TURUN YLIOPISTO
Kauppakorkeakoulu

Tekoälyn hyödyntäminen yritysten ennaltaehkäisevän kyberturvallisuuden vahvistamisessa

Tietojärjestelmätieteen kandidaatintutkielma

Laatija:
Nikolas Luojus

Ohjaaja:
FT Samuli Laato

17.5.2025
Turku

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.

Kandidaatintutkielma

Oppiaine: Tietojärjestelmätiede
Tekijä(t): Nikolas Luojus
Otsikko: Tekoälyn hyödyntäminen yritysten ennaltaehkäisevässä kyberturvallisuudessa
Ohjaaja(t): FT Samuli Laato
Sivumäärä: 66 sivua
Päivämäärä: 17.5.2025

Kyberrikollisuuden kasvu ja hyökkäysten yhä monimutkaisemmat muodot pakottavat yritykset etsimään innovatiivisia ratkaisuja tietoturvahaasteisiin. Tekoäly tarjoaa tehokkaita keinoja erityisesti ennaltaehkäisevän suojauksen toteuttamiseen, mikä auttaa vähentämään kyberrikollisuudesta aiheutuvia taloudellisia tappioita, jotka ovat jatkuvassa kasvussa.

Tämä kandidaatintutkielma keskittyy tutkimaan, kuinka tekoälypohjaisia syväoppimis- ja koneoppimismenetelmiin pohjautuvia ratkaisuja voidaan hyödyntää yritysten ennaltaehkäisevässä kyberturvallisuudessa. Tutkielma on kirjallisuuskatsaus, jossa tarkastellaan tekoälyyn ja kyberturvallisuuteen liittyviä tieteellisiä artikkeleita. Se analysoi erilaisia kyberuhkia ja -hyökkäyksiä sekä tarkastelee koneoppimisen ja syväoppimisen tekoälytekniikoita ja niiden sovelluksia. Tutkimuksen päähuomio on näiden teknologioiden hyödyntämisessä ennaltaehkäisevän kyberturvallisuuden parantamiseksi.

Tutkielma selvittää tekoälyn toiminnallisuuksia ja sen tuomia etuja ennaltaehkäisevälle kyberturvallisuudelle, mutta tuo myös esille mahdolliset haitat ja rajoitteet. Lisäksi tarkastellaan tulevaisuuden tekoälypohjaisia kyberturvallisuusratkaisuja.

Tutkielmassa toteutetun synteesin perusteella tekoälypohjaiset ratkaisut voivat merkittävästi parantaa yritysten kyberturvallisuutta ja tehostaa kyberuhkien ennaltaehkäisyä. Koneoppimisen ja syväoppimisen menetelmät tarjoavat kehittyneitä työkaluja arkaluontoisen tiedon ja toimintojen suojaamiseen. Lisäksi näihin menetelmiin pohjautuvat kyberturvallisuusratkaisut voivat nopeuttaa kyberturvallisuusprosesseja ja lisätä niiden tarkkuutta ja luotettavuutta.

Avainsanat: kyberturvallisuus, koneoppiminen, syväoppiminen, verkkourkinta, haittaohjelmahyökkäys, palvelunestohyökkäys

SISÄLLYS

KUVIOT

TAULUKOT

1 Johdanto

Generatiivisen tekoälyn innovaatiot ovat nopeuttaneet merkittävästi koneoppimisen ja syväoppimisen kehitystä (Google Cloud, 2024). Se avaa uusia mahdollisuuksia sekä kyberhyökkäysten torjuntaan että niiden toteuttamiseen. Tässä tutkielmassa tekoälyllä tarkoitetaan erityisesti koneoppimisen ja syväoppimisen menetelmiä. Kyberhyökkäyksistä tulee entistä tehokkaampia, automatisoidumpia ja kohdennetumpia. Lähitulevaisuudessa tekoälyn nopea kehitys todennäköisesti luo parempia mahdollisuuksia hyökkäysten automatisointiin sekä käyttäjämanipulointiin ja tiedonkeruuseen. Ajan myötä tekoälyteknologioiden sekä taitojen ja työkalujen saatavuus tulee helpottumaan, mikä osaltaan kannustaa hyökkääjiä lisäämään tekoälyn käyttöä osana kyberhyökkäyksiä (Huoltovarmuuskeskus, 2022). Vuoden 2015 Sony Inc. -hakkerointi on esimerkki onnistuneesta kyberhyökkäyksestä verkkourkinnan muodossa, ja se osoittaa, kuinka laajaa vahinkoa tällainen hyökkäys voi aiheuttaa. Kyseisessä hyökkäyksessä Pohjois-Koreaa syytettiin Sony Pictures Entertainmentiin kohdistetun kyberhyökkäyksen toteuttamisesta. Hyökkäys oli osa laajempaa kampanjaa vahingoittaa Yhdysvaltain taloutta ja sisälsi muita kyberrikoksia, kuten kiristyshaittaohjelmahyökkäyksen sekä yrityksiä varastaa rahaa kansainvälisiltä finanssilaitoksilta. Pelkästään tässä tapauksessa vahingot olivat New York Timesin mukaan satojen miljoonien Yhdysvaltain dollarien luokkaa. (Sanger & Benner, 2018.) Jo vuonna 2013 Wall Street Journal arvioi koko kyberrikollisuuden kokonaiskustannusten olevan vuositasolla Yhdysvalloissa noin 100 miljardia Yhdysvaltain dollaria (Gorman, 2013). Vaikka tilastot osoittavat kyberturvallisuuden haasteiden kasvavan, organisaatiot voivat nyt hyödyntää tekoälyä uhkien tunnistuksessa monin eri tavoin käyttäytymismallien analysoinnista uhkakuvien kartoitukseen (Skwarczek, 2023; Cybersecurity Ventures, 2022).

Kyberhyökkäys on yleensä pahantahtoinen ja suunnitelmallinen yritys, jossa yksittäinen henkilö tai organisaatio pyrkii murtautumaan toisen henkilön tai organisaation tietojärjestelmään. Tavallisimpia kyberhyökkäyksen muotoja ovat haittaohjelmat, kiristysohjelmat, palvelunestohyökkäykset, tietojenkalastelu, sosiaalinen manipulointi sekä SQL-injektiohyökkäykset (Sarker ym., 2020). Nykyajan kyberturvallisuuden monimuotoisten ongelmien ratkaisemiseksi voidaan hyödyntää suosittuja tekoälymenetelmiä, kuten koneoppimista ja syväoppimista. Näiden lisäksi merkittäviä tekniikoita ovat luonnollisen kielen käsittely, tiedon esitys ja päättely sekä sääntöpohjaisten asiantuntijajärjestelmien mallinnus. Näitä tekniikoita voidaan soveltaa muun muassa haitallisten toimintojen tunnistamiseen, petosten havaitsemiseen, kyberhyökkäysten

ennustamiseen, pääsynhallinnan hallintaan sekä kyberanomalian tai tunkeutumisten havaitsemiseen. (Sarker ym., 2021.)

Samalla kun tekoälyn hyödyntäminen kyberrikollisuudessa on yleistynyt, yritykset ovat alkaneet ottaa tekoälyä myös omaan käyttöönsä kyberturvallisuutensa vahvistamiseksi (Tao ym., 2021). Tällä hetkellä tekoälyyn perustuvat sovellukset ovat erityisen edistyksellisiä uhkien tunnistamisen ja havaitsemiseen, mikä tekee niistä keskeisiä työkaluja kyberhyökkäysten torjunnassa (Marchal, Nawrotek & WithSecure, 2024).

Kyberhyökkäykset monimutkaistuvat ja niiden kehittämiseen löytyy motivaatiota suuren potentiaalisen rahallisen hyödyn takia. Rikolliset toimivat yhä organisoidummin ja hyödyntävät työnjakoa, jossa eri toimijat vastaavat esimerkiksi haittaohjelmien kehittämisestä, jakelusta ja rahanpesusta. Hyökkäyksistä on tullut kohdennetumpia, ja niiden toteuttamisessa hyödynnetään taustatutkimusta sekä esimerkiksi yksilöllisesti räätälöityjä kalasteluviestejä. Lisäksi hyökkäyksiä tehostetaan automatisoinnilla ja koneoppimisella, mikä mahdollistaa kyberpuolustuksen kiertämisen. (Peersman, Williams, Edwards, & Rashid, 2022.) Tästä syystä hyökkäysten torjuminen on muuttunut yhä haastavammaksi. Tekoäly tarjoaa kuitenkin uusia lupaavia ratkaisuja, kuten poikkeavuuksien reaaliaikaisen havaitsemisen, uhkien ennakoinnin ja automaattisen reagoinnin. Tekoäly kykenee myös oppimaan jatkuvasti uusista uhkista, mikä tehostaa puolustautumisen sopeutumista jatkuvasti kehittyvään toimintaympäristöön. Tekoälypohjaiset menetelmät voivat siis merkittävästi vahvistaa yritysten kyberturvallisuutta. (Khan ym., 2024.) Tämän lisäksi organisaatioiden on omaksuttava ennaltaehkäiseviä toimenpiteitä, kuten jatkuva valvonta ja riskienhallinta, vastatakseen kehittyviin uhkiin tehokkaasti (Narayanan & Venkatraman, 2024). Tuore raportti osoittaa, että tekoälyn hyödyntäminen voi kuitenkin jopa kaksinkertaistaa organisaatioiden kyvyn vastustaa hyökkäyksiä ja vähentää hyökkäysten aiheuttamia kustannuksia 20 % (The Sunday Times, 2024).

Tekoälyn soveltaminen kyberturvallisuudessa on kasvanut merkittävästi viime vuosina, ja tutkimusaktiviteetti on noussut alalle yhä suuremmaksi. Vuosien 2004–2023 välillä julkaistiin yli 9 352 tutkimusta, jotka käsittelivät tekoälyn soveltamista kyberturvallisuuteen ja yksityisyyden suojaukseen. Näistä tutkimuksista noin 13 % keskittyi tunkeutumisen havaitsemiseen tekoälyn avulla, ja 10 % käsitteli haittaohjelmien luokittelua koneoppimisen menetelmillä. Lisäksi tekoälyä hyödynnettiin IoT-turvallisuudessa, DDoS-hyökkäysten torjunnassa sekä yksityisyyden suojaamisessa hajautetun oppimisen avulla (Achuthan ym., 2023). Tämä osoittaa tekoälyn merkityksen kyberturvallisuuden kehittämisessä ja sen potentiaalin vastata kehittyviin uhkiin.

Markkinat tekoälyn käytölle kyberturvallisuudessa ovat myös laajentuneet. Vuonna 2022 markkinoiden arvo oli 19,2 miljardia dollaria ja sen ennustetaan kasvavan 154,8 miljardiin dollariin vuoteen 2032 mennessä (PR Newswire, 2023). Tekoälyn rooli kyberturvallisuudessa on siis kasvamassa ja tutkimusta tehdään huomattavasti.

1.1 Tutkielman tarkoitus ja tutkimuskysymykset

Tämä kandidaatintutkielma tarkastelee tekoälyn roolia ennaltaehkäisevässä kyberturvallisuudessa ja erityisesti sen soveltamista jatkuvassa valvonnassa. Aihe on tieteellisessä tutkimuksessa suhteellisen uusi, ja ala kehittyy jatkuvasti, mistä syystä lisätutkimukselle nähdään olevan tarvetta. Erityisesti uhkia ja niihin reagoimista on tutkittu jo hyvin laajasti. Ennaltaehkäisyyn liittyvät kyberturvallisuusratkaisut ovat kokeneet suuria edistysaskelia viime aikoina juuri koneoppimis- ja syväoppimistekniikoihin perustuvien mallien kehittymisen myötä. Kirjallisuuskatsauksessa pyritään tuomaan esiin näitä uusimpia kehityskulkuja tarjoamalla siihen systemaattisen tavan kartoittaa olemassa olevaa tutkimusta. Lisäksi yritetään löytää ristiriitaisuuksia aiemmassa tutkimuksessa, mikä voi toimia perustana uusille tutkimusaloitteille. Yksi keskeinen haaste tekoälyn hyödyntämisessä yritysten ennaltaehkäisevässä kyberturvallisuudessa on ollut esimerkiksi väärien positiivisten hälytysten suuri määrä, joita voidaan kuitenkin vähentää hyödyntämällä syväoppimiseen pohjautuvia menetelmiä (Al Jallad, Aljnidi, & Desouki, 2022). Tutkielman tutkimuskysymykset ovat seuraavat:

1. Millaisia kyberuhkia yritykset kohtaavat ja miten niiltä voidaan suojautua ennaltaehkäisevästi?

2. Miten yritykset voivat hyödyntää syväoppimisen ja koneoppimisen menetelmiin pohjautuvia tekniikoita ennaltaehkäisevässä kyberturvallisuudessa?

3. Millainen tulevaisuus tekoälyllä on yritysten ennaltaehkäisevässä kyberturvallisuudessa?

Tutkielmassa aihetta lähestytään yrityksen näkökulmasta. Kyberuhkia on paljon erilaisia, mutta eritoten keskitytään palvelunestohyökkäyksiin, verkkourkintaan ja haittaohjelmiin niiden yleisyyden takia (Falowo ym., 2024; Basit ym., 2020). Perinteisistä kyberpuolustuskeinoista tuodaan esiin muun muassa palomuurit, virustorjuntaohjelmistot ja hyökkäyksen havaitsemisjärjestelmät, koska ne tarjoavat monikerroksisen suojan erilaisia uhkia vastaan. Nämä teknologiat täydentävät toisiaan ja muodostavat yhdessä tehokkaan puolustusstrategian. (Roopesh., 2024.) Näihin ratkaisuihin voidaan tuoda kokonaisvaltaisesti osaksi koneoppimisen ja syväoppimisen menetelmiä tietyille osa-alueille, joista tutkielmassa tarkastellaan haitallisen

liikenteen tunnistamista, käyttäytymisen poikkeavuuksien analyysiä reaaliaikaisessa seurannassa sekä uhkamallinnusta ja ennakoivaa riskianalyysiä.

Tutkielman tavoitteena on tuoda esiin tekoälyn merkitys kyberturvallisuudessa, tarkastella sen käyttömahdollisuuksia ja rajoituksia sekä pohtia sen vaikutuksia ihmisen roolin kehittymiseen kyberturvallisuuden toteuttamisessa. Tutkielma tarjoaa arvokasta tietoa yrityksille, jotka pyrkivät vahvistamaan kyberturvallisuuttaan tekoälyteknologioiden avulla. Se avaa tekoälyn ja eritoten kone- ja syväoppimismenetelmien mahdollisuuksia ennaltaehkäisevän kyberturvallisuuden saralla, kuten haitallisen verkkoliikenteen tunnistamisessa, käyttäjäanalyysissä ja uhkamallinnuksessa. Näin se auttaa organisaatioita ymmärtämään, miten tekoäly voi toimia keskeisenä työkaluna nykyaikaisissa turvallisuusratkaisuissa niin organisaation sisäisten kuin ulkoisten uhkien torjunnassa. Lisäksi tällä jaottelulla voidaan tarkastella kone- ja syväoppimismenetelmiin pohjautuvien tekoälyteknologioiden ja -järjestelmien mahdollisuuksia jatkuvan valvonnan välineenä niin organisaation, kuin myös yksilön tasolla. Jaottelun lisäksi tuodaan esille jo markkinoilla olevia ennaltaehkäisevään kyberturvallisuuteen liittyviä järjestelmiä, kuten Amazon Macie. Tutkielma tarjoaa myös akateemiselle yhteisölle systemaattisen katsauksen tekoälyratkaisujen nykytilasta, haasteista ja tulevista kehityssuunnista.

Tutkielma käsittelee myös keskeisiä haasteita, kuten laskennalliset resurssivaatimukset ja väärien hälytysten minimointi, tarjoten näin ratkaisuja yritysten kohtaamiin käytännön ongelmiin. Kokonaisuutena tutkielma toimii pohjana tuleville tutkimusaloitteille sekä tarjoaa konkreettisia suosituksia yrityksille, jotka haluavat integroida tekoälyratkaisut osaksi kyberturvallisuusstrategioitaan.

Tutkielmassa luodaan katse tulevaisuuden näkymiin tekoälyn hyödyntämisessä yritysten ennaltaehkäisevässä kyberturvallisuudessa. Erityisesti tarkastellaan, miten nopeasti kehittyvät tekoäly- ja koneoppimisratkaisut voivat muuttaa yritysten kyberturvallisuusstrategioita entistä ennakoivammiksi ja dynaamisemmiksi. Lisäksi pohditaan tulevia painopistealueita, kuten jatkuvaa mallien kehittämistä, adaptiivisia järjestelmiä, automaattisen uhkatiedon vaihdon merkitystä, lainsäädännön kehitystä, suurten kielimallien hyödyntämistä sekä ihmisen ja tekoälyn yhteistyötä. Tutkielmassa arvioidaan, millaisia haasteita ja mahdollisuuksia uudet teknologiat tuovat yritysten riskienhallintaan ja tietoturvaan, sekä miten organisaatiot voivat varautua tekoälyn lisääntyvään käyttöön kyberturvallisuuden kentällä.

1.2 Tutkielman rakenne ja rajaukset

Perinteinen lähestymistapa, jossa keskitytään yksittäisten uhkien kuten verkkourkinnan, haittaohjelmien tai palvelunestohyökkäysten ehkäisyyn, tarjoaa tärkeää käytännön tietoa, mutta jää usein ilmiöpohjaiseksi eikä tuo esiin tekoälyn toimintalogiikkaa eri konteksteissa. Sen sijaan tässä tutkielmassa käytetään jäsentelyä, joka tarkastelee tekoälyn tapoja havaita poikkeavuuksia verkon normaalissa käyttäytymisessä, analysoida uhkakuvia ennakoivasti ja tunnistaa laajamittaista haitallista liikennettä. Jäsentely mahdollistaa tekoälyn roolin ja arvon ymmärtämisen kyberturvallisuuden eri tasoilla kokonaisvaltaisesti. Näkökulma ei ainoastaan kuvaa, mitä tekoäly voi tehdä, vaan selittää, miksi ja miten se toimii, mitä dataa se hyödyntää ja millaisia riskejä siihen liittyy. Lisäksi se avaa mahdollisuuden kriittiseen arvioon siitä, missä konteksteissa tekoäly on tehokas ja missä sen rajat tulevat vastaan. Näin ollen lähestymistapa tukee myös liiketoimintanäkökulmaa ja tarjoaa rakenteen, jonka avulla yritykset voivat tarkastella tekoälyyn perustuvien suojausratkaisujen soveltuvuutta omassa riskienhallinnassaan.

Tutkielmassa ei syvennytä julkisen sektorin näkökulmiin eikä kuluttajatasoiseen kyberturvallisuuteen, elleivät ne palvele suoraan yritysten ennaltaehkäisevää näkökulmaa tekoälyn hyödyntämisessä. Lisäksi tutkielmassa ei paneuduta teknisten algoritmien syvälliseen matemaattiseen esitykseen. Tiettyihin ajankohtaisiin kyberturvallisuusuhkiin tai kyberpolitiikkaan liittyvä keskustelu sivuutetaan, mikäli se ei liity tutkimuskysymyksissä määriteltyihin näkökulmiin. Näin rajattu lähestymistapa mahdollistaa keskittymisen tekoälyn konkreettiseen arvoon modernissa yritysturvallisuudessa.

Tutkielma etenee seuraavasti: toinen luku käsittelee ensimmäistä tutkimuskysymystä, joka koskee yritysten kohtaamia kyberuhkia ja niiltä suojautumista ennaltaehkäisevästi. Luvussa tarkastellaan, mitä kyberturvallisuus ja kyberhyökkäykset ovat sekä miten organisaatiot voivat reagoida ennaltaehkäisevästi kyberuhkiin. Kolmas luku keskittyy kolmanteen tutkimuskysymykseen, joka käsittelee koneoppimisen ja syväoppimisen käsitteitä. Luvussa selvitetään, kuinka ja miksi näitä tekoälytekniikoita hyödynnetään ennaltaehkäisevässä kyberturvallisuudessa sekä millaisia uhkia tekoälyn käyttöön liittyy. Neljännessä luvussa pohditaan tekoälyn tulevaisuutta ennaltaehkäisevässä kyberturvallisuudessa. Viides luku sisältää yhteenvedon, jossa vastataan tutkielman ensimmäiseen ja päätutkimuskysymykseen: miten yritykset voivat hyödyntää tekoälyä ennaltaehkäisevässä kyberturvallisuudessa. Tässä luvussa esitetään myös työn keskeiset johtopäätökset.

2 Yritysten kohtaamat keskeiset kyberuhat ja niiltä suojautuminen ennaltaehkäisevästi

2.1 Kyberturvallisuuden määritelmä

Tässä tutkielmassa kyberturvallisuudelle käytetään Mijwilin ja Aljanabin (2023) määritelmää, jonka mukaan kyberturvallisuus koostuu tekniikoista ja lähestymistavoista, joiden avulla suojataan tietojärjestelmät ja data kyberhyökkäyksiltä sekä estetään haittaohjelmien pääsy järjestelmien hallintaan. Kyberturvallisuus on laaja käsite, jossa huomioidaan niin tietoturva kuin myös hyökkäyksistä toipuminen sekä loppukäyttäjien koulutus (Martínez Torres ym., 2019). Sen tarkasta määrittelystä on kuitenkin olemassa erilaisia hieman toisistaankin poikkeavia näkemyksiä (Craigen ym., 2014). Kyberturvallisuuden ydin on järjestelmien ja datan suojaaminen haavoittuvuuksilta ja haittaohjelmien pääsyltä järjestelmään sekä kyberrikollisuuden torjuntaan ja erinomaisen sähköisen toimintaympäristön luomiseen. Lisäksi tiedot on suojattava varkaudelta, vahingoittamiselta ja luvattomalta käytöltä sekä erilaisilta luonnonilmiöiltä, kuten pölyltä ja kosteudelta. (Mijwil ja Aljanabi., 2023).

Kyberturvallisuuden käsitteen määrittely tuottaa tietyn haasteen myös siksi, että se sekoitetaan usein muiden aiheeseen liittyvien käsitteiden kuten tietoturvallisuuden kanssa (Fischer, 2016). Molemmat käsitteet sisältävät myös samankaltaisuuksia, mutta kyberturvallisuutta pidetään tietoturvallisuuteen nähden huomattavasti laajempana käsitteenä, sillä se ottaa huomioon myös inhimilliset tekijät ja ihmisen mahdollisena uhkatekijänä turvallisuudelle (von Solms & van Niekerk, 2013). Kyberavaruus on keskeinen kyberturvallisuuteen liittyvä käsite, jolla tarkoitetaan erilaisten laitteiden ja niiden välisten yhteyksien muodostamaa virtuaalista ulottuvuutta. Sen suojaamiseen on käytettävissä monia eri keinoja, kuten salasanojen käyttö, palomuurit ja perusosaaminen turvallisesta internetin käytöstä.

International Organization for Standardization (ISO):n määritelmän mukaan kyberturvallisuus keskittyy tiedon luottamuksellisuuden, yhtenäisyyden ja saavutettavuuden varmistamiseen kyberavaruudessa. Kyberturvallisuus linkittyy ISO:n mukaan viiteen keskeiseen käsitteeseen, joita ovat tietoturvallisuus, verkkosuojaus, internet-turvallisuus sekä kriittisen tietorakenteiden turvaaminen ja kyberrikollisuus. (ISO, 2012). Tietoturva pyrkii varmistamaan tiedon luottamuksellisuuden, yhtenäisyyden ja saatavuuden, kun taas verkkosuojaus liittyy verkkojen suunnitteluun, toteutukseen ja hallintaan tietoturvallisuuden saavuttamiseksi. Internet-turva kohdistuu internetin palveluiden ja järjestelmien suojaamiseen sekä niiden saatavuuden

varmistamiseen. Kriittisen tietoinfrastruktuurin suojaus puolestaan tarkoittaa toimenpiteitä, joilla varmistetaan, että tärkeät järjestelmät ovat suojattuja tieto-, verkko-, ja kyberturvariskeiltä. Tehokas kyberturvallisuus edellyttää, että kaikki nämä osa-alueet toimivat saumattomasti yhdessä. Teknologian ja prosessien on tässä kehikossa tuettava toisiaan, jotta kyberhyökkäyksiltä voidaan suojautua mahdollisimman tehokkaasti (Cisco, 2024a).

2.2 Kyberrikollisuus ja erilaiset kyberuhat

Kyberrikollisuus on rikollista toimintaa, jossa kyberavaruuden palveluita tai sovellutuksia käytetään rikoksen tekemiseen tai ne toimivat rikoksen kohteena, lähteenä tai paikkana (International Organization for Standardization, 2012). Tyypillisesti kyberhyökkäys on digitaalinen rikollisessa mielessä tehty teko, jonka tavoitteena on varastaa, vahingoittaa tai tunkeutua organisaation suojattuun dataan (Basit ym., 2021). Hyökkäysten motiivit voivat liittyä taloudelliseen hyötyyn, poliittisten päämäärien saavuttamiseen tai tietojen kalasteluun ja tuhoamiseen (Mijwil ym., 2023; AL-Hawamleh, 2023). Kyberhyökkäykset hyödyntävät sähköisten järjestelmien ja verkkojen haavoittuvuuksia (Mijwil ym., 2023). Niiden määrä kasvaa jatkuvasti samalla kun niiden tehokkuus ja monimutkaisuus lisääntyvät. Kyberhyökkäyksissä hyödynnetään aukkoja sähköisten järjestelmien tietoturvassa (Mijwil ym., 2023). Ne ovat nopeasti kasvava rikosten muoto ja niiden kompleksisuus ja tehokkuus aiheuttavat jo miljoonien kustannuksia (Abdullah & Mohd, 2019).

Kyberhyökkäysten taloudelliset vaikutukset voivat olla merkittäviä. Ne voivat hidastaa yrityksen toimintaa ja jopa lamaannuttaa sen kokonaan. Samalla ne voivat vahingoittaa yrityksen asiakkaita, mikä taas heikentää yrityksen luotettavuutta. Erilaisten kyberhyökkäysten toteutusta ymmärretään varsin puutteellisesti, ja suojautuminen niitä vastaan vaatii yleensä kattavaa asiantuntemusta ja perehtymistä aiheeseen (AL-Hawamleh, 2023). Jotta kyberhyökkäyksiltä voidaan suojautua tehokkaasti, pitääkin niitä pystyä luokittelemaan ja ymmärtämään paremmin. Monet yritykset ja organisaatiot ovat kohdanneet kyberhyökkäyksiä ja toisinaan myös joutunut niiden uhriksi. Ciscon entisen toimitusjohtajan mukaan yritykset voidaankin jakaa kahteen ryhmään: niihin, jotka ovat jo joutuneet hyökkäyksen kohteeksi, ja niihin, jotka eivät vielä tiedä olevansa hyökkäyksen kohteena (Cisco, 2018).

Kyberuhkia on useita erilaisia, mutta tässä työssä keskitytään kolmeen erityisen merkittävään: verkkourkintaan, palvelunestohyökkäyksiin ja haittaohjelmiin. Näiden uhkien valinta perustuu kolmeen pääperusteeseen: esiintyvyyteen, vaikutuspotentiaaliin ja monipuolisuuteen uhkatyyppien näkökulmasta. Verkkourkinta edustaa sosiaalista manipulointia, palvelunestohyökkäykset kohdistuvat järjestelmien saatavuuteen, ja haittaohjelmat vaikuttavat suoraan järjestelmien tai

tiedostojen eheyttä ja toimivuutta vastaan. Näin valittu kolmijako tarjoaa kattavan yleiskuvan keskeisistä kyberturvallisuusuhista, jotka koskettavat laajasti sekä yksityishenkilöitä että organisaatioita.

Vaikka tällaista kolmijaottelua ei aina esitetä muodollisesti akateemisessa kirjallisuudessa, juuri nämä kolme uhkaa esiintyvät toistuvasti keskeisissä kansainvälisissä kyberturvallisuusraporteissa ja tutkimuksissa. Esimerkiksi ENISA:n (2023) ja IBM:n (2023) julkaisuissa nämä uhkamuodot nousevat esiin organisaatioiden merkittävimpiä riskejä kartoitettaessa. Myös (Sarker ym., 2020) listaavat verkkourkinnan, haittaohjelmat sekä palvelunestohyökkäykset yleisiksi uhiksi. Lisäksi niiden välinen yhteys on olennainen: verkkourkinta voi toimia porttina haittaohjelmien asentamiselle, ja haittaohjelmat puolestaan mahdollistavat esimerkiksi palvelunestohyökkäysten toteuttamisen bottiverkkojen eli haittaohjelmalla kaapattujen tietokoneiden tai muiden internetiin kytkettyjen laitteiden avulla. Näin ollen valittu jäsentely ei ainoastaan kata kolme keskeistä uhkaa, vaan myös tuo esiin niiden keskinäisiä suhteita ja yhteisvaikutuksia.

2.2.1 Verkkourkinta

Verkkourkinnan (engl. Phishing) tavoitteena on varastaa käyttäjien luottamuksellisia ja salassa pidettäviä tietoja verkossa. Kohdistetussa verkkourkinnassa pyritään saamaan haltuun luottamuksellisia tietoja joltain erityiseltä kohteelta, kuten yritykseltä. Kuusi yleisintä verkkourkinnan kohdetta ovat finanssilaitokset, sähköpostipalvelut, sosiaalinen media, kuljetuspalvelut, maksupalvelut ja verkkokaupat, ja nämä kattavat noin 76,4 % kaikista verkkourkintahyökkäyksistä (Petrosyan, 2022). Verkkourkinta tapahtuu usein sähköpostin välityksellä, ja se perustuu kohdehenkilön manipulaatioon (Abdullah & Mohd, 2019). Kohdistetussa verkkourkinnassa viestit ovat henkilökohtaisempia, mikä tekee niistä uskottavampia ja siten vaikeammin tunnistettavia. Viesti voi sisältää linkin, joka ohjaa käyttäjän haitalliselle sivustolle. Tämä muistuttaa usein oikeaa verkkosivua mutta tallentaa syötetyt tiedot hyökkääjän käyttöön. Järjestelmään pääseminen antaa hyökkääjälle mahdollisuuden joko myydä saatuja tietoja tai hyödyntää niitä muulla tavoin, erityisesti jos ne sisältävät kohdehenkilön luottokorttitietoja (Banday & Qadri, 2011.)

Verkkourkinta on kasvava huolenaihe, koska monet internetin käyttäjät yrityksistä aina yksityishenkilöihin sortuvat niihin. Kyse on ennen kaikkea sosiaalisesta manipulaatiosta, jossa hyökkääjä pyrkii saamaan uhrinsa luovuttamaan arkaluontoisia tietoja. Tietojenkalastelijat käyttävät useimmiten hyväkseen ihmisten henkisiä haavoittuvuuksia sen sijaan että turvautuisivat kehittyneisiin teknisiin menetelmiin. Suurin osa tietojenkalasteluhyökkäyksistä alkaa

sähköpostiviestillä, jossa hyödynnetään laittomasti luotettavia yrityksiä uhrin luottamuksen saavuttamiseksi. Viestiin upotettu linkki ohjaa uhrin verkkosivustolle, jossa tietojen kalastelu tapahtuu (Alkhalil ym., 2021). Kalastelijat voi myös lisätä sähköpostiin tiedoston, joka asentaa haittaohjelman uhrin tietokoneelle ja voi näin varastaa esimerkiksi kirjautumistietoja tai maksutietoja (AL-Hawamleh, 2023).

2.2.2 Palvelunestohyökkäykset

Palvelunestohyökkäykset alkavat yleensä yksittäisestä lähteestä, mutta myöhemmin niitä on alettu toteuttaa hajautetusti (Kaur Chahal ym., 2019). Hajautetut palvelunestohyökkäykset (engl. Distributed Denial of Service, DDoS) ovat tällä hetkellä yleisimpiä ja monimutkaisimpia uhkia organisaatioille, ja niiden estäminen on yhä vaikeampaa (Sahoo ym., 2019; Zargar ym., 2013; Conti & Gangwal, 2017). Esimerkiksi vuonna 2018 GitHub joutui yhden historian suurimman DDoS-hyökkäyksen kohteeksi (Kottler, 2018). Hyökkääjät käyttävät tuhansia päätelaitteita, koneita ja bottiverkkoja käynnistääkseen samanaikaisesti DDoS-hyökkäyksiä, jotka lopulta kuluttavat kohdejärjestelmän pääresurssit tehden koko palveluista käyttökelvottomia. DDoS-hyökkäyksessä kohteena olevan yrityksen palvelimet kuormitetaan valtavalla määrällä liikennettä, jolloin ne ylikuormittuvat ja kaatuvat. Tästä seurauksena myöskään hyväksyttyjä prosesseja ei pystytä käsittelemään (Saravanan & Bama, 2019). Vuonna 2019 jopa 24 % yrityksistä raportoi joutuneensa DDoS-hyökkäyksen kohteeksi (Saravanan & Bama, 2019). Hyökkäyksen torjunta on haastavaa, koska hyökkäykset voivat tulla useista eri IP-osoitteista eri puolilta maailmaa, mikä jo omalta osaltaan vaikeuttaa niiden alkuperän selvittämistä (AL-Hawamleh, 2023).

2.2.3 Haittaohjelmat

Haittaohjelmat ovat tietokoneohjelmia, jotka asennetaan käyttäjän tietokoneelle ilman heidän suostumustaan. Niiden tarkoituksena voi olla tietojen varastaminen, tietokoneen vahingoittaminen tai käyttöoikeuksien kaappaaminen (Jawhar, 2023). Haittaohjelmat voidaan luokitella eri kategorioihin, joista yksi tapa on jakaa ne staattisiin eli ensimmäisen sukupolven ja dynaamisiin eli toisen sukupolven haittaohjelmiin infektiostrategian perusteella. Ensimmäisen sukupolven haittaohjelmat säilyvät muuttumattomina tartunnan jälkeen, kun taas toisen sukupolven haittaohjelmat mukautuvat ja muuttavat muotoaan jokaisen tartunnan yhteydessä (Alenezi ym., 2020).

Haittaohjelmat voidaan jakaa myös niiden tarkoituksen ja leviämistavan mukaan. Takaportti (engl. Backdoor) on ohjelma, joka kiertää tietoturvamekanismit ja antaa hyökkääjälle pääsyn

järjestelmään. Botit suorittavat automaattisesti erilaisia toimintoja, kuten levittävät muita haittaohjelmia tai suorittavat palvelunestohyökkäyksiä. Kiristyshaittaohjelmat (engl. Ransomware) lukitsevat käyttäjän tiedostot tai järjestelmän ja vaativat maksua niiden vapauttamisesta. Troijalainen on haittaohjelma, joka naamioituu harmittomaksi ohjelmaksi ja huijaa käyttäjän asentamaan sen järjestelmään. Virus on haittaohjelma, joka siirtyy laitteesta toiseen itsestään. Mato on myös eräänlainen virus, joka hyödyntää käyttöjärjestelmän haavoittuvuuksia levitäkseen. Suurin ero matojen ja virusten välillä on se, että madot pystyvät itsenäisesti lisääntymään ja leviämään, kun taas virukset vaativat leviämiseen ihmisen toimintaa. (Gibert ym., 2020).

Yleisimmin haittaohjelmat leviävät linkkien kautta, jotka lataavat haitallisen tiedoston käyttäjän tietokoneelle. Ne voivat levitä myös fyysisesti esimerkiksi USB-muistitikun avulla. Haittaohjelmahyökkäykset yleistyvät merkittävästi ja rikollisille ne ovat erityisen kiinnostavia niiden taloudellisen potentiaalin vuoksi. Haittaohjelmat ovat monimutkaistuneet ja tuottavat rikollisille miljardien dollarien voittoja, minkä vuoksi niiden torjunta on entistä haasteellisempaa (Alenezi ym., 2020).

2.3 Kyberuhilta suojautuminen ennaltaehkäisevästi

Kyberturvallisuusuhkien ennaltaehkäisy ja havaitseminen vaativat huolellisesti suunniteltujen strategioiden ja tehokkaiden työkalujen käyttöönottoa. Näiden avulla voidaan suojata organisaation järjestelmät vaaroilta, jotka uhkaavat järjestelmien luottamuksellisuutta, eheyttä ja saatavuutta. (Marchal ym., 2024.) Kyberturvallisuus voidaan jäsentää kolmeen keskeiseen käsitteeseen, joiden avulla ennaltaehkäisevänkin kyberturvallisuuden tavoitteita voi hahmottaa. Niitä ovat estää, tunnistaa ja vastata (engl. Protect, Identify, Respond). Tämä joukko kattaa toimenpiteet, joilla pyritään suojaamaan organisaation järjestelmiä, tunnistamaan mahdolliset hyökkäykset ja reagoimaan niihin tehokkaasti. (Bayuk ym. 2012.)

Perinteisesti turvallisuuden päämääränä on ollut estää vastustajan hyökkäykset ennen kuin ne onnistuvat. Alan asiantuntijat kuitenkin myöntävät, ettei kaikkia hyökkäyksiä voida täysin estää. Tämän vuoksi turvallisuussuunnitteluun on sisällytettävä tehokkaat menetelmät käynnissä olevien hyökkäysten havaitsemiseen, mieluiten ennen kuin ne aiheuttavat vahinkoa. Mikäli hyökkäys tunnistetaan ja järjestelmä joutuu kohteeksi, sen on tärkeää kyetä reagoimaan ja torjumaan hyökkäyksen aiheuttamat haitat. (Bayuk ym. 2012.)

Yritysten ennaltaehkäisevä kyberturva perustuu monikerroksiseen lähestymistapaan, jossa yhdistyvät teknologiset ratkaisut, prosessien optimointi ja henkilöstön kouluttaminen.

Ennaltaehkäisevä strategia on välttämätön nykyaikaisessa toimintaympäristössä, jossa kyberuhat kehittyvät jatkuvasti ja voivat aiheuttaa merkittävää vahinkoa organisaatioiden toiminnalle ja maineelle. Organisaatioiden tulee suojata tietovarantojaan hyödyntämällä niin teknisiä kuin koulutuksellisia keinoja.

Teknisellä puolella teknologioiden, kuten pääsynvalvonnan, hiekkalaatikoiden ja SIEM-järjestelmien, käyttö auttaa luomaan vahvan puolustusmekanismin uhkia vastaan (González-Granadillo, González-Zarzosa, & Diaz, 2021; El Demerdash, 2023). Esimerkiksi pääsynvalvonta, erityisesti nollaluottamusmallia hyödyntäen, varmistaa, että arkaluonteisiin tietoihin pääsevät käsiksi vain valtuutetut henkilöt. SIEM-järjestelmien avulla voidaan saavuttaa kattava tilannekuva organisaation tietoturvasta, mikä mahdollistaa nopean reagoinnin mahdollisiin uhkiin. Hiekkalaatikot puolestaan tarjoavat eristetyn ympäristön epäluotettavien ohjelmien testaamiseen ilman, että ne vaarantavat järjestelmän turvallisuutta (Tiwari & Kumari, 2022).

Ennaltaehkäisyssä ei voida kuitenkaan unohtaa inhimillistä näkökulmaa. Organisaatioiden kyberturvan parantamisessa henkilöstön rooli on keskeinen, sillä monet kyberuhat, kuten tietojenkalastelu ja haittaohjelmat, ovat osittain mahdollisia inhimillisten virheiden kautta (Bada & Nurse, 2019). Tietoturvakoulutusten avulla voidaan lisätä henkilöstön tietämystä ja kykyä tunnistaa ja estää erilaisia hyökkäyksiä. Esimerkiksi kampanjat, joissa hyödynnetään luovia oppimismenetelmiä, kuten pelillistämistä ja virtuaalitodellisuutta, voivat vahvistaa tietoturvatietoisuutta tehokkaasti (Adinolf ym., 2019; Prümmer ym., 2024).

Ennaltaehkäisevä turvallisuus korostuu jo lainsäädännönkin vaatimusten kautta. Yrityksillä on merkittävä vastuu henkilötietojen käsittelyssä, ja tätä säädellään useilla laeilla. Yksi keskeisimmistä säädöksistä on EU:n yleinen tietosuoja-asetus (engl. General Data Protection Regulation, GDPR), joka asettaa tiukat vaatimukset henkilötietojen käsittelylle ja suojaamiselle kaikissa EU-maissa (Tietosuojavaltuutetun toimisto, 2023). Tietoturvan varmistamiseksi yritysten on noudatettava erilaisia protokollia, kuten tietojen varmuuskopiointia ja salausta, jotka suojaavat sekä asiakastietoja että organisaation järjestelmiä.

3 Tekniset ratkaisut, koneoppimisen ja syväoppimisen menetelmien sovellutukset sekä niiden rooli ennaltaehkäisevässä kyberturvallisuudessa

Ghelanin (2022) mukaan suurin osa yrityksistä hyödyntää perustason tietoturvatoimenpiteitä, kuten virustorjuntaohjelmistoja, palomuureja, vakoiluohjelmien torjuntaohjelmia, virtuaalisia erillisverkkoja (VPN), haavoittuvuuksien hallintaa sekä salattua tietoliikennettä. Näitä toimenpiteitä pidetään perustason ratkaisuina, sillä ne ovat yleisesti tunnettuja, helposti käyttöönotettavia ja tarjoavat suojan tunnetuimpia kyberuhkia vastaan. Perustason tietoturvatoimet ovat kuitenkin tyypillisesti reaktiivisia teknisiä ratkaisuja, jotka yksinään eivät riitä vastaamaan nykypäivän kehittyneisiin ja jatkuvasti muuttuviin uhkiin.

Edistyneemmät tietoturvatoimenpiteet, kuten ennaltaehkäisyn, pelotteiden, valvonnan ja harhautuksen hyödyntäminen, laajentavat tietoturvaa teknologian ulkopuolelle strategisemmalle tasolle. Ne keskittyvät muun muassa ihmisten käyttäytymisen ohjaamiseen, turvallisuuskulttuurin vahvistamiseen ja jatkuvaan tilannekuvan ylläpitämiseen. (Steingartner ym., 2021.) Näiden lähestymistapojen yhdistelmällä voidaan luoda monikerroksinen ja mukautuva tietoturvastrategia, joka tarjoaa kattavamman suojan organisaation tietovarannoille ja parantaa merkittävästi kyberuhkien torjuntakykyä (Marchal ym., 2024).

Virustorjuntaohjelmistot ovat myös osa tietoturvaa, ja ne voidaan jakaa kahteen päätyyppiin: allekirjoituspohjaisiin ja heuristiikkapohjaisiin (Jawhar, 2023). Allekirjoituksiin perustuva tunnistus on yleisin strategia, jota saatavilla olevat virustentorjuntaohjelmat hyödyntävät. Tämä tekniikka käyttää haittaohjelmatiedostoa, joka otetaan talteen, ja siitä erotetaan tunnusomainen allekirjoitus. Allekirjoitus mahdollistaa samanlaisten tunnisteiden omaavien haittaohjelmien löytämisen (Shijo & Salim, 2015). Useimmat virustentorjuntaohjelmat käyttävät tätä menetelmää. Menetelmä perustuu haittaohjelman omiin ominaisiin allekirjoituksiin, jotka voivat olla esimerkiksi tiedoston tiiviste tai tietty tavusarja. Tämän lähestymistavan etuna on alhainen väärien positiivisten havaitsemismäärä. Menetelmän heikkoutena on kuitenkin sen kyvyttömyys havaita uusia haittaohjelmia tai sellaisia, jotka muuttavat allekirjoituksiaan. (Jawhar, 2023.) Allekirjoituksiin perustuva lähestymistapa käyttää staattista jäsentämistä erityisten peräkkäisten tavujen tunnistamiseen, joita kutsutaan tageiksi (Yunus & Ngah, 2020).

Heuristinen tunnistus, joka tunnetaan myös epätavallisuuden tai käyttäytymiseen perustuvana tunnistuksena, on menetelmä, jossa haittaohjelman havaitseminen perustuu sen toiminnan

analysointiin sen suorittamisen aikana. Ensimmäisessä vaiheessa kerätään tietoa haittaohjelmasta harjoitteluvaiheen aikana. Seuraavaksi kerättyä tietoa tulkitaan ja siitä erotetaan tärkeimmät yksityiskohdat, jotka ryhmitellään käyttäytymismallin muodostamiseksi. Lopuksi testausvaiheessa haittaohjelma tunnistetaan vertaamalla sen käyttäytymismallia haitalliseksi luokiteltujen ohjelmien käyttäytymiseen. (Jawhar, 2023.)

Tämän lähestymistavan merkittävä etu on sen kyky tunnistaa tuntemattomat haittaohjelmat, joita perinteiset menetelmät eivät välttämättä havaitse. Menetelmän haasteena on korkea väärien positiivisten havaintojen määrä, mikä voi johtaa haitattomien ohjelmien virheelliseen luokitteluun haitallisiksi. (Jawhar, 2023.) Heuristisen tunnistuksen vahvuus on siis erityisesti modernien haittaohjelmien havaitsemisessa. Ominaisuus on erityisen tärkeä ennaltaehkäisevässä kyberturvassa, jossa pyritään estämään uhkia, ennen kuin ne ehtivät aiheuttaa vahinkoa.

Hyökkäyksen havaitsemisjärjestelmät (engl. Intrusion Detection System, IDS) ovat myös tärkeä osa yritysten tietoturvaa. Ne tarkkailevat järjestelmiä ja palvelimia havaitakseen epäilyttävää toimintaa, kuten poikkeuksellista verkkoliikennettä tai luvattomia tunnistautumisyrityksiä. IDS-järjestelmät eivät kuitenkaan estä hyökkäyksiä suoraan, vaan niiden tehtävänä on havaita uhkia ajoissa ja mahdollistaa niiden torjuminen ennen merkittäviä vahinkoja. Haasteena on, että IDS-järjestelmät tunnistavat ensisijaisesti vain aiemmin havaittuja hyökkäysmalleja eivätkä välttämättä havaitse täysin uusia uhkia. (Sarker ym., 2021.)

Yritysten kyberturvallisuusstrategiat keskittyvät suojaamaan tietokonejärjestelmiä, -verkkoja, ohjelmistoja ja dataa tietomurroilta ja tietoturvapoikkeamilta. Tähän pyritään hyödyntämällä erilaisia suojausmekanismeja, kuten pääsynvalvontaa, palomuureja, virustorjuntaohjelmistoja, hiekkalaatikkoja, tietoturvatapahtumien hallintaa sekä salaustekniikoita (Sarker ym., 2021).

Pääsynvalvonnan avulla organisaatiot rajoittavat tietyille käyttäjille myönnettyjä käyttöoikeuksia, jolloin vain tarvittavat henkilöt pääsevät käsiksi arkaluontoisiin tietoihin. Tämä vähentää merkittävästi tietoturvariskejä. Erityisesti nollaluottamusmalli (engl. Zero Trust Model) korostaa oletusta, ettei mikään käyttäjä tai laite ole oletusarvoisesti luotettava riippumatta siitä, sijaitseeko se organisaation sisä- vai ulkoverkossa. Tällainen lähestymistapa yhdistettynä tarkkaan pääsynvalvontaan auttaa ehkäisemään sisäisiä uhkia ja estämään luvattoman pääsyn myös silloin, kun ulkoinen suojaus on murrettu. (El Demerdash, 2023.)

Hiekkalaatikot (engl. sandboxing) tarjoavat eristetyn ja hallitun testausympäristön, jossa epäluotettavia tai mahdollisesti haitallisia ohjelmia voidaan ajaa ilman, että ne vaarantavat

varsinaisen käyttöjärjestelmän tai verkon turvallisuutta (Tiwari & Kumari, 2022). Hiekkalaatikkotekniikat ovat osoittautuneet tehokkaiksi erityisesti haittaohjelmien analysoinnissa ja uusien uhkien torjunnassa, sillä ne mahdollistavat reaaliaikaisen tarkkailun ja käyttäytymisanalyysin turvallisessa kontekstissa. Tämä tekee niistä keskeisen osan nykyaikaista ennaltaehkäisevää kyberturvapuolustusta.

Tietoturvatapahtumien hallinta (engl. Security Information and Event Management, SIEM) yhdistää tietoturvahälytyksiä analysoivat järjestelmät ja tapahtumien hallinnan. SIEM-järjestelmät tarjoavat organisaatioille reaaliaikaisen kokonaiskuvan tietoturvatilanteesta, mikä mahdollistaa uhkien havaitsemisen ja niihin reagoimisen tehokkaasti. (González-Granadillo, González-Zarzosa, & Diaz, 2021.) Salausmenetelmät puolestaan varmistavat, että arkaluonteiset tiedot pysyvät suojattuina, sillä vain valtuutetut osapuolet voivat purkaa niiden sisällön. (Sarker ym., 2021)

Perustason ratkaisut, kuten virustorjuntaohjelmistot ja palomuurit, tarjoavat siis suojan tunnettuja uhkia vastaan, mutta edistyneemmät menetelmät, kuten heuristinen tunnistus ja hiekkalaatikot, ovat elintärkeitä tuntemattomien uhkien torjunnassa. SIEM-järjestelmät (engl. Security Information and Event Management), jotka yhdistävät tietoturvahälytysten analysoinnin ja tapahtumien hallinnan, mahdollistavat yrityksille jatkuvan tilannekuvan kyberturvallisuudestaan (González-Granadillo ym., 2021). Näiden järjestelmien avulla yritykset voivat paitsi havaita uhat reaaliajassa myös reagoida niihin nopeasti.

Bayuk ym. 2012 alussa määrittelemien periaatteiden mukaan kyberturvallisuusstrategiassa yhdistetään teknologia, prosessit ja ihmisten rooli kattavaksi tietoturvaratkaisuksi. Pääsynvalvonta, kuten nollaluottamusmallin käyttö, varmistaa, että käyttöoikeudet myönnetään vain tarpeellisille henkilöille (El Demerdash, 2023). Samoin salausmenetelmät suojaavat arkaluonteisia tietoja valtuuttamattomilta käyttäjiltä (Sarker ym., 2021). Hiekkalaatikot tarjoavat turvallisen testausympäristön uusien uhkien analysointiin, kun taas IDS-järjestelmät täydentävät tätä tarkkailemalla järjestelmien toimintaa mahdollisten hyökkäysten havaitsemiseksi (Sarker ym., 2021). Näiden teknologisten ratkaisujen tehokas hyödyntäminen luo organisaatiolle vankan kyberturvakehyksen. Alla taulukossa 1 on listattuna yllä mainittuja ennaltaehkäisevän kyberturvallisuuden tekniikoita.

Taulukko 1 Ennaltaehkäisevän kyberturvallisuuden tekniikat

Toimenpide	Kuvaus	Esimerkkejä	Edut	Haasteet
Perustason tekniset ratkaisut	Reaktiivisia, yleisesti tunnettuja ja helposti käyttöönotettavia suojausmenetelmiä	Virustorjunta, palomuurit, vakoiluohjelmien torjunta, VPN, haavoittuvuuksien hallinta, salaus	Suojaavat tunnetuimmilta uhkilta; helppo ylläpitää	Eivät löydä täysin uusia uhkia; rajoittuvat tunnettuun hyökkäysmalliin
Edistyneet toimenpiteet	Strategisia ja ennakoivia keinoja, jotka suuntautuvat ihmisiin ja prosesseihin	Pelote- ja harhautustekniikat, käyttäytymiseen perustuva valvonta, jatkuva tilannekuva	Vahvistavat turvallisuuskulttuuria; ohjaavat käyttäytymistä	Vaativat jatkuvaa ylläpitoa ja kattavaa organisaatiotukea
Virustorjuntaohjelmistot	Tunnistavat haittaohjelmat allekirjoitusten tai heuristiikan avulla	Allekirjoituspohjainen ja heuristinen tunnistus	Allekirjoitus: vähän vääräpositiivisia; heuristiikka: löytää uusia uhkia	Allekirjoitus: ei huomaa allekirjoitustaan muuttavia; heuristiikka: paljon vääräpositiivisia
Hyökkäyksen havaitsemisjärjestelmät (IDS)	Tarkkailevat ja ilmoittavat epäilyttävästä verkko- tai järjestelmätoiminnasta	Verkkoliikenteen poikkeavuuksien ja luvattomien kirjautumisyritysten seuranta	Ajoissa havaittavat hyökkäykset; parantaa reagointikykyä	Tunnistavat pääosin vain tunnettuja hyökkäysmalleja
Pääsynvalvonta & Zero Trust	Rajoittaa käyttöoikeuksia vain niille, jotka niitä todella tarvitsevat	Monivaiheinen tunnistus, vähimmän oikeuden periaate	Estää sisäiset ja ulkoiset uhkat tehokkaammin	Korkea käyttöönoton ja hallinnan monimutkaisuus
Hiekkalaatikot (Sandboxing)	Eristetty ympäristö haitallisten ohjelmien testaamiseen ja käyttäytymisen analysointiin	Konttiteknologia tai virtuaalikoneet epäluotettaville sovelluksille	Turvallinen analyysi ja reaaliaikainen käyttäytymisanalyysi	Resurssivaatimus; monimutkainen hallita suurissa ympäristöissä
SIEM	Yhdistää tietoturvahälytykset ja tapahtumien hallinnan yhdelle työkalulle	Keskitetty lokien keruu, analytiikka ja hälytykset	Tarjoaa reaaliaikaisen tilannekuvan ja nopean reagoinnin	Suuret tietomäärät voivat aiheuttaa haasteita
Salaustekniikat	Suojaavat tiedonsiirtoa ja tallennettua dataa	End-to-end-salaus, TLS/SSL, levysalaus	Estää tiedon sieppauksen ja manipuloinnin	Avainhallinnan ja suorituskyvyn haasteet

Tekoälyteknologiat ovat viime vuosina muodostuneet keskeiseksi osaksi yritysten ja organisaatioiden kyberturvallisuusratkaisuja. Tekoälyn avulla on mahdollista havaita poikkeavuuksia, estää kehittyneitä hyökkäyksiä ja parantaa järjestelmien sopeutumiskykyä yhä monimutkaisemmaksi muuttuvassa uhkaympäristössä (Marchal ym., 2024). Tekoälypohjaiset järjestelmät pystyvät analysoimaan suuria määriä verkkoliikennettä ja tunnistamaan epätyypillistä käyttäytymistä nopeammin ja luotettavammin kuin ihmisvalvonta (Puthal ym., 2021; Szepesvári, 2015), mikä mahdollistaa myös varhaisessa vaiheessa tapahtuvat toimenpiteet uhkien torjumiseksi (Chen ym., 2016; Hassanien ym., 2021). Erityisen hyödyllisiä nämä ratkaisut ovat kehittyneiden kyberuhkien torjunnassa, sillä tekoäly kykenee havaitsemaan myös aiemmin tuntemattomia hyökkäysmenetelmiä mukautumalla uusiin uhkiin jatkuvan oppimisen avulla (Hassanien ym., 2021).

Tärkeimmät tekoälyn sovellusalueet kyberturvallisuudessa liittyvät koneoppimisen ja syväoppimisen menetelmiin. Koneoppimismenetelmillä viitataan tässä yhteydessä menetelmiin, joissa järjestelmä oppii tunnistamaan uhkia ja riskitekijöitä analysoimalla suuria tietomassoja ilman tarkkoja valmiiksi määriteltyjä sääntöjä (Marchal ym., 2024). Esimerkiksi tukivektorikoneet, satunnaismetsät ja päätöspuumallit ovat osoittautuneet tehokkaiksi työkaluiksi erityisesti

poikkeavuuksien tunnistamisessa ja haittaohjelmien luokittelussa (Muhammad ym., 2025; Midighe Usoh ym., 2023). Koneoppimismenetelmien hyödyntäminen mahdollistaa järjestelmille jatkuvan parantamisen aiempien tapahtumien pohjalta ja vähentää virheellisten positiivisten tunnistusten määrää kokonaisvaltaisessa tietoturvassa (Dale, 1995; Goldfarb ym., 2022).

Syväoppimismenetelmät laajentavat koneoppimisen mahdollisuuksia hyödyntämällä monikerroksisia tekoälymalleja, jotka kykenevät tunnistamaan erittäin monimutkaisia datarakenteita ja uhkamalleja, kuten hyökkäysketjuja tai haitallisten verkkotunnusten tarkkoja tunnuspiirteitä (Le ym., 2018; Marchal ym., 2024). Syväoppimismenetelmät vähentävät asiantuntijatyön tarvetta automatisoidun piirreanalyysin ansiosta ja soveltuvat erityisesti tilanteisiin, joissa tietomäärät ovat erittäin suuria ja reagointinopeus on kriittistä, kuten reaaliaikaisessa haittaohjelmavalvonnassa (Rigaki & Garcia, 2020; Hou ym., 2020). Vaikka syväoppimismenetelmät tuovat huomattavia etuja, liittyy niihin myös haasteita, kuten mallien tulkittavuuden ja datan laadun vaatimukset, sekä riski, että mallit voivat joutua manipuloivien hyökkäysten kohteeksi (Ozkan-Okay ym., 2024; Makkar ym., 2020; Papernot ym., 2016).

Syvä- ja koneoppimismenetelmät ovat osoittautuneet tehokkaiksi keinoiksi vahvistaa tietoturvaa. Ne mahdollistavat ennakoivaa suojautumista, ja tehostavat resurssien käyttöä muuttuvassa kyberympäristössä (Salem ym., 2024). Niiden hyödyntämisessä tulee kuitenkin huomioida teknologian tuomat haasteet, kuten mallien selitettävyys ja jatkuvan valvonnan sekä tietosuojan tarve (Marchal ym., 2024; Euroopan unionin neuvosto, 2021).

Koneoppimisen ja syväoppimisen erilaiset menetelmät ovat yleistyneet nopeasti kyberturvallisuuden sovellusalueilla, mutta kaikkien menetelmien merkitys ei ole tietoturvan kannalta yhtä olennainen. Kyberturvallisuuden erityispiirteenä on toimintaympäristön jatkuva muutos. Hyökkäystavat kehittyvät nopeasti, datamäärät kasvavat ja uhkat voivat olla täysin uusia sekä aiemmin havaitsemattomia. Tavanomaiset sääntöpohjaiset järjestelmät eivät riitä tällaisen dynaamisen ja moniulotteisen ympäristön hallintaan. Siksi painopiste tulee asettaa niihin kone- ja syväoppimismenetelmiin, joiden on tieteellisessä kirjallisuudessa ja käytännön sovelluksissa osoitettu parhaiten ratkaisevan näitä haasteita. (Marchal ym., 2024.)

Erityisesti haitallisen verkkoliikenteen tunnistuksessa, käyttäytymisanalyysissä, uhkamallinnuksessa ja reaaliaikaisessa riskianalyysissä vaaditaan järjestelmiä, jotka pystyvät sopeutumaan nopeasti, tunnistamaan poikkeavuuksia ja oppimaan uusia uhkia erittäin suurista ja monimuotoisista tietoaineistoista. Näillä osa-alueilla neuroverkkojen, tukivektorikoneiden, satunnaismetsien ja yhdistelmäalgoritmien on havaittu tarjoavan kyvykkyyksiä, jotka ovat

olennaisen tärkeitä nykyaikaisessa kyberturvassa. Ne mahdollistavat automaattisen uhkien tunnistamisen myös tuntemattomissa tilanteissa, vähentävät väärien hälytysten määrää ja tukevat tietoturva-asiantuntijoiden päätöksentekoa (Muhammad ym., 2025; Rigaki & Garcia, 2020; Midighe Usoh ym., 2023).

Tässä luvussa keskitytään niihin koneoppimisen ja syväoppimisen ratkaisuihin, joita kirjallisuuden ja empiiristen tutkimusten perusteella voidaan pitää tietoturvan kannalta kriittisimpinä. Tarkastelu perustuu sekä menetelmäkohtaiseen analyysiin että niiden käytännön yhteyteen kyberturvallisuuden ydintoimintoihin. Näin voidaan kuvata, miten ja miksi nämä ratkaisut mahdollistavat tehokkaimman uhkien havaitsemisen sekä nykyaikaisen, reagointikykyisen ja ennakoivan kyberturvallisuuden rakentamisen.

3.1.1 Haitallisen verkkoliikenteen tunnistaminen

Haitallisen verkkoliikenteen tunnistaminen on yksi keskeisimmistä sovellusalueista, jossa koneoppimista ja syväoppimista hyödynnetään kyberturvallisuuden ennaltaehkäisyssä. Koneoppimismallit analysoivat verkkoliikenteen käyttäytymismalleja ja oppivat tunnistamaan poikkeavuuksia, jotka voivat viitata haitalliseen toimintaan. Toisin kuin perinteiset tunnistusmenetelmät, jotka perustuvat ennalta määriteltyihin sääntöihin ja tunnettuihin uhkakuviin, koneoppimismallit voivat havaita epänormaalia toimintaa verkossa ilman, että kyseistä hyökkäystä on aiemmin esiintynyt. Tämä mahdollistaa ennakoivan uhkien tunnistamisen, jolloin järjestelmä voi reagoida uusiin ja tuntemattomiin kyberuhkiin ennen kuin ne ehtivät aiheuttaa merkittävää vahinkoa. (Ansari ym., 2022; Zhao ym., 2024; Fahim ym., 2025; Wang ym., 2023.) Neuroverkkoihin perustuvat mallit ovat erityisen tehokkaita haitallisen verkkoliikenteen tunnistamisessa, sillä ne kykenevät analysoimaan monimutkaisia liikennemalleja ja havaitsemaan poikkeavuuksia, kuten DoS-hyökkäyksiä ja haitallista ohjelmakoodin lataamista, tietojen kalastelua ja muita tunkeutumisyrityksiä (Radford ym., 2018). Tutkimuksessa verkkoliikenteen metatiedot, kuten siirrettyjen tavujen määrä, muunnettiin sanoiksi, jotka muodostivat lauseita tietokoneiden välisestä viestinnästä. Nämä lauseet syötettiin syväoppimismalliin, joka hyödyntää pitkämuistisia neuroverkkoja (engl. Long Short-Term Memory, LSTM) ja toistoneuroverkkoja (engl. Recurrent Neural Networks, RNN). Malli oppi verkkoliikenteen kielen rakenteen, mikä mahdollisti poikkeamien ennustamisen ja kyberuhkien tunnistamisen. Juuri pitkämuistisiin ja toistoneuroverkkoihin pohjautuvia sekventiaalista rakennetta eli esimerkiksi tapahtumasarjoja, kuten pyyntö, vastaus, komento hyödyntäviä malleja on useita (Kim ym., 2016; Yin ym., 2017; Almiani ym., 2019; Kolosnjaji ym., 2016). Mallien tehtävänä on tyypillisesti ennustaa seuraava

tapahtuma viestintäketjussa, ja poikkeamat ennusteesta viittaavat mahdollisiin poikkeavuuksiin tai hyökkäyksiin. Poikkeuksena Wang ym., 2017 kehittämä malli, jossa toiminta perustui konvoluutionaalisiin verkkoihin (engl. Convolutional Neural Network, CNN). Tutkimuksessa muutettiin verkkopakettien eli pienten tietoyksikköjen, joiden avulla tietoa siirretään verkossa sisältämä raakadata taulukoiksi, jotka muistuttivat kuvia. Nämä kuvat syötettiin konvoluutioverkkoihin perustuvalle tekoälyjärjestelmälle, joka oli koulutettu tunnistamaan niistä poikkeuksia ja haitallisia malleja samalla tavalla kuin ihminen erottaa epätavallisen muodon kuvasta. Wangin ym. 2017 kehittämällä mallilla tarkkuus, väärien positiivisten määrä sekä tunnistuskyky oli parempi RNN- ja LSTM-pohjaisiin malleihin verrattuna. RNN, LSTM sekä CNN pohjaisten mallien automaattinen tunnistus ilman ohjattua oppimista mahdollistaa reaaliaikaisen uhkien torjunnan, jolloin hyökkäykset voidaan estää ennen kuin ne ehtivät aiheuttaa vahinkoa. Tämä tekee neuroverkkoihin perustuvista ratkaisuista keskeisen työkalun ennakoivassa kyberpuolustuksessa, jossa tavoitteena on havaita ja neutraloida uhkia jo niiden alkuvaiheessa.

Myös Zhao ym. (2024), Fahim ym. (2025) ja Wang ym. (2023) ovat tutkineet tekoälyyn perustuvia ratkaisuja haitallisen verkkoliikenteen tunnistamiseksi, mutta heidän tavassaan lähestyä aihetta on huomattavia eroavaisuuksia niin mallien arkkitehtuurissa, datankäsittelyssä kuin soveltuvuudessa erityyppisiin verkkoympäristöihin. Zhao ym. (2024) esittelevät tutkimuksessaan syvävahvistusoppimiseen perustuvan menetelmän haitallisen verkkoliikenteen tunnistamiseen. Heidän lähestymistapansa yhdistää kaksi yleistä tekniikkaa. Toinen on koneoppimisen menetelmiin lukeutuva päätöspuualgoritmi, joka tekee luokittelusta selkeää ja selitettävää sekä syvävahvistusoppimisen Proximal Policy Optimisation (PPO) -algoritmi, joka mahdollistaa mallin jatkuvan parantamisen ja sopeutumisen uusiin uhkiin.

Keskeinen innovaatio tutkimuksessa on entropian käyttäminen osana mallin koulutusta. Tässä yhteydessä entropia voidaan tulkita epävarmuudeksi tai moninaisuudeksi mallin tekemissä päätöksissä. Käytännössä siis miten paljon se uskaltaa tutkia erilaisia ratkaisuja verrattuna siihen, että se tyytyy valitsemaan aina saman totutun toimintatavan. Korkea entropia koulutuksen aikana kannustaa mallia kokeilemaan uusia strategioita ja tutustumaan erilaisiin mahdollisiin ratkaisuihin. Tämä estää mallia jämähtämästä liian aikaisin yhteen lähestymistapaan, joka ei välttämättä ole paras mahdollinen, ja antaa mahdollisuuden löytää tehokkaampia tapoja tunnistaa haitallinen liikenne, myös silloin kun kohtaamiset ovat uusia tai poikkeavia.

PPO-algoritmin ja entropian yhdistelmällä voidaan saavuttaa erittäin lupaavia tuloksia, kun malli pystyy sekä hyödyntämään oppimiaan tehokkaita ratkaisuja, että edelleen tutkimaan ja kokeilemaan

uusia mahdollisuuksia. Tämä korostuu nykyisessä uhkakentässä, jossa hyökkäysmuodot vaihtelevat nopeasti ja uusia tuntemattomia uhkia syntyy jatkuvasti.

Uutta tulokulmaa testattiin sekä simuloiduilla että oikeilla verkkoliikenteen aineistoilla. Heidän mallinsa osoitti sekä korkeaa tarkkuutta että kykyä sopeutua uusiin ja kehittyviin uhkatilanteisiin. Malli onnistui havaitsemaan haitallisen liikenteen tehokkaammin ja joustavammin kuin perinteiset staattiset koneoppimismallit. Entropian hyödyntäminen paransi erityisesti mallin kykyä löytää aiemmin tunnistamattomia hyökkäysmalleja.

Aiempi kirjallisuus on painottanut vahvistusoppimisen mahdollisuuksia kyberturvallisuudessa, mutta Zhao ym. (2024) tutkimus tuo uuden näkökulman entropiapohjaisen säännöstelyn hyödyntämisestä, jolloin malli ei ainoastaan pysty sopeutumaan, vaan aktiivisesti etsii erilaisia ratkaisuvaihtoehtoja, mikä lopulta tuottaa tehokkaamman puolustuksen verkkoympäristöön.

Fahim ym. (2025) vertaavat klassisia koneoppimismalleja, kuten satunnaismetsiä, syväoppimiseen pohjautuviin ratkaisuihin kuten syviin neuroverkkoihin (engl. Deep Neural Network, DNN). Heidän analyysinsa osoittaa, että syväoppimismallit ylittävät selvästi perinteiset mallit haitallisen liikenteen havaitsemisen tarkkuudessa, erityisesti tilanteissa, joissa liikennedatan piirteet ovat moniulotteisia ja epälineaarisia. Tämä tulos on linjassa vallitsevan näkemyksen kanssa, joka korostaa, että syvät neuroverkot kykenevät tunnistamaan ja mallintamaan monimutkaisia suhteita, joita perinteiset koneoppimismenetelmät eivät tavoita yhtä hyvin. Merkille pantavaa on kuitenkin, että syväoppimismallien tehokas käyttö edellyttää riittävästi korkealaatuista harjoitusdataa ja niiden selitettävyyttä voi olla vaikea arvioida.

Wang ym. (2023) lähestyvät haitallisen verkkoliikenteen tunnistamisen haastetta erityisesti salatun liikenteen näkökulmasta. He kehittämässään mallissa on kyse kaksitasoisuudesta, jossa yhdistyvät sekä kone- että syväoppimistekniikat, mahdollisimman tehokkaaseen piirteiden louhintaan ilman tarvetta purkaa verkkoliikennettä. Tämä lähestymistapa on erittäin ajankohtainen nykyisessä kyberturvallisuusympäristössä, jossa salattu liikenne on yhä yleisempää ja haittaliikenteen tunnistus ilman purkamista on kriittinen haaste. Menetelmä on erittäin tarkka salatun haitallisen liikenteen tunnistamisessa, mutta mallin heikkoutena on tarve laajalle ja monipuoliselle harjoitusaineistolle. Lisäksi uudenlaisten uhkien havaitseminen saattaa olla haastavaa, jos harjoitusdata ei anna riittävän tarkkaa ja todenmukaista kuvaa tapahtuneista hyökkäyksistä.

Tekoälypohjaiset, erityisesti syväoppimista ja yhdistelmämenetelmiä hyödyntävät mallit tarjoavat selkeitä etuja perinteisiin menetelmiin nähden haitallisen verkkoliikenteen tunnistamisessa.

Syvävahvistusoppiminen ja joustavat hybridimallit mahdollistavat paremman uhkien havainnoinnin dynaamisessa ja monimutkaisessa verkkoympäristössä, mutta niiden tehokas hyödyntäminen edellyttää korkeaa laskentakapasiteettia ja laadukasta harjoitusdataa. Lisäksi mallien joustavuus ja kyky mukautua uusiin uhkakuviin ovat edelleen kehityksen kohteena. Tähän nojaten optimaalinen ratkaisu vaatii kompromissin mallien laskennallisten vaatimusten, harjoitusaineiston saatavuuden ja käytännön sovellettavuuden välillä (Zhao ym., 2024; Fahim ym., 2025; Wang ym., 2023).

Tekoäly on tuonut paljon haitallisen verkkoliikenteen tunnistamiseen ja analysointiin, erityisesti monimutkaisissa verkkoympäristöissä, joissa perinteiset tietoturvaratkaisut ovat osoittautuneet riittämättömiksi. Tunkeutumisen havaitsemis- ja ehkäisyjärjestelmät (engl. Intrusion Detection System, IDS ja Intrusion Prevention System, IPS) muodostavat kyberturvallisuuden kovan ytimen, ja tekoälypohjaisten menetelmien integrointi näihin järjestelmiin on lisännyt niiden tehokkuutta merkittävästi. Perinteiset tunnistusmenetelmät perustuvat ennalta määriteltyihin sääntöihin ja tunnettuja uhkia kuvaaviin allekirjoituksiin, mutta tämä lähestymistapa ei riitä torjumaan uusia ja hienovaraisia kyberuhkia. Koneoppimismallit kykenevät oppimaan verkkoliikenteen normaaleja käyttäytymismalleja ja havaitsemaan poikkeavuuksia, jotka voivat viitata luvattomaan käyttöön, tietomurtoihin tai haittaohjelmien leviämiseen. (Salem ym., 2024; Sowmya ym., 2023.)

Poikkeamien havaitsemisjärjestelmät perustuvat normaalin verkkoliikenteen mallintamiseen, jolloin poikkeamat voidaan tulkita mahdollisiksi hyökkäyksen merkeiksi. Tämä lähestymistapa on tehokas, kun kyseessä ovat näkyvät ja selkeät verkkoliikenteen muutokset, kuten porttiskannaukset tai palvelunestohyökkäykset. Poikkeamien havaitsemisjärjestelmät eivät kuitenkaan kykene tunnistamaan tehokkaasti hienovaraisia hyökkäyksiä tai haittaohjelmaviestintää, jotka muistuttavat tavanomaista viestinvaihtoa. Tällaisia uhkia varten tekoälypohjaiset järjestelmät hyödyntävät herkempää poikkeamien havaitsemista, mikä voi kuitenkin lisätä väärien hälytysten määrää. (Salem ym., 2024; Marchal ym., 2024.) Väärien hälytysten hallinta on kriittinen osa haitallisen liikenteen analysointia, sillä liiallinen hälytyskuormitus voi vaikeuttaa tehokasta päätöksentekoa.

Väärien hälytysten suuren määrän vuoksi havaitsemis- ja ehkäisyjärjestelmät hyödyntävät tarkempia luokittelumenetelmiä, jotka parantavat uhkien tunnistustarkkuutta. Binääriluokittelu erottaa haitallisen verkkoviestinnän normaalista liikenteestä, jolloin järjestelmä voi tunnistaa ja estää epäilyttävät tapahtumat. Moniluokkaluokittelu taas mahdollistaa erilaisten hyökkäystyyppien ja haitallisten käyttäytymismallien erottamisen toisistaan, mikä parantaa uhkien analysointia ja torjuntaa. Hahmontunnistus puolestaan tunnistaa käyttäjän, laitteen tai sovelluksen verkkoviestinnän perusteella. Menetelmä perustuu yksityiskohtaiseen mallinnukseen, jossa

viestintäkuvioita analysoidaan ja verrataan tunnettuun normaaliin käyttäytymiseen (engl. pattern matching). (Salem ym., 2024; Marchal ym., 2024; Sowmya ym., 2023.)

Koneoppimisalgoritmit hyödyntävät verkkoliikenteen tunnusmerkkejä analysoidakseen viestinnän tyypin ja tunnistaakseen mahdollisia uhkia. Tämä prosessi toteutetaan eri tarkkuustasoilla, jotka määräytyvät havaittavan toiminnan, analyysin viiveen, käytettävissä olevan laskentatehon ja protokollatiedon saatavuuden perusteella. Verkkoliikenteen salaus rajoittaa analysoitavissa olevaa tietoa, minkä vuoksi tiedon poiminta voi kohdistua yksittäisiin verkkopaketteihin, pakettiluokkiin, viestintävirtoihin tai koko verkkoliikenteeseen. (Marchal ym., 2024.)

Yksi merkittävimmistä haasteista verkkoturvallisuudessa on verkkoliikenteen monimuotoisuus. Erilaisten viestintäprotokollien suuri määrä vaikeuttaa kompleksisten ympäristöjen ja käyttäytymismallien tarkkaa mallintamista. Verkkoliikenteen salauksen yleistyminen on huomattava haaste poikkeavuuksien havaitsemiselle, koska se estää suoran pääsyn tietoliikenteen sisältöön. Salattujen yhteyksien vuoksi tietoturva-analyytikot eivät voi tarkastella pakettien sisältöä hyökkäysten tunnistamiseksi, vaan he joutuvat turvautumaan metatietoihin, kuten pakettien kokoon, ajalliseen jakaumaan, suuntaan ja protokollaan. Tämän rajoituksen vuoksi poikkeavuuksien tunnistaminen perustuu koneoppimismalleihin, jotka analysoivat normaalin liikenteen käyttäytymistä ja havaitsevat siitä poikkeavat mallit. Esimerkiksi poikkeavuus voi olla epätavallisen suuri tiedonsiirto ulkoiselle palvelimelle tai poikkeuksellinen viestintäkuvio, joka ei vastaa tyypillistä verkkokäyttäytymistä. Salattu liikenne ja verkkoviestinnän monimuotoisuus luovat suotuisan ympäristön kehittyneille kyberhyökkäyksille, jotka voivat naamioitua ja mukautua normaalilta vaikuttavaan verkkoliikenteeseen. (Marchal ym., 2024; Ibraheem ym., 2022; Wang ym., 2023.)

Tekoälypohjaiset verkkoturvaratkaisut soveltuvat parhaiten ympäristöihin, joissa verkkoliikenne on rakenteeltaan yksinkertaista ja ennustettavaa. Näihin kuuluvat esimerkiksi esineiden internet (engl. Internet of Things, IoT) -verkot sekä teollisuuden ohjausjärjestelmät (engl. Industrial Control System, ICS), joissa laitteiden käyttäytymismallit ovat vakaita ja helposti analysoitavissa. Näissä ympäristöissä koneoppimismallit voivat hyödyntää tehokkaita neuroverkkopohjaisia tunnistusmenetelmiä, jotka parantavat tunkeutumisen havaitsemista ja uhkien torjuntaa. (Abdullahi ym., 2022; Marchal ym., 2024.)

Tekoälyn merkitys haitallisen verkkoliikenteen tunnistamisessa on kasvussa, mutta sen täysi potentiaali ei ole vielä täysin toteutunut. Verkkoympäristöjen monimutkaisuus ja niiden jatkuva kehitys luovat haasteita tekoälytekniikoiden soveltamiselle kyberturvallisuudessa. Tulevaisuudessa

tekoälypohjaiset ratkaisut voivat saavuttaa paremman kypsyystason, mikä tekee mahdolliseksi entistä tarkemmat ja joustavammat tunnistusmenetelmät. Yhdistämällä tekoälyyn perustuvat analyysimenetelmät tehokkaisiin tietoturvamekanismeihin voidaan saavuttaa entistä parempi suojaus kehittyviä kyberuhkia vastaan.

Viime vuosien systemaattiset kirjallisuuskatsaukset (Dobler ym., 2024; Abdulganiyu ym., 2023; Jin ym., 2024) valottavat hyvin tutkimuskentän yleistä tilaa. Dobler ym. (2024) tarkastelivat kirjallisuuskatsauksessaan 62 vertaisarvioitua artikkelia ja konferenssijulkaisua, jotka oli julkaistu vuosina 2018–2023. Tutkimukset valittiin analysoitavaksi, mikäli ne käsittelivät kone- tai syväoppimisen menetelmiä haitallisen verkkoliikenteen tunnistuksessa, ja tarkasteluun otettiin ainoastaan tutkimuksia, joissa oli selkeästi raportoitu käytetyt aineistot ja suorituskykymittarit. Abdulganiyu ym. (2023) analysoivat puolestaan 54 kansainvälistä tutkimusartikkelia, jotka olivat ilmestyneet vuosina 2017–2022. Heidän keskeinen valintakriteerinsä oli, että tutkimukset käsittelivät erityisesti dataintegraatiota, tiedon laatua sekä koneoppimismallien selitettävyyttä verkkoliikenteen kontekstissa. Jin ym. (2024) laativat katsauksensa perustaksi 49 alkuperäistutkimusta, jotka kattoivat julkaisuajankohdat vuosilta 2019–2023. Näihin sisältyi artikkeleita, joissa on kehitetty tai arvioitu adaptiivisia ja kontekstuaalisia malleja haitallisen verkkoliikenteen tunnistamisessa. Kaikissa kolmessa katsauksessa tutkimusartikkelit valittiin systemaattisella kirjallisuushaulla, hyödyntäen kansainvälisiä tietokantoja, ja ne arvioitiin osuvuutensa, laadun sekä raportointitarkkuuden perusteella.

Keskeisin muutos tutkimuskentässä on kone- ja syväoppimismenetelmien hyödyntämisen yleistyminen. Dobler ym. (2024) katsaus osoittaa, että suurimmassa osassa uutta tutkimusta koneoppimisen ja syväoppimisen mallien käyttö johtaa selvästi parempaan tarkkuuteen haitallisen liikenteen tunnistuksessa, kun niitä verrataan perinteisiin sääntöpohjaisiin menetelmiin. Erityisesti mallien kyky käsitellä laajoja ja monimutkaisia tietoaineistoja sekä oppia itsenäisesti uusia uhkamalleja luo pohjan entistä automaattisemmille ja kattavammille tietoturvaratkaisuille. Samalla adaptiivisten ja itseoppivien mallien kehitys mahdollistaa Jin ym. (2024) katsauksen mukaan nopean reagoinnin muuttuviin uhkiin ja verkko-olosuhteisiin. Tämä dynaamisuus on nähtävissä sekä alkuperäisissä tutkimuksissa että alan meta-analyyseissä. Tiedon integroinnin kehittyminen on myös tuonut uusia mahdollisuuksia, sillä Abdulganiyu ym. (2023) korostavat monimuotoisten datalähteiden hyödynnettävyyden ja yhdistelyn merkitystä parantamaan turvallisuuden kokonaiskuvaa ja järjestelmien tarkkuutta.

Toisaalta systemaattinen kirjallisuus osoittaa myös huomattavia heikkouksia, joiden ratkaiseminen on välttämätöntä tutkimusalan eheyden ja jatkuvuuden kannalta. Erityisen merkittäväksi ongelmakohdaksi nousee datan sirpaleisuus ja laatu. Abdulganiyu ym. (2023) nostavat esiin, että suuri osa arvioiduista tutkimuksista kärsii puutteellisesta dataintegraatiosta ja heikkolaatuisista aineistoista. Hajautettu, formaateiltaan erilainen data heikentää mallien luotettavuutta ja vaikeuttaa tulosten yleistämistä laajempiin ympäristöihin. Selitettävyys on toinen keskeinen rajoite, sillä etenkin syväoppimismallit jäävät usein niin sanotuksi mustaksi laatikoksi, joiden päätösten perustelu vaatii vielä lisäkehitystä. Tämä tehdään erityisen näkyväksi Abdulganiyu ym. (2023) sekä Dobler ym. (2024) analysoimien alkuperäistutkimusten kautta, joissa mallien tulkittavuus ja päätösten validointi osoittautuvat käytännön soveltamisen haasteiksi.

Reaaliaikaisuuden ja skaalautuvuuden vaatimukset ovat erityisesti Dobler ym. (2024) tarkastelemissa tutkimuksissa yksi alan akuuteista pullonkauloista. Suurten mallien pyörittäminen ja jatkuva oppiminen edellyttävät huomattavia laskentaresursseja, eivätkä nykyiset ratkaisut useinkaan sovellu suoraan hajautettuihin ympäristöihin, kuten IoT-laitteisiin tai reunalaskentaan. Adaptiivisten algoritmien käyttöön liittyy myös aina riski yliennustamisesta ja itsepäivittymisen aiheuttamasta suorituskyvyn vaihtelusta, kuten Jin ym. (2024) huomauttavat.

Tulevaisuuden kehityssuunnat kumpuavat juuri niistä rajoitteista, jotka kirjallisuudessa selvästi tunnistetaan. Tutkimuksen painopisteen siirtäminen laajojen, standardoitujen ja laadukkaiden tietoaineistojen saatavuuteen ja yhdistämiseen parantaisi mallien yleistettävyyttä ja vertailtavuutta. Samoin selitettävää tekoälyä koskevat tutkimukset ovat alalla yhä kriittisempiä, jotta päätöksenteko olisi luotettavaa ja hyväksyttävissä myös käytännön tietoturvaympäristöissä. Laskenta- ja energiatehokkuuden optimoinnin kehittäminen mahdollistaisi monimutkaisten mallien hyödyntämisen myös resursoiduissa ympäristöissä. Lisäksi adaptiivisten järjestelmien vakauden ja luotettavuuden varmistaminen edellyttää uusia teknisiä ja teoreettisia ratkaisuja, jotka mahdollistavat jatkuvan kehityksen ilman ylisovittuneisuuden riskiä.

Kaiken kaikkiaan edellä kuvatut systemaattiset kirjallisuuskatsaukset osoittavat, että haitallisen verkkoliikenteen tunnistamisen tutkimus on kehittynyt teknologisesti nopeasti, mutta kenttä kohtaa edelleen keskeisiä ratkaisemattomia haasteita. Nämä liittyvät datan laatuun, mallien läpinäkyvyyteen ja järjestelmien käytännön sovellettavuuteen. Kysymysten myöhempi ratkaiseminen ohjaa tutkimusalaa kohti entistä moniulotteisempia, selitettävämpiä ja resurssitehokkaampia ratkaisuja, joiden kehitys nojautuu edelleen vahvasti tutkimusnäytön kriittiseen tarkasteluun.

3.1.2 Käyttäytymisen poikkeavuuksien analyysi reaaliaikaisessa seurannassa

Useat tuoreet tutkimukset osoittavat, että syväoppimismenetelmillä, erityisesti toistoneuroverkoilla (RNN) ja pitkämuistisilla LSTM-verkoilla, on kasvavaa käyttöä myös käyttäytymisanalytiikassa, jossa pyritään tunnistamaan poikkeamia yksilön tai järjestelmän normaalista toiminnasta. Menetelmiä hyödynnetään esimerkiksi käyttäjätoiminnan ajallisen rakenteen mallintamiseen, mikä mahdollistaa esimerkiksi epätyypillisten kirjautumisyritysten, tiedostotoimintojen ja prosessikäyttäytymisen tunnistamisen yritysverkoissa reaaliaikaisesti (Mohammed ym., 2025; Cui ym., 2022). LSTM-mallit kykenevät oppimaan hienovaraisempia ajallisia riippuvuuksia, jotka perinteisiltä malleilta jäävät huomaamatta Mallit kuitenkin edellyttävät huomattavan määrän hyvin esikäsiteltyä tapahtumadataa (Marchal ym., 2024).

UEBA (User and Entity Behavior Analytics) voidaan luokitella riskienhallintajärjestelmäksi, joka hyödyntää koneoppimista, algoritmeja ja tilastollista analyysiä reaaliaikaisten verkkouhkien havaitsemiseksi (Marchal ym., 2024). Toisen määritelmän mukaan UEBA on koneoppimiseen ja tilastollisiin malleihin perustuva lähestymistapa, joka oppii käyttäjän tai entiteetin eli esimerkiksi laitteen tai prosessin normaalin toimintaprofiilin ja tunnistaa siitä poikkeamat (Rashid ym., 2021; Khan ym., 2022). UEBA auttaa organisaatioita hallitsemaan käyttäjäturvallisuutta tehokkaasti esimerkiksi käyttäjäkoulutuksen avulla, jossa käyttäjä saa palautetta virheellisestä toiminnastaan ja oppii turvallisempia käytäntöjä. Jatkuva reaaliaikainen seuranta vähentää luvattomia pääsyjä ja epänormaalia käyttäytymistä, joko tahallisesti tai vahingossa, mikä mahdollistaa turvallisen ympäristön. Tahalliset uhat, kuten tietomurrot, palvelunestohyökkäykset ja haittaohjelmat, voidaan tunnistaa analysoimalla verkkoliikenteen kaavoja, kun taas inhimilliset virheet, kuten vahingossa tapahtuvat tiedostovuodot, tunnistetaan ja estetään automaattisesti, jos ne rikkovat tietoturvakäytäntöjä. (Marchal ym., 2024; Shashanka ym., 2016.)

Khan ym. (2022) ja Marchal ym. (2024) mukaan UEBA-järjestelmien käyttö on erityisen hyödyllistä identiteetinhallinnan riskienhallinnassa. Järjestelmien tulee kuitenkin oppia normaalikäyttäytymistä ennen kuin poikkeamat voidaan tunnistaa. Tästä syystä ne ovat alttiita virheille erityisesti oppimisvaiheessa. (Khan ym., 2022.) Esimerkiksi uudet käyttäjät, joista ei vielä ole muodostunut profiilia, voivat mahdollisesti jäädä vaille suojaa tai tulla virheellisesti luokitelluksi riskitekijöiksi. Marchal ym. (2024) havaitsivat, että ilmiö nousee esiin tilanteissa, joissa käyttäytymisprofiiliin päätyy poikkeavaa toimintaa jo oppimisvaiheessa.

Myös käyttäytymisen luonnollinen vaihtelu ajan myötä, kuten työnkuvan tai sijainnin muutokset, voivat johtaa väärien hälytysten määrän kasvuun. Hälytykset voidaan virheellisesti tulkita uhiksi.

Tästä syystä UEBA-järjestelmät tarvitsevat säännöllistä uudelleenkoulutusta ja kontekstitietoista sopeutumista. (Nguyen ym., 2019; Tang ym., 2017; Marchal ym., 2019.) Käyttäytymismallien päivittämisen on oltava jatkuvaa, jotta järjestelmä ei vanhenisi eikä muodostaisi vääriä oletuksia normaalista toiminnasta (Marchal ym., 2024).

Tang ym. (2017) ja Khan ym. (2022) korostavat, että poikkeavuuden pelkkä havaitseminen ei riitä vaan järjestelmän on kyettävä arvioimaan myös niiden merkityksellisyyttä. Näin korostuu mallien kyky yhdistää tekninen havainto kulloiseenkin kontekstiin. Esimerkiksi yön aikana epätavallisesta sijainnista tehty kriittinen lataus voidaan tulkita vaarattomaksi tai erittäin vakavaksi tilanteen mukaan. Tällöin voidaan suositella erilaisia ennaltaehkäiseviä toimintoja, kuten toiminnon esto, monivaiheinen todennus tai hiljainen hälytys (Tang ym., 2017; Marchal ym., 2024).

Useat tutkimukset kritisoivat UEBA-järjestelmiä siitä, että niissä toimenpiteiden määrittely tehdään usein manuaalisesti, eikä mukautuminen eri käyttäjärooleihin ole automaattista. (Marchal ym., 2024; Khan ym., 2022.) Tämän lisäksi on mahdollisuus, että väärien positiivisten hälytysten yleisyys heikentää järjestelmän käytettävyyttä ja uskottavuutta, ellei käyttäjällä ole mahdollisuutta antaa palautetta järjestelmän päätöksistä (Marchal ym., 2024). Tästä syystä Tang ym. (2017) varoittavat, että UEBA-sovelluksia ei tulisi käyttää ilman inhimillistä valvontaa kriittisessä päätöksenteossa, jossa väärä negatiivinen tai positiivinen johtaa vakaviin seurauksiin.

Raj ym., (2019) laatima laaja systemaattinen kirjallisuuskatsaus osoittaa, että UEBA-ratkaisut hyödyntävät sekä tilastollisia malleja että koneoppimisalgoritmeja. Näistä mainittakoon klusterointi ja anomalioiden tunnistus, joita hyödynnetään erityisesti mallinnuksessa, riskipisteytyksissä ja hälytyksissä. Katsauksessa analysoitiin 150 kansainvälistä tutkimusta ja alan raporttia vuosilta 2008–2018 koskien käyttäjän ja entiteetin käyttäytymisanalytiikkaa (UEBA). Katsauksen keskeisenä tavoitteena oli tuoda esiin, miten UEBA-menetelmiä on sovellettu tietoturvakontekstissa, mitä teknologioita ja analytiikkaratkaisuja on hyödynnetty, sekä mitä haasteita ja parhaita käytäntöjä tutkimus on paljastanut.

Kirjallisuuskatsauksen perusteella UEBA:n keskeinen vahvuus on sen kyky analysoida käyttäjän toimintaa kokonaisvaltaisesti. Yksittäisten tapahtumien sijaan huomio kiinnittyy käyttäytymisen poikkeamiin suhteessa historialliseen ja vertailukelpoiseen kontekstiin. Tyypillisesti esimerkiksi niin sanottu "first access anomaly" eli käyttäjän ensimmäinen pääsy uuteen resurssiin epätavallisessa tilanteessa voi toimia ensimerkkinä haitallisesta toiminnasta. Tämänkaltaiset käyttäytymisindikaattorit mahdollistavat pidemmälle kehittyneen riskienhallinnan kuin perinteinen

lokianalyysi, jossa tutkitaan järjestelmien tuottamia lokitietoja, kuten kirjautumisyrityksiä ja tiedostojen siirtoja.

Katsaus osoittaa myös, että UEBA-menetelmät nojautuvat tyypillisesti tilastollisiin malleihin ja koneoppimiseen, erityisesti valvomattoman oppimisen tekniikoihin. Näiden avulla voidaan muodostaa käyttäytymisprofiileja, joiden avulla tunnistetaan poikkeavuuksia ilman valmiita sääntöjä. Samanaikaisesti huomioitiin, että tehokkaan analytiikan toteuttaminen edellyttää toimiakseen skaalautuvia järjestelmiä, jotka pystyvät käsittelemään suuria tietomassoja reaaliaikaisesti.

Katsauksessa tunnistettiin useita keskeisiä haasteita UEBA:n käyttöönotossa ja tutkimuksessa. Näitä ovat muun muassa väärien positiivisten hälytysten suuri määrä, standardien ja yhteismitallisuuden puute eri ratkaisujen välillä sekä integraatiohaasteet muun tietoturvainfrastruktuurin, kuten SIEM-järjestelmien, kanssa. Lisäksi UEBA:n tutkimuskenttä kärsii edelleen teoreettisen viitekehyksen hajanaisuudesta, sillä lähestymistavat perustuvat usein heuristiikkaan tai kontekstisidonnaisiin käytäntöihin.

UEBA:n tehokkuus on vahvasti riippuvaista käytetyn datan laadusta ja kattavuudesta sekä algoritmien kyvystä mukautua alati muuttuvaan toimintaympäristöön, mikä on koettu monissa tutkimuksissa haasteeksi. Tehokkaimmat järjestelmät vaativat pidemmälle kehittynyttä datan esikäsittelyä, tehokkaita laskentamenetelmiä ja syklistä mallien optimointia, jotta analyysin nopeus ja luotettavuus säilyvät hyväksyttävällä tasolla. (Raj ym., 2019; Rashid ym., 2021; Khan ym., 2022.)

Vaikka UEBA-teknologian hyödyntäminen kasvaa nopeasti, tutkimuskentässä on edelleen selkeitä aukkoja. Vertailut eri menetelmien ja järjestelmien välillä ovat rajallisia, mikä vaikeuttaa objektiivisempaa arviota sovellettavimmista ja parhaista käytännöistä (Grand View Research, 2023; Tang ym., 2017). Standardoidut arviointimittarit puuttuvat, jonka takia UEBA-ratkaisujen todellinen arvo on usein konteksti- ja ympäristöriippuvaista (Ahmed ym., 2016; Tang ym., 2017). Lisäksi datan anonymisointi ja yksityisyyden suoja tuottavat jatkuvia käytännön ja eettisiä haasteita, erityisesti herkkiä henkilötietoja analysoitaessa (Chio & Freeman, 2019; Rashid ym., 2021).

UEBA ja käyttäjäanalyysi tarjoavat lupaavia mahdollisuuksia kehittyneen kyberturvallisuuden rakentamiseksi erityisesti sisäisten ja vaikeasti ennakoitavien uhkien torjunnassa. Tutkimusnäytön perusteella niiden tehokkuus on kuitenkin edelleen vahvasti riippuvainen käytettyjen mallien laadusta, datan kattavuudesta ja järjestelmien joustavuudesta nopeasti muuttuvassa

toimintaympäristössä. (Tang ym., 2017; Rashid ym., 2021; Khan ym., 2022.) Tulevaisuuden tutkimuksen tulisi vahvemmin suuntautua suorituskyvyn objektiiviseen ja systemaattiseen vertailuun, datan esikäsittelyn kohentamiseen sekä automaattisen mallinpäivityksen ja yksityisyydenhallinnan ratkaisuihin.

3.1.3 Uhkamallinnus ja ennakoiva riskianalyysi

Uhkamallinnus ja ennakoiva riskianalyysi ovat muodostuneet keskeisiksi välineiksi kyberuhkien ennustamisessa ja kokonaisvaltaisessa riskien hallinnassa. Uhkamallinnuksessa tavoitteena on tunnistaa mahdolliset uhkatilanteet etukäteen analysoimalla uhkien syntyprosesseja ja niiden vaikutuksia organisaatioon (Seaman ym., 2022). Yksiselitteistä määritelmää uhkamallinnukselle ei ole mutta eräs suosituimmista kuvaa uhkamallinnuksen prosessina, jota voidaan käyttää analysoimaan mahdollisia hyökkäyksiä tai uhkia, ja sitä voidaan tukea uhkakirjastoilla tai hyökkäystaksonomioilla (Uzunov ym., 2014). Kyberturvallisuuden uhkamallinnuksessa hyödynnetään järjestelmällistä ja rakenteellista lähestymistapaa, jonka tavoitteena on tunnistaa järjestelmän mahdolliset uhkat, haavoittuvuudet ja hyökkäysreitit. Uhkamallinnus voidaan jakaa karkeasti kahteen pääkategoriaan: manuaaliseen ja automatisoituun mallinnukseen. Molempiin lähestymistapoihin voidaan nykyään liittää koneoppimisen ja syväoppimisen menetelmien soveltaminen, mikä mahdollistaa uhkien tunnistamisen laajemmin ja tehokkaammin kuin perinteisillä keinoilla (Xiong ym., 2019). Vaikka monet uhkamallinnusmenetelmät perustuvat yhä ihmisasiantuntijoiden käsin tekemiin arvioihin ja mallinnuksiin, automaatiota pidetään kiistatta merkittävänä tulevaisuuden kehityssuuntana (Saulaiman, 2025). Jo varhain järjestelmän suunnitteluvaiheessa toteutettu järjestelmän arkkitehtuurin tarkka mallintaminen on mahdollistanut riskialttiiden kohtien systemaattisen dokumentoinnin ja arvioinnin. Näin turvallisuusasiantuntijat ja järjestelmäkehittäjät saavat selkeän kokonaiskuvan siitä, millä eri tavoilla hyökkääjä voisi etenemismahdollisuuksiaan hyödyntää järjestelmän heikkouksien kautta. Mallintaminen auttaa tunnistamaan sellaisia monitahoisia uhkaskenaarioita ja hyökkäyskombinaatioita, jotka saattaisivat jäädä tyystin huomaamatta, mikäli arviointi perustuisi vain perinteiseen, pääosin testaukseen nojaavaan riskianalyysiin. (Xiong ym., 2019; Saulaiman ym., 2025.)

Graafipohjaiset mallinnusmenetelmät, kuten hyökkäys- ja uhkakäyrät, ovat osoittautuneet erityisen hyödyllisiksi monimutkaisten järjestelmien tarkastelussa. Graafit mahdollistavat järjestelmän komponenttien välisen tietovirran ja riippuvuuksien havainnollistamisen sekä hyökkäyspolkujen kartoittamisen yksityiskohtaisesti. Tämän ansiosta voidaan tunnistaa kriittiset komponentit, joiden kautta hyökkääjä todennäköisimmin pystyy etenemään järjestelmässä. Graafipohjainen

lähestymistapa soveltuu erityisesti laajoihin ja kompleksisiin ympäristöihin, joissa manuaalinen riskianalyysi ei enää skaalaudu järjestelmän kaikkiin mahdollisiin hyökkäyspintoihin. (Meng ym., 2025; Saulaiman ym., 2025; Xiong ym., 2019; Xu ym., 2012.)

Uhkamallinnuksen keskeisenä tavoitteena on muodostaa kattava ja priorisoitu kokonaiskuva järjestelmän todennäköisimmistä uhkakuvista ja hyökkäyspoluista. Näin eri riskiskenaariot voidaan asettaa järjestykseen niiden todennäköisyyden ja vaikutusten perusteella, mikä mahdollistaa tietoturvallisuuden kohdentamisen potentiaalisesti merkittävimpiin haavoittuvuuksiin. Resurssit voidaan näin ollen jakaa tehokkaasti ja puolustusratkaisut suunnata niihin kohtiin, joissa niille on suurin tarve. (Šijan, 2024.)

Perinteisesti uhkamallinnuksen, erityisesti hyökkäysgraafien, laatiminen on ollut pitkälti manuaalista, aikaa vievää ja altistunut näin ollen myös ihmisten tekemille virheille. Monimutkaisten järjestelmien ja kasvavan uhkien määrän vuoksi tällainen lähestymistapa on kuitenkin tullut tiensä päähän nykyisessä uhkaympäristössä. Koneoppimisen ja syväoppimisen menetelmät, kuten graafineuroverkot ja konvoluutioneuroverkot, mahdollistavat uhkien ja hyökkäyspolkujen automaattisen tunnistamisen ja analysoinnin sekä historiallisesta että reaaliaikaisesta datasta. (Meng ym., 2025; François ym., 2025.)

Meng ym. (2025) rakensivat hyökkäysgraafit koneoppimismenetelmien avulla käyttäen historialliseen verkkoliikenteeseen perustuvaa datasarjaa. Heidän mallinsa tunnisti tunnettuja hyökkäyspolkuja hyvin, mutta sen haasteena oli alttius väärille positiivisille tulkinnoille epätäydellisissä ympäristöissä, jossa mallin käytössä oleva data on erinäisin tavoin puutteellista verrattuna todelliseen verkkoympäristöön.

François ym. (2025) puolestaan hyödynsivät syväoppimiseen perustuvia graafineuroverkkoja rakentamaan dynaamisia hyökkäysgraafeja, joita pystyttiin päivittämään reaaliaikaisella verkko- ja lokidatalla. Heidän lähestymistapansa osoitti vahvaa kykyä tunnistaa uusia, aiemmin tuntemattomia hyökkäyspolkuja erityisesti monimutkaisissa ja suurissa järjestelmissä, malli saavutti suurimman tunnistustarkkuuden nollapäivähyökkäysten kaltaisissa hyökkäyksissä. Tutkimuksessa kuitenkin havaittiin, että menetelmä vaatii suuren määrän monipuolista opetusaineistoa sekä lisäkeinoja, joilla voidaan vähentää harhaanjohtavien, epäluotettavien yhteyksien automaattista muodostumista graafiin tilanteissa, joissa alkudata on puutteellista tai erittäin heterogeenistä.

Tekoälyn hyödyntäminen tuo uhkamallinnukseen ja riskianalyysiin jatkuvuutta sekä päivittyvyyttä. Syväoppimismenetelmillä voidaan simuloida myös uusia, vielä realisoitumattomia uhkia sekä

arvioida puolustusratkaisujen tehokkuutta dynaamisissa ja muuttuvissa olosuhteissa. Tällainen lähestymistapa on kriittinen erityisesti laajoissa ja nopeasti muuttuvissa verkko- ja järjestelmäkokonaisuuksissa, kuten esimerkiksi 5G-teknologiaa hyödyntävissä ajoneuvoissa, joissa hyökkäyspinta on erityisen suuri (Saulaiman ym., 2025).

Uhkamallinnuksen ja ennakoivan riskianalyysin merkitys korostuu yhä enemmän kyberturvallisuuden saralla, sillä nykyaikaiset organisaatiot tarvitsevat keinoja tunnistaa ja torjua monimutkaisia ja jatkuvasti kehittyviä uhkia. Näiden tavoitteiden saavuttaminen edellyttää siirtymistä perinteisistä, reaktiivisista ja manuaalisista menetelmistä kohti ennaltaehkäiseviä, automaattisia ja tekoälypohjaisia lähestymistapoja, joissa koneoppimisen ja syväoppimisen menetelmät ovat keskeisessä asemassa (Shi ym., 2025; Junquera ym., 2024).

Sun ym. (2022) ja Shi ym. (2025) osoittavat, että syväoppimismallit kykenevät tunnistamaan verkko- ja lokidatasta poikkeavuuksia, jotka voivat jäädä huomaamatta perinteiseltä analyytikolta vaikkakin lähestymistavoissa on selviä eroja. Shi ym. (2025) tarkastelevat haitallisen verkkoliikenteen tunnistamista hyökkäyspintojen näkökulmasta hyödyntäen verkon topologiaa, eli sitä, miten verkon eri osat, laitteet ja yhteydet ovat järjestyneet ja linkittyneet toisiinsa. Sun ym. (2022) taas painottavat käyttäjäkohtaisen historian ja käyttäytymisen hyödyntämistä riskiluokittelun perustana. Heidän näkemyksensä mukaan pelkkä teknisiin ominaisuuksiin perustuva arviointi ei riitä kyberriskin ennakointiin, vaan myös käyttäjän toimintahistoria kuten kirjautumistottumukset ja käyttöajat, täydentävät analyysiä. Malli osaa tunnistaa poikkeavan käyttäytymisen kuten epätavalliset kirjautumispaikat ja -ajat, ja käyttää näitä tietoja yksilöllisen riskiprofiilin tekemiseen, jolloin riskiluokituksesta tulee dynaamisempaa ja paremmin käyttäjäkohtaisia uhkia havainnoivaa. Näiden lähestymistapojen yhdistäminen voisi tuoda merkittäviä hyötyjä, mutta on toistaiseksi ollut vähäistä.

Tekoäly tukee myös uhkatiedustelua monin tavoin. Se kykenee automatisoimaan kyberuhkatiedusteluraporttien, kuten avoimen lähdekoodin uhkatiedon, jatkuvan haun ja integroinnin, mikä pitää analyysin ajantasaisena (Pantelis ym., 2021). Lisäksi syväoppimistekniikoita hyödyntämällä on mahdollista tunnistaa relevantteja uhkatietoja myös sellaisista epätyypillistä lähteistä kuin sosiaalinen media, jossa haavoittuvuustiedot leviävät nopeasti (Iorga ym., 2021). Koneoppimismenetelmien avulla pystytään myös tunnistamaan ajallisia trendejä ja uusia nousevia teemoja, jotka edesauttavat uhkien hallinnan ennakoivuutta ja tukevat strategista puolustussuunnittelua (Kim ym., 2020). Uhkatiedustelun kehityksestä huolimatta tekoälyteknologioiden soveltamiseen liittyy myös tiettyjä rajoitteita. Ratkaisevaa on, että käytettävä

data on korkealaatuista ja riittävän monipuolista. Heikkolaatuinen tai puutteellinen data voi johtaa virheellisiin päätelmiin, mikä vaikuttaa negatiivisesti koko uhkatiedustelun laatuun. (Marchal ym., 2024.)

Haavoittuvuuksien hallinnan näkökulmasta Huff ym. (2021) ja Jeon & Kim (2021) painottavat tekoälymenetelmien kykyä tehostaa automaatiota ja varmistaa nopea reagointi nouseviin ohjelmistohaavoittuvuuksiin. Tulokset perustuvat systemaattisiin vertailuihin, joissa tekoäly analysoi suuria haavoittuvuuskantoja, kuten CVE-tietokantaa ja ohjaa riskien priorisointia luonnollisen kielen käsittelyn ja syväoppimisen avulla. Sekä Huff ym. että Jeon & Kim tuovat esiin, että syväoppimismallit osoittavat lupaavaa suorituskykyä automaattisessa haavoittuvuuksien tunnistuksessa ja luokittelussa, mutta sovellukset ovat useimmiten vielä kokeilu- ja kehitysasteella. Tekoäly on jo nyt analyytikoiden apuna tunkeutumistestauksessa ja laskee manuaalisen haavoittuvuusanalyysin kustannuksia. Generatiivisten tekoälymenetelmien kehittyessä ne voivat tulevaisuudessa ottaa merkittävämpää roolia haavoittuvuuksien hallinnassa. (Marchal ym., 2024.)

Saha ym. (2021) nostavat esiin, että tekoälyllä on laajat käyttömahdollisuudet erityisesti toimitusketjun riskienhallinnassa, jossa suurten tietomassojen analysointi ja automaattisella mallinnuksella löydettävät riippuvuudet ovat ratkaisevassa asemassa monimutkaisten uhkien hallinnassa. Esimerkiksi voidaan löytää monimutkaisia riippuvuuksia ja kriittisiä tekijöitä, kuten toimitusviiveitä, joita olisi vaikeaa tunnistaa manuaalisesti. Myös tässä tutkimuksessa painottuu kuitenkin tekoälyratkaisujen merkittävä riippuvuus laadukkaasta syötedatasta sekä mallien luotettavuudesta.

Xiong ja Lagerström (2019) toteuttivat systemaattisen kirjallisuuskatsauksen, jossa analysoitiin 54 keskeistä tutkimusta uhkamallinnuksesta vuosilta 2007–2018. Tutkimuksen perusteella voidaan sanoa alan olevan edelleen kehittyvä ja jakautunut omiin osa-alueisiinsa, mikä näyttäytyy erityisesti menetelmien ja arviointikäytäntöjen hajanaisuutena. Suurin osa tarkastelluista malleista perustuu manuaaliseen työskentelyyn, mikä rajoittaa niiden skaalautuvuutta ja soveltuvuutta nopeaa reagointia vaativiin uhkiin.

Katsauksessa havaittiin, että graafiset menetelmät, kuten hyökkäyspuut ja tiedonvirtakaaviot, ovat edelleen keskeisimpiä välineitä uhkamallinnuksessa. Menetelmien suosio juontuu niiden intuitiivisuudesta ja havainnollisuudesta, mutta ne kärsivät haasteista, kuten vaikeudesta kuvata järjestelmien monimutkaisia ja dynaamisia uhkakuva-alueita. Vastaavasti formaaliset menetelmät, joita käytetään mallien tarkemmassa analysoinnissa, ovat marginaalisessa asemassa, mikä johtuu osittain niiden korkeamman oppimiskynnyksen ja soveltamisvaikeuksien vuoksi.

Automaation vähäisyys uhkamallinnuksessa on myös keskeinen ongelma. Xiong ja Lagerström (2019) huomauttavat, että manuaaliseen mallinnukseen liittyy merkittävää työvoimakustannusta ja riski ihmisen virheistä, mikä heikentää mallien ajantasaisuutta ja luotettavuutta. Lisäksi kvantitatiivisten validointimenetelmien puute heikentää kykyä arvioida mallien tehokkuutta objektiivisesti. Tämä tuo rajoitteita eri menetelmien vertailtavuuteen ja uusien ratkaisujen kehityksen ohjaamiseen.

Aihealueen moniulotteisuus piilee myös siinä, että mallinnuskohteet ja uhkaluokat ovat erittäin heterogeenisiä. Vaikka eri tutkimukset kattavat laajan joukon sovellusalueita aina ohjelmistosovelluksista kriittisiin infrastruktuureihin, yhtenäistä lähestymistapaa ei ole vielä pystytty muodostamaan. Tämä vaikeuttaa standardoitujen käytäntöjen syntymistä ja rajoittaa tutkimuksen ja käytännön yhteistyötä.

Reaaliaikainen uhkamallinnus ja riskianalyysi esitellään potentiaalisina ratkaisuina näihin ongelmiin, mutta niiden kehittäminen edellyttää merkittäviä teknisiä innovaatioita automaation, datan integroinnin ja analyysimenetelmien skaalautuvuuden lisäämiseksi. Myös organisaatioiden prosessien ja päätöksenteon tulisi mukautua hyödyntämään näitä uusia työkaluja mahdollisimman tehokkaasti.

Kriittisesti tarkasteltuna katsauksen tulokset nostavat esiin useita alan kompastuskiviä, jotka liittyvät erityisesti metodologiseen hajanaisuuteen, puutteelliseen standardisointiin ja vähäiseen automaatioon. Nämä seikat nostavat esiin tarpeen systemaattiselle lähestymistavalle, joka yhdistää teoreettisen kehityksen ja käytännön sovellukset. Vaikka Xiong ja Lagerström (2019) tarjoavat kattavan yleiskuvan, katsauksessa olisi voinut lisätä syvempää analyysiä eri mallien suorituskyvystä ja käytännön vaikutuksista.

Cheimonidis ym. (2023) sekä Bratsas ym. (2024) ovat toteuttaneet erilliset, mutta toisiaan täydentävät systemaattiset kirjallisuuskatsaukset kahdesta kyberturvallisuuden osa-alueesta: reaaliaikaiseen riskianalyysiin luettavasta dynaamisesta riskien arvioinnista (engl. Dynamic Risk Assesment, DRA) ja uhkatiedustelusta (engl. Cyber Threat Intelligence, CTI). Näiden katsauksien vertaileva tarkastelu tuo esiin paitsi lähestymistapojen eroja, myös systemaattisen katsausmenetelmän käyttöä tiedonkeruussa, analyysissa ja synteesissä.

Cheimonidis ym. (2023) analysoivat 50 vertaisarvioitua tutkimusta vuosilta 2001–2023, jotka esittelivät tai sovelsivat DRA-malleja. Artikkelit valittiin systemaattisen hakuprosessin tuloksena, ja ne luokiteltiin kolmen kriteerin mukaan. Ensimmäisenä oli käytetty analyysimenetelmä eli vaikkapa

koneoppiminen tai matemaattiset mallit, toisena sovellusalue eli toimitusketjujen hallinta, teollisuuden ohjausjärjestelmät, tieto- ja viestintäteknologia ja mallin kypsyys idean tasosta toteutuneeseen malliin. Tutkimukset eriteltiin myös siihen perustuen, miten ne hyödynsivät riskinarvioinnin syötteitä eli IDS-lokeja, haavoittuvuustietokantoja, asiantuntija-arvioita ja muita lähteitä. Katsauksen systemaattisuus näkyy erityisesti siinä, miten jokaiselle mallille on määritelty eksplisiittiset arviointikriteerit, joiden avulla pystytään vertailemaan mallien kypsyyttä, reaaliaikaisuutta ja datalähtöisyyttä.

Bratsas ym. (2024) puolestaan analysoivat 74 tutkimusta, jotka käsittelivät semanttisten teknologioiden käyttöä uhkatiedustelussa. Katsauksessa tarkasteltiin erityisesti tietämysverkkoja (engl. Knowledge graphs) ja ontologioita eli eräänlaisia jäsentämismenetelmiä uhkatiedustelun jäsentämiseen ja hyödyntämiseen. Kirjallisuushaku perustui useisiin hakutermeihin ja rajattiin vertaisarvioituihin artikkeleihin vuosilta 2010–2023. Tutkimukset luokiteltiin ontologiatyypin, lähdeaineistojen kuten MITRE:n mukaan, käyttötarkoituksen eli uhkien tunnistuksen, tiedon yhdistämisen ja automaation sekä validointimenetelmän kuten empiirisen testauksen mukaan. Katsausrakenteen perustana on laaja vertailtavuus. Eri ontologioita pyrittiin arvioimaan suhteessa kykyyn jäsentää uhkatietoa, integroida lähteitä ja tukea reaaliaikaista päätöksentekoa.

Kummassakin katsauksessa tunnistetaan merkittäviä metodologisia haasteita. Dynaamisen riskien hallinnan kirjallisuudessa ongelmallisiksi osoittautuivat mallien reaktiivisuus, selitettävyys ja rajallinen syötedatan monipuolisuus. Uhkatiedustelun kirjallisuudessa taas havaittiin standardoinnin puute, tiedon semanttinen epäyhtenäisyys ja rajallinen integrointi operatiivisiin järjestelmiin. Näitä puutteita analysoitiin molemmissa katsauksissa kriittisesti, ja niihin ehdotettiin konkreettisia ratkaisuja, kuten luotettavuusarvioiden sisällyttämistä riskienhallinnan malleihin ja standardien semanttista laajentamista uhkatiedustelumalleissa.

Systemaattisen katsausmenetelmän näkökulmasta molemmat tutkimukset täyttävät keskeiset kriteerit eli eksplisiittisen tutkimuskysymyksen, kattavan lähdehaun, todennettavan sisäänotto- ja poissulkuprosessin, sekä jäsennellyn tulosten analyysin. Bratsas ym. painottavat katsauksessaan myös ontologian elinkaaren vaiheita eli suunnittelua, validointia sekä käyttöönottoa, kun taas Cheimonidis ym. korostavat enemmän mallien operatiivista kypsyyttä ja suorituskykyä. Molemmissa katsauksissa on pyritty synteesiin, joka tuo esiin paitsi nykytilan, myös tutkimukselliset ja käytännölliset kehityskulut.

Nämä kaksi systemaattista kirjallisuuskatsausta osoittavat tutkimusalueidensa olevan aktiivisia, mutta rakenteellisesti kehittymättömiä. DRA-mallit hyötyisivät CTI:n semanttisista rakenteista ja

tiedon yhdistämisen menetelmistä, kun taas CTI-mallinnus voisi integroitua paremmin operatiivisiin riskinarviointijärjestelmiin, jos sen tiedonrakenteet yhdistetään osaksi reaaliaikaista päätöksentekoa. Systemaattinen katsausmenetelmä on näissä tutkimuksissa mahdollistanut kehityskohteiden tunnistamisen, mutta jatkossa tarvitaan myös vertailevaa meta-analyysia, joka nousisi yksittäisten näkökantojen yläpuolelle ja pyrkisi yhdistäviin, monitieteellisiin malleihin.

Kaikista systemaattisista kirjallisuuskatsauksista, jotka käsittelevät uhkamallinnusta, uhkatiedustelua sekä ennakoivaa riskianalyysiä, nousee esiin yhteneväisiä huomioita. Tiedon laatu ja heterogeenisuus sekä sen integrointi muodostavat keskeisen teknisen pullonkaulan kaikilla näistä alueista. Automatisoinnin ja koneoppimisen nopea kehitys tarjoaa mahdollisuuksia laaja-alaisesti uhkamallinnuksessa ja ennakoivassa riskianalyysissä, mutta samalla tulee ottaa huomioon eettiset, yksityisyyteen liittyvät sekä väärien positiivisten minimointiin liittyvät kysymykset. Myös organisaatiokohtaiset toimintaympäristöt ja prosessit vaikuttavat merkittävästi teknologioiden soveltuvuuteen ja tehokkuuteen, mikä puolestaan luo tarpeen räätälöidyille ratkaisuille ja joustaville kehityskäytännöille. Uhkamallinnuksen tutkimus on vielä kehitysvaiheessa, ja seuraavat tutkimusaskeleet vaativat standardoitujen ja automaattisten ratkaisujen kehittämistä erityisesti reaaliaikaisen uhkien hallinnan vahvistamiseksi.

3.1.4 Käytännön esimerkkejä yrityksiltä ja toteutuksista

Tekoälyyn ja koneoppimiseen pohjautuvien kyberturvallisuusratkaisujen käyttöönotto osoittautuu usein moniportaiseksi ja huolellista valmistelua vaativaksi prosessiksi teknologian ja organisaation näkökulmasta (Abdeldaiem, 2024). Tekoälyprojekteilla on korkea epäonnistumisriski, jota selittävät tuotantoympäristöjen monimutkaisuus, datan jatkuva muuttuminen (engl. data drift) ja se, että suuri osa projekteista osoittautuu resurssien kannalta liian raskaaksi tai kalliiksi (Yusuff, 2023; Marchal ym., 2024). Onnistuneen ratkaisun edellytyksenä on selkeä, liiketoimintalähtöinen ongelmanmäärittely. Tekoälyn odotettu lisäarvo perinteisiin ratkaisuihin nähden tulee arvioida todenmukaisesti, ja tavoitteet pitää pystyä muuntamaan helposti seurattaviksi suorituskykymittareiksi (engl. Key Performance Indicator, KPI), joiden toteutumista voidaan sitten monitoroida koko käyttöönoton ajan (Gilbert, 2024).

Teknisen toteutettavuuden ja integraation arviointi olemassa oleviin järjestelmiin tulisi aloittaa varhaisessa vaiheessa. Datan laadun ja määrän jatkuva seuranta on yhtä lailla välttämätöntä, sillä koneoppimismallit ovat täysin riippuvaisia oppimisdatasta (Razzaq ym., 2025). Muutokset datan rakenteessa tai määrässä voivat heikentää järjestelmän tarkkuutta nopeasti, ellei mallia päivitetä säännöllisesti (Sarker ym., 2020). Onkin siis erittäin suositeltavaa toteuttaa käyttöönotto ensin

rajatussa laajuudessa pienimuotoisena pilotointina, jotta ongelmat tunnistetaan varhaisessa vaiheessa ja kehitysprosessi pysyy kustannustehokkaana.

Liiallista teknologista monimutkaisuutta tulisi välttää. Usein yksinkertaisemmat koneoppimismallit ovat sekä realistisempia että taloudellisempia kuin syvät neuroverkot, erityisesti kriittisissä sovelluksissa, joissa vaaditaan manuaalista auditoinnin mahdollisuutta tai hybridiratkaisuja (Veeramachaneni ym., 2016). Parhaiten menestyvät kehityshankkeet ovat niitä, joissa tiimityössä yhdistyy kyberturvallisuuden domain-osaaminen sekä koneoppimisen ja data-analytiikan vahva tuntemus (Marchal ym., 2024).

Jotta tekoälyratkaisut voisivat yleistyä tuotantoympäristöissä, tarvitaan prosessien ja työkalujen vakiointia sekä kehityskirjastoja, jotka mahdollistavat mallien uudelleenkäytettävyyden ja tehokkaan ylläpidon. Yhdistetty tietovaranto, kattavat monitorointivälineet ja helposti hallittava infrastruktuuri lisäävät toistettavuutta ja mittakaavaetuja. Näin voidaan toteuttaa kustannustehokkaita tekoälyprojekteja monipuolisissa kyberturvallisuuden käyttökohteissa (Marchal ym., 2024).

Koneoppimisen ja syväoppimisen menetelmien hyödyntäminen parantaa kyberturvallisuuden kykyä havaita ennakoivasti uusia ja monimutkaisia uhkia. Järjestelmät, jotka yhdistävät automaattisen poikkeavuuksien tunnistuksen ja asiantuntijoiden palautteen, saavuttavat korkean tunnistustarkkuuden pystyen samalla rajoittamaan väärien positiivisten määrää (Vähäkainu ym., 2018). Kyseessä on merkittävä parannus perinteisiin sääntö- ja allekirjoituspohjaisiin ratkaisuihin verrattuna. Alla kuvassa 1 kootusti tekoälyn hyödyntämisen keskeiset edellytykset yritysten kyberturvallisuudessa.

Kuva 1 Keskeiset edellytykset tekoälyn hyödyntämiseen yritysten kyberturvallisuudessa (koostettu lähteistä Marchal ym., 2024; Vähäkainu ym., 2018)

PatternEx AI2 on esimerkki järjestelmästä, jossa yhdistetään ohjaamaton ja ohjattu oppiminen ihmisanalyytikoiden tukemana. Alusta analysoi valtavat määrät verkkolokidataa havaitakseen poikkeavaa käyttäytymistä ja klusteroi datan itsenäisesti merkityksellisiin malleihin. Tällaiset järjestelmät hyödyntävät ohjaamatonta oppimista säädelläkseen poikkeavuuksia. Niiden merkityksen arvioi lopulta asiantuntija, jonka palautetta hyödynnetään jatkossa ohjatussa mallinnuksessa. PatternEx AI2:n iteratiivinen rakenne mahdollistaa järjestelmän ripeän kehittymisen sekä tunnistusprosessin hienosäädön, jolle on tunnusomaista väärien positiivisten hälytysten määrän systemaattinen pieneneminen järjestelmän oppimisen edetessä. Kun järjestelmä aluksi tuottaa runsaasti hälytyksiä asiantuntijakäsittelyyn, alkaa niiden määrä laskea merkittävästi nopeasti saavutetun oppimissyklin ansiosta, mikä puolestaan vapauttaa analyytikkojen resursseja. Testitulokset osoittavat AI2:n mahdollistavan jopa kolminkertaisen parannuksen tunnistustarkkuudessa aiempiin ratkaisuihin verrattuna. (Veeramachaneni ym., 2016; vrt. Vähäkainu ym., 2018.)

Amazon Macie on hyvä esimerkki siitä, miten koneoppimisen menetelmiä voidaan hyödyntää sensitiivisen tiedon suojaamisessa pilvipalveluissa. Macie luokittelee Amazon S3 -ympäristön datan automaattisesti ja tunnistaa yksityisyyden kannalta herkäksi määriteltävää tietoa, kuten

henkilötietoja. Palvelu rakentaa käyttäytymisprofiileja ja tarkkailee sekä tiedostojen että käyttäjien toimintaa, tarkastellen muun muassa tiedonsiirtoja, kopiointeja ja käyttöoikeuksien muutoksia. Macie kykenee automaattisesti tuomaan esiin poikkeamat, jonka jälkeen se laukaisee hälytyksiä potentiaalisista tietovuodoista tai luvattomista toimista. Tällainen automatisoitu käyttäytymisanalyysi mahdollistaa varhaiset varoitukset ja antaa organisaatiolle mahdollisuuden reagoida nopeasti riskeihin. (Amazon Macie, 2025; Vähäkainu ym., 2018.)

Cyberlytic Profilerin esimerkki painottaa ohjaamattoman koneoppimisen ja käyttäytymisen analyysin voimaa verkkoliikenteen riskien arvioinnissa. Profiler analysoi HTTP-protokollaan pohjautuvaa verkkoliikennettä ja vasteita, rakentaen yksilöllisiä malleja siitä, mikä on tavanomaista kyseisessä verkkoympäristössä. Muutos alkuperäiseen mahdollistaa poikkeavien ilmiöiden nostamisen esiin, ja analysointiin hyödynnetään myös puolittain ohjattuja menetelmiä erilaisten hyökkäysmuotojen tunnistamiseen. Verkkosovelluksia profiloimalla Profiler kykenee tunnistamaan, ovatko lähetetyt pyynnöt peräisin tavanomaisesta tietyn verkkosovellusalueen sovelluksen jakelusta. Tämä lähestymistapa määrittää, miltä normaali näyttää jollekin tietylle organisaatiolle. Sen seurauksena päätetään, mikä voi aikaansaada poikkeuksellista liikennettä alkuperäiseen lähtötilanteeseen verrattuna. Anomalioita tunnistamalla voidaan painottaa erittäin todennäköistä häiriötoimintaa, jonka perusteella voidaan tehdä riskiarvioita. Profiler tarjoaa myös selainpohjaisen kojelautanäkymän reaaliajassa, mahdollistaen näin kokonaistilanteen, uhkatyylien ja ajallisten trendien jatkuvan seurannan. Järjestelmä kykenee erottelemaan hyökkäyksiä esimerkiksi monimutkaisuuden, kyvykkyyden ja tehokkuuden perusteella, tuottaen syvällistä tietoa esimerkiksi siitä, käyttääkö hyökkäyksen toteuttaja automatisoituja työkaluja vai onko kyseessä inhimillinen toimija (Vähäkainu ym., 2018).

Yhteistä kaikille edellä mainituille esimerkeille on niiden iteratiivisuus, jatkuva oppiminen ja kyky suodattaa tietomassoista oleellinen uhka- ja riskitieto luotettavalla tavalla. Automatisoitu poikkeavuuksien analyysi, koneoppimisen ja syväoppimisen tehokas hyödyntäminen sekä asiantuntijapalautteen ja selittävyyden tuominen järjestelmiin nostavat ennakoivan kyberturvallisuuden valmiutta huomattavasti perinteisiä ratkaisuja korkeammalle. Kriittistä on, että järjestelmät on rakennettu skaalautuviksi datamäärän suhteen, ja ne mahdollistavat nopean sopeutumisen uusiin uhkakuvioihin ja toimintaympäristöihin (Vähäkainu ym., 2018; Meng ym., 2023).

4 Tekoälyn tulevaisuus yritysten ennaltaehkäisevässä kyberturvallisuudessa

Koneoppimisen ja syväoppimisen tekoälyteknologioiden käyttöönotto yritysten ennaltaehkäisevässä kyberturvallisuudessa on viimeisen vuosikymmenen aikana muuttanut ratkaisevasti kyberuhilta puolustautumisen paradigmaa. Traditioista, staattisiin sääntöihin ja allekirjoituksiin perustuvista ratkaisuista on siirrytty kohti automatisoitua, sopeutuvaa ja skaalautuvaa suojautumista, jossa korostuu datan laajamittaisen käsittelyn, oppivan mallinnuksen ja ihmisen sekä koneen yhteistyö (Mohamed ym., 2025; Sarker ym., 2020). Nykyaikaiset hyökkäykset ovat yhä vaikeammin tunnistettavissa, monimuotoisempia ja luonteeltaan dynaamisia, minkä vuoksi pelkkä reagoiminen niihin ei enää riitä. Järjestelmien on kyettävä ennakoimaan, tunnistamaan poikkeamat ja osattava tunnistaa jopa uusia hyökkäystyyppejä ennen vahingon toteutumista (Hassanien ym., 2021; Kim ym., 2020).

Tekoälypohjaisten järjestelmien kyvyt ennaltaehkäisevässä kyberturvallisuudessa ylittävät monin paikoin perinteiset menetelmät, eritoten haittaohjelmien ja poikkeavan verkkoliikenteen tunnistuksessa sekä uhkatiedustelun ennakoinnissa (Jeon & Kim, 2021; Pantelis ym., 2021). Uudessa tilanteessa nousee tarve monikerroksisille, itsenäisesti oppiville ja kontekstuaalisille ratkaisuille, joissa yhdistyvät erilaiset tietolähteet, automaatio ja ihmisanalyysin tuottama lisäarvo (Saha ym., 2021; Deng ym., 2023).

Uutena, merkittävänä murrosvaiheena voidaan pitää suurten kielimallien (engl. Large Language Model, LLM) integrointia ja sulauttamista kyberturvallisuuteen. Näiden mallien kyky käsitellä ja tuottaa luonnollista kieltä tuo uusia mahdollisuuksia tiedon yhdistelyssä, selittämisessä ja vuorovaikutuksessa ihmisen kanssa. Suuret kielimallit tehostavat erityisesti uhkatiedustelua, turvallisuusanalyysejä ja tietoturvakoulutusta, sillä ne kykenevät taustoittamaan suuria tietomassoja sekä nostamaan järjestelmien ymmärrettävyystasoa ja selitettävyyttä loppukäyttäjille (Marchal ym., 2024; Xu ym., 2024). Lisäksi suuret kielimallit mahdollistavat turvallisuusprosessien läpinäkyvyyden. Analyytikot voivat ymmärtää, miksi jokin toimi estettiin tai tietty tiedosto luokiteltiin haitalliseksi, mikä edesauttaa luottamuksen syntymistä järjestelmiin. (Xu ym., 2024.)

Kuitenkin samalla korostuvat riskit ja rajoitteet, jotka on ratkaistava, jotta tekoälyn täysi potentiaali pystyttäisiin hyödyntämään. Suurten kielimallien kyky tunnistaa ja havainnoida uhkia on rajallinen verrattuna syvästi teknisiin ratkaisuihin, kuten IDS:ään. Niiden arvo on enemmän analyysin, yhteenvetojen ja selitysten tuotannossa kuin esimerkiksi päätelaitetason havaintojen

automatisoinnissa (Vähäkainu ym., 2018). Suurten kielimallien hallusinaatioherkkyys rajoittaa niiden käyttöä automatisoiduissa uhkien torjuntaprosesseissa. Hallusinaatiolla tarkoitetaan virheellisiä tai harhaanjohtavia vastauksia, joita mallit voivat tuottaa, erityisesti monimutkaisissa ja dynaamisissa kyberturvallisuusympäristöissä. Tuotettujen vastausten luotettavuus ei vielä kaikissa tilanteissa riitä täysautomaatioon (Marchal ym., 2024; Xu ym., 2024). Lisäksi mallien alttius vihamielisille hyökkäyksille, uhkaa järjestelmien eheyden ja tietosuojan toteutumista. Esimerkiksi Mallin saastutus (engl. Model Poisoning) voi manipuloida tekoälymallin koulutusdataa, jolloin malli oppii virheellisiä kaavoja ja altistuu takaporttihyökkäyksille (engl. Backdoor attacks), joissa malli toimii pääosin normaalisti mutta mutta tietyissä olosuhteissa antaa hyökkääjän haluaman vastauksen. Evasion-hyökkäykset puolestaan muuttavat syötettä niin, että haitallinen toiminta pysyy näkymättömänä tekoälyjärjestelmälle. Näistä voidaan nostaa adversaariset esimerkit, joissa hyökkääjä muokkaa syötettä niin hienovaraisesti, että malli tekee virheellisen luokituksen. (Paracha ym., 2024.)

Yhdeksi keskeiseksi havainnoksi tämän tutkimuksen pohjalta nousee sääntelyn ja hallitun käyttöönoton kasvava tarve. Sääntelykeskustelu eli niin Euroopan unionin tekoälyasetus, tietosuoja-asetukset (GDPR) kuin kansainvälisesti kehittyvät tekniset standardit, asettaa lähivuosina uusia rajoja ja mahdollisuuksia tekoälyn käytölle kyberturvallisuudessa (Euroopan unionin neuvosto, 2024; NIST, 2023; MITRE, 2023). Organisaatioiden tulee vastaisuudessa pystyä todentamaan mallien toimintaperiaatteet, selitettävyys sekä kyky tunnistaa ja hallita riskejä. Tämän ohella data governance eli tiedon laadun, saatavuuden ja yksityisyydensuojan jatkuva varmistaminen tulee kriittiseksi menestystekijäksi. Tekoälyn rooli yritysten ennaltaehkäisevässä kyberturvallisuudessa korostuu jatkuvasti, kun digitaaliset uhat monimutkaistuvat ja laajenevat. Tulevaisuudessa myös EU:n keskeisessä intressissä on vahvistaa tekoälyinfrastruktuuria ja edistää tekoälyn käyttöönottoa yhteiskunnassa ja yrityksissä. Komission 9.4.2025 julkaisemaa AI Continent Action Plania lainaten: "Tekoälyn käyttöönotto taloutemme keskeisillä sektoreilla on vasta alussa, mutta se auttaa jo nyt ratkaisemaan aikamme kiireellisimpiä haasteita. Vaikka tämän mullistavan muutoksen koko vaikutus on yhä kehittymässä, Euroopan on toimittava kunnianhimoisesti, nopeasti ja ennakoivasti muokatakseen tekoälyn tulevaisuutta tavalla, joka vahvistaa kilpailukykyämme ja suojelee demokraattisia arvojamme" (Euroopan komissio, 2025). Tekoäly ei siis ole vain teknologinen edistysaskel, vaan strateginen voimavara myös kyberturvallisuuden saralla tulevaisuudessa.

Tekoälyteknologioiden, kuten koneoppimisen ja syväoppimisen, yleistyminen yritysten ennaltaehkäisevässä kyberturvallisuudessa on mullistanut tavan, jolla organisaatiot voivat vastata alati monimutkaisempiin ja kehittyvämpiin uhkakuviin (Mohamed ym., 2025). Siirtyminen pelkästä

reaktiivisesta raportoinnista ja sääntöpohjaisesta suojaamisesta kohti sopeutuvia ja ennakoivia syväoppimispohjaisia ratkaisuja on parantanut puolustuksen nopeutta, kattavuutta ja tehokkuutta (Okoli ym., 2024; Vähäkainu ym., 2018; Shi ym., 2025; Junquera ym., 2024).

Yksi keskeisimmistä eduista on koneoppimisen ja syväoppimisen kyky analysoida paljon suurempia ja monimuotoisempia tietomassoja kuin mihin ihminen kykenee. Uhkatiedustelussa koneoppimismenetelmiin pohjautuvat ratkaisut mahdollistavat valtavien määrien sisäisten ja ulkoisten uhkatietosyötteiden järjestelmällisen seulonnan, tärkeiden signaalien esiin nostamisen ja trendien tunnistamisen jo ennen kuin niistä syntyy konkreettinen riski organisaatiolle (Kim ym., 2020; Sarker ym., 2020). Syväoppimismallit kykenevät myös käsittelemään epästrukturoitua dataa, kuten uutisia, foorumikeskusteluja tai sosiaalisen median viestejä (Iorga ym., 2021). Näin ne tunnistavat automaattisesti uusia hyökkäystekniikoita, rikollisia toimijoita ja haavoittuvuuksia, mikä tehostaa merkittävästi organisaation tilannetietoisuutta ja auttaa kohdentamaan resurssit täsmällisemmin (Marchal ym., 2024).

Haitallisen verkkoliikenteen tunnistamisessa koneoppimisen kehittyneet luokittelu- ja anomaliantunnistusmenetelmät ylittävät perinteiset allekirjoituspohjaiset lähestymistavat etenkin, kun yritykset kohtaavat ennen näkemättömiä tai polymorfisia hyökkäysmuotoja (Vähäkainu ym., 2018; Zhao ym., 2024). Syväoppimismallit, kuten toistuvat neuroverkot (RNN) ja konvoluutioneuroverkot (CNN), voivat suoraan analysoida verkon raakadatan ja löytää piileviä riippuvuuksia tai poikkeavuuskaavoja, joita muut menetelmät eivät tavoita. (Fahim ym., 2025; Kim ym., 2016; Yin ym., 2017; Almiani ym., 2019; Kolosnjaji ym., 2016; Radford ym., 2018). Tämä kyvykkyys mahdollistaa esimerkiksi nollapäiväuhkien havaitsemisen nopeasti sekä haitallisten bottiverkkojen, tietomurtojen tai edistyksellisten jatkuvien uhkien varhaisen tunnistamisen (Buczak ym., 2016; Ahmad ym., 2023).

Käyttäjäanalyysissä tekoäly mahdollistaa yksilökohtaisen normaalin toiminnan dynamiikan mallintamisen, jolloin poikkeamat, kuten poikkeavat kirjautumisajat, epätavalliset resurssien käyttökäyttäytymiset tai epäilyttävät tiedonsiirrot, havaitaan automaattisesti ja oikea-aikaisesti. Käyttäytymisanalytiikkaa hyödyntävä UEBA (User and Entity Behavior Analytics) vähentää väärien hälytysten määrää kohdistamalla valvonnan oikeisiin poikkeamiin, ja samalla auttaa löytämään niin sisäpiiririkollisuudesta kuin tilikaappauksista johtuvia uhkia ennen kuin ne eskaloituvat. (Shashanka ym., 2016; Khan ym., 2022; Tang ym., 2017; Marchal ym., 2024.)

Tärkeänä huomiona kone- ja syväoppimisjärjestelmät vaativat jatkuvaa seurantaa, laadukasta opetusaineistoa sekä kykyä mukautua ympäristön muutoksiin. Ne eivät myöskään ole

haavoittumattomia vaan altistuvat muun muassa datan muutosilmiöille, vinoumille sekä uusille vihamielisille hyökkäysvaaroille, jotka kohdistuvat nimenomaan tekoälyyn. (Paracha ym., 2024; Malatji ym., 2024.) Siksi kehitystyötä tulee ohjata vastuullisen tiedonkeruun, mallin läpinäkyvyyden, sääntelyn ja eettisten periaatteiden mukaisesti.

Tekoäly sekä koneoppimisen ja syväoppimisen menetelmät ovat muodostumassa yritysten kyberturvallisuuden selkärangaksi, ei pelkästään teknologisena apuvälineenä, vaan myös strategisena voimavarana, joka tukee organisaatioiden ennaltaehkäisevää sietokykyä nopeasti muuttuvassa uhkaympäristössä (Euroopan unionin neuvosto, 2024; NIST, 2023; MITRE, 2023). Järjestelmien laajan käyttöönoton onnistumisen edellytyksiä ovat selitettävyys, jatkuva validointi, monitieteinen yhteistyö sekä tiiviisti kehittyvä sääntely- ja standardointikehys. Tietoiset ja harkitut panostukset näillä osa-alueilla luovat perustan sille, että yritykset kykenevät hyödyntämään tekoälyn täyden potentiaalin turvallisuutensa tulevaisuuden rakentamisessa. (Marchal ym., 2024; Vähäkainu ym., 2018.)

5 Yhteenveto ja johtopäätökset

Yritysten ennaltaehkäisevä kyberturvallisuus ja erityisesti tekoälyn hyödyntäminen siinä on viime vuosina noussut keskeiseksi tutkimuskohteeksi. Nykyisessä tutkimuksessa ja yrityskäytännöissä ennaltaehkäisevä näkökulma tekoälyn hyödyntämisessä yritysten kyberturvallisuudessa on verrattain uusi. Valtaosa olemassa olevasta kirjallisuudesta ja käytännöistä painottuu ratkaisuihin, joissa tekoälyä hyödynnetään pääosin vasta toteutuneiden uhkien tunnistamiseen ja torjumiseen. Tämä tutkielma pyrkii paikkaamaan tätä tutkimusaukkoa tarjoamalla kokonaisvaltaisen katsauksen siihen, miten tekoälyn ja erityisesti koneoppimisen ja syväoppimisen menetelmiä voidaan soveltaa yritysten ennaltaehkäisevään kyberturvallisuuteen.

Ensimmäinen tutkimuskysymys, johon tutkielmassa vastataan, on millaisia kyberuhkia yritykset kohtaavat ja miten niiltä voidaan suojautua ennaltaehkäisevästi. Yritykset altistuvat hyvin monenlaisille uhille, joihin lukeutuvat muun muassa haittaohjelmat, palvelunestohyökkäykset, tietomurrot ja verkkourkinta. Nämä hyökkäykset voivat kohdistua niin tietojärjestelmiin kuin prosesseihin, ja niillä on usein taloudellisia tai poliittisia motiiveja. Yritysten ennaltaehkäisevä kyberturvallisuus rakentuu nykyaikana teknologisista ratkaisuista, kuten uhkien havaitsemisjärjestelmistä, sekä organisaation tietoturvakulttuurista ja henkilöstön osaamisesta. Keskeistä on kuitenkin kyky havaita uhkia mahdollisimman varhaisessa vaiheessa, minimoida järjestelmien haavoittuvuudet ja kehittää jatkuvasti puolustusvalmiutta. Yritysten on rakennettava kokonaisvaltaisesti päivittyvä ja monipuolinen kyberturvastrategia, jolla vastataan jatkuvasti muuntuviin uhkiin.

Toinen tutkimuskysymys on miten yritykset voivat hyödyntää syväoppimisen ja koneoppimisen menetelmiä ennaltaehkäisevässä kyberturvallisuudessa. Koneoppimisen ja syväoppimisen menetelmät mahdollistavat suurten tietomassojen reaaliaikaisen käsittelyn, mikä tukee uusien uhkien ja poikkeavuuksien tunnistamista, joiden havaitseminen perinteisin menetelmin olisi haastavaa. Käytännön sovellutuksia ovat esimerkiksi haitallisen verkkoliikenteen ja hyökkäysten automaattinen tunnistus, käyttäytymisen poikkeavuusanalyysi sekä uhkamallinnuksen ja riskianalyysin automatisointi. Näissä menetelmissä hyödynnetään edistyneitä malleja, kuten neuroverkkoja (CNN, LSTM), jotka kykenevät analysoimaan monimutkaisia, moniulotteisia datakokonaisuuksia. Tekoäly lisää ennakoivan kyberturvallisuuden tehokkuutta, mutta asettaa samalla uusia vaatimuksia datan laadulle, mallien luotettavuudelle ja jatkuvalle järjestelmäkehitykselle. Alla taulukossa 2 on kuvattuna tarkemmin keskeiset tutkimushavainnot ja tekoälyn hyödyntämisen osa-alueet.

Taulukko 2 Syväoppimisen ja koneoppimisen menetelmien hyödyntäminen ja käytännön sovellutukset

Sovellusalue	Tekoälyn ja syvä-/koneoppimisen menetelmien hyödyt	Menetelmät ja sovellutukset	Käytännön esimerkit	Keskeiset tutkimushavainnot
Haitallisen verkkoliikenteen havaitseminen	Laajojen ja monimuotoisten datavirtojen automaattinen analyysi, poikkeavuuksien nopea tunnistus, ennakoiva reagointi	Syväoppiminen (CNN, RNN), Koneoppiminen (päätöspuut, SVM), anomaliantunnistus	Amazon Macie, PