallisuudessa?

Tutkimuksen tulokset osoittivat, että tekoälyn hyödyntäminen kyberturvallisuuden parantamisessa voi tehostaa uhkien havaitsemista ja torjuntaa. Nopea uhkan havaitseminen voi säästää yrityksen kalliilta tietovuodolta.

Tekoäly parantaa verkkoturvallisuutta monin tavoin. Se tehostaa verkkoliikenteen poikkeavuuksien tunnistamisen ja tietojen suojaamista, havaitsemalla epätavallisia kirjautumisia ja estämällä datan

manipulaatiota. Lisäksi tekoälyjärjestelmät voivat tunnistaa palvelunestohyökkäyksiä ja aktivoida automatisoituja puolustusmekanismeja. Reaaliaikainen uhan analyysi mahdollistaa haitallisen verkkoliikenteen nopean tunnistamisen ja auttaa uhan estämisessä, mikä vähentää tietoturvariskejä.

Mitä riskejä tai haasteita tekoäly tuo kyberturvallisuuteen?

Tekoäly tuo mukanaan myös kyberturvallisuuteen liittyviä haasteita. Yksi merkittävä haitta on tekoälypohjaiset hyökkäykset, jossa tekoälyn avulla voidaan toteuttaa entistä vaikeammin havaittavia hyökkäyksiä.

Mitä eettisiä puolia pitää ottaa huomioon tekoälyn käytössä kyberturvallisuudessa?

Tutkimuksen tulokset osoittavat, että tekoälyn eettiset haasteet ja tekoälyn omat riskit kulkevat käsi kädessä. Esimerkiksi tekoälyn huonosti koulutettudata voi altistua herkemmin hyökkäyksille tai aiheuttaa tahattomia tietovuotoja. Eettisyyden näkökulmasta on keskeistä pohtia sen läpinäkyvyyttä ja tietosuojaa, jotta sen käyttö pysyy turvallisena ja vastuullisena.

Loppujen lopuksi olen tyytyväinen saamiini tuloksiin ja sain mielestäni hyvin hahmoteltua, miten eri tavoin tekoäly auttaa kyberturvallisuutta sekä sain vastauksen kaikkiin tutkimuksen kysymyksiin. Keskeisin tulos on se, että tekoäly ei ole vain itsenäinen ratkaisu, vaan se toimii parhaiten ihmisten tukena.

6.5 Oma oppiminen

Tutkimusprosessi on tarjonnut minulle syvällistä ymmärrystä tekoälystä ja sen osa-alueista. Erityisesti tekoälyn historia osoittautui kiinnostavaksi. Olen oppinut tunnistamaan tekoälyn eettisiä haasteita, jotka ovat olennaisia sen vastuullisessa hyödyntämisessä.

Tämän tutkimuksen kautta tieto ei ole ainoastaan laajentanut teknologista osaamistani, vaan myös vahvisti ymmärrystä erilaisista tutkimusmenetelmistä, kuten kirjallisuuskatsauksesta sekä laadullisista analysointitavoista.

Uskon että oppimani asiat tulevat olemaan arvokkaita, niin tulevassa työelämässä, kuin jatko-opinnoissa, sillä ne tarjoavat vankan pohjan kriittiselle ajattelulle ja tiedonkäsittelylle.

6.6 Jatkotutkimus

Tekoäly on kaksijakoinen kyberturvallisuudessa: sitä voidaan hyödyntää puolustamisessa sekä hyökkäämisessä. Tämä herättää kysymyksen siitä, onko tulevaisuus tekoälyjärjestelmien vastakkaistaistelu, jossa tekoälyjärjestelmät taistelevat toisiaan vastaan kyberturvallisuudessa.

Tämä aihe voisi olla hyvä jatkotutkimuksen aihe. On hyvä tutkia, miten tekoälypohjaisten puolustusmekanismit voisivat kehittyä vastaamaan yhä kehittyneempiä tekoälypohjaisia hyökkäyksiä vastaan.

Lähteet

Aksela, M., Marchal, S., Patel, A., Rosenstedt, L. & WithSecure. 2022. The security threat of AI-enabled cyberattacks. Liikenne- ja viestintävirasto Traficom, Kyberturvallisuuskeskus. Helsinki. Luettavissa: https://www.traficom.fi/sites/default/files/media/publication/TRAFICOM_The_security_threat_of_AI-enabled_cyberattacks%202022-12-12_en_web.pdf Luettu: 7.4.2025.

Bhandari, P. 19.6.2020. What Is Qualitative Research? | Methods & Examples. Luettavissa: https://www.scribbr.com/methodology/qualitative-research/ Luettu: 22.4.2025.

China, C, R. & Goodwin, M. 11.6.2024. What is the OSI model? Luettavissa: https://www.ibm.com/think/topics/osi-model Luettu: 15.3.2025.

Cisco. s.a. What Is a Hacker? Luettavissa: https://www.cisco.com/c/en/us/products/security/what-is-a-hacker.html Luettu: 27.4.2025.

Cisco. s.a. What Is Network Security? Luettavissa: https://www.cisco.com/c/en/us/products/security/what-is-network-security.html Luettu: 1.4.2025.

CloudFlare. s.a. What is natural language processing (NLP)? Luettavissa: https://www.cloudflare.com/en-gb/learning/ai/natural-language-processing-nlp/ Luettu: 15.4.2025.

CloudFlare. s.a. What is the OSI Model? Luettavissa: https://www.cloudflare.com/en-gb/learning/ddos/glossary/open-systems-interconnection-model-osi/ Luettu: 15.3.2025.

CloudFlare. s.a. What is a DDoS attack? Luettavissa: https://www.cloudflare.com/en-ca/learning/ddos/what-is-a-ddos-attack/ Luettu: 15.3.2025.

Copeland, B, J. 16.5.2025. Methods and goals in AI. Luettavissa: https://www.britannica.com/technology/artificial-intelligence/Methods-and-goals-in-AI Luettu: 15.3.2025.

Darktrace. s.a. What is network security? Luettavissa: https://www.darktrace.com/cyber-ai-glossary/network-security Luettu: 1.4.2025.

Deep Instinct. s.a. Deep Learning. Luettavissa: https://www.deepinstinct.com/glossary/deep-learning Luettu: 15.4.2025.

Dinu, A., Vasile, P. C. & Georgescu, A. 2024. AI-driven solutions for cybersecurity: comparative analysis and ethical aspects. Romanian Journal of Information Technology and Automatic Control, 34, 3, s. 35-48. Luettavissa: https://rria.ici.ro/documents/1203/art._Dinu_Vasile_Georgescu.pdf Luettu: 12.5.2025.

Editage. 3.4.2024. What is Literature Review? Definition, Types and Examples. Editage blogi. Luettavissa: https://www.editage.com/blog/what-is-literature-review-definition-types-and-examples/ Luettu: 22.4.2025.

Elastic Platform Team. 2024. NLP vs. LLMs: Understanding the differences. Luettavissa https://www.elastic.co/blog/nlp-vs-llms Luettu: 15.4.2025.

Fasulo, P. 1.9.2021. What is the CIA Triad? Definition, Importance, & Examples. SecurityScorecard blogi. Luettavissa: https://securityscorecard.com/blog/what-is-the-cia-triad/ Luettu: 1.4.2025.

F-Secure. 21.7.2022, Mikä on palvelun-esto-hyökkäys (DDoS)? Luettavissa: https://www.f-secure.com/fi/articles/what-is-ddos Luettu:15.3.2025.

F-Secure. 18.2.2022. Mikä on haittaohjelma? Näin pysyt turvassa vaarallisilta ohjelmilta. Luettavissa: https://www.f-secure.com/fi/articles/what-is-malware Luettu: 15.3.2025.

Ferdous, J., Islam, R., Mahboubi, A. & Islam, M, Z. 2024. AI-Based Ransomware Detection: A Comprehensive Review. IEEE Access, 12. Luettavissa: https://ieeexplore.ieee.org/abstract/document/10681072 Luettu: 1.5.2025.

Fortinet. s.a. Malware: Definition, Types and Methods of Detection & Prevention. Luettavissa: https://www.fortinet.com/resources/cyberglossary/malware Luettu: 15.3.2025.

Fortinet. s.a. What Is Cybersecurity? Luettavissa: https://www.fortinet.com/resources/cyberglossary/what-is-cybersecurity Luettu: 18.3.2025.

Fortinet. s.a. What Is A Data Leak? Causes And Prevention. Luettavissa: https://www.fortinet.com/resources/cyberglossary/data-leak Luettu: 14.3.2025.

Fortinet. s.a. What Is A Data Breach? Luettavissa: https://www.fortinet.com/resources/cyberglossary/data-breach Luettu: 17.3.2025.

Fortinet. s.a. What Is DDOS Attack? https://www.fortinet.com/resources/cyberglossary/ddos-attack Luettu: 15.3.2025.

Fortinet. s.a. CIA Triad. Luettavissa: https://www.fortinet.com/resources/cyberglossary/cia-triad Luettu: 1.4.2025.

Geeks for Geeks. 28.6.2024. What is the difference between a strong AI and a weak AI? Luettavissa: https://www.geeksforgeeks.org/what-is-the-difference-between-a-strong-ai-and-a-weak-ai/ Luettu: 1.4.2025.

Golda, A., Mekonen, K., Pandey, A., Singh, A., Hassija, V., Chamola, V. & Sikdar, B. 2024. Privacy and Security Concerns in Generative AI: A Comprehensive Survey. IEEE Access, 12. Luettavissa: https://ieeexplore.ieee.org/document/10478883 Luettu: 1.5.2025.

Hamon, R., Junklewitz, H., Garrido, J, S. & Sanchez, I. 2024. Three Challenges to Secure AI Systems in the Context of AI Regulations. IEEE Access, 12. Luettavissa: https://ieeexplore-ieee-org.ezproxy.haaga-helia.fi/document/10506836 Luettu: 1.5.2025.

IBM. 6.10.2021. What is a neural network? Luettavissa: https://www.ibm.com/think/topics/neural-networks Luettu: 27.4.2025.

IBM. 14.6.2022. What is social engineering? Luettavissa: https://www.ibm.com/think/topics/social-engineering Luettu: 17.3.2025.

IBM. 19.4.2023. What is an intrusion detection system (IDS)? Luettavissa: https://www.ibm.com/think/topics/intrusion-detection-system Luettu: 1.4.2025.

Institute of Data. 26.3.2024. The Importance of Network Security. Institute of Data blogi. Luettavissa: https://www.institutedata.com/blog/the-importance-of-network-security/ Luettu: 1.4.2025.

Islam, M. R. 2024. Generative AI, Cybersecurity, and Ethics. Wiley. Hoboken, New Jersey. E-kirja. Luettavissa: https://learning.oreilly.com/library/view/generative-ai-cybersecurity/9781394279265/ Luettu: 12.5.2025.

Kufel, J., Bargieł-Łączek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik, M., Czogalik, Ł., Dudek, P., Magiera, M., Lis, A., Paszkiewicz, I., Nawrat, Z., Cebula, M. & Gruszczyńska, K. 2023. What Is Machine Learning, Artificial Neural Networks and Deep Learning? Examples of Practical Applications in Medicine. Diagnostics (Basel),13, 15. Luettavissa: https://www.mdpi.com/2075-4418/13/15/2582 Luettu: 15.5.2025.

Holdsworth, J. & Kosinski, M. 26.7.2024. What is information security (InfoSec)? Luettavissa: https://www.ibm.com/think/topics/information-security Luettu: 1.4.2025.

Holdsworth, J. & Scapicchio, M. 17.6.2024. What is deep learning? Luettavissa: https://www.ibm.com/think/topics/deep-learning Luettu: 15.3.2025.

Kanade, V. 4.4.2022. What Is Machine Learning? Definition, Types, Applications, and Trends. Luettavissa: https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/ Luettu: 15.4.2025.

Kolosnjaji, B., Xiao, H., Xu, P. & Zarras, A. 2024. Artificial Intelligence for Cybersecurity. Packt Publishing. Birmingham, United Kingdom. E-kirja. Luettavissa: https://learning.oreilly.com/library/view/artificial-intelligence-for/9781805124962/ Luettu: 1.5.2025.

Kosinski, M. 29.10.2024. What is black box artificial intelligence (AI)? Luettavissa: https://www.ibm.com/think/topics/black-box-ai Luettu: 1.5.2025.

Kosinski, M. 24.5.2024. What is a data breach? Luettavissa: https://www.ibm.com/think/topics/data-breach Luettu: 17.3.2025.

Kosinski, M. 4.6.2024. What is ransomware? Luettavissa: https://www.ibm.com/think/topics/ransomware Luettu: 15.3.2025.

Kaspersky. s.a. What is VPN? How It Works, Types of VPN. Luettavissa: https://www.kaspersky.com/resource-center/definitions/what-is-a-vpn Luettu: 1.4.2025.

Lee, I. 7.4.2025. CIA Triad Definition. Examples of Confidentiality, Integrity, and Availability. Luettavissa: https://www.wallarm.com/what/cia-triad-definition Luettu: 14.3.2025.

Marchal, S., Nawrotek, B. & WithSecure. 2024. Applying artificial intelligence in cybersecurity. Liikenne- ja viestintävirasto Traficom, Kyberturvallisuuskeskus. Helsinki. Luettavissa: https://www.kyberturvallisuuskeskus.fi/sites/default/files/media/file/Applying%20AI%20in%20cybersecurity_EN.pdf Luettu: 7.4.2025.

Mohamed, N. 2025. Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms. Knowledge and information systems. Luettavissa: https://link.springer.com/article/10.1007/s10115-025-02429-y Luettu: 28.4.2025.

Mucci, T. 21.10.2024. The history of AI. Luettavissa: https://www.ibm.com/think/topics/history-of-artificial-intelligence Luettu: 7.4.2025.

National Cyber Security Centre. 21.1.2019. What is an antivirus product? Do I need one? Luettavissa: https://www.ncsc.gov.uk/guidance/what-is-an-antivirus-product Luettu: 1.4.2025.

Numminen, L. 19.10.2023. Mikä on Turingin testi eli Turing-koe? Luettavissa: https://www.finnishup.com/mika-on-turing-testi/ Luettu 18.3.1025.

Reed, J. 30.1.2025. When ransomware kills: Attacks on healthcare facilities. Luettavissa: https://www.ibm.com/think/insights/when-ransomware-kills-attacks-on-healthcare-facilities Luettu: 20.4.2025.

Santos, O., Salam, S. & Dahir, H. 2024. The AI Revolution in Networking, Cybersecurity, and Emerging Technologies. Addison-Wesley Professional. Boston (USA). E-kirja. Luettavissa: https://learning.oreilly.com/library/view/the-ai-revolution/9780138293703/ Luettu: 1.5.2025.

Stryker, C. & Scapicchio, M. 22.3.2024. What is generative AI? Luettavissa: https://www.ibm.com/think/topics/generative-ai Luettu: 14.3.2025.

Stryker, C. & Kavlakoglu, E. 9.8.2024 What is artificial intelligence (AI)? Luettavissa: https://www.ibm.com/think/topics/artificial-intelligence Luettu: 14.3.2025.

Zeadally, S., Adi, E., Baig, Z. & Khan, I, A. 2020. Harnessing Artificial Intelligence Capabilities to Improve Cybersecurity. IEEE Access, 8. Luettavissa: https://ieeexplore.ieee.org/abstract/document/8963730 Luettu: 1.5.2025.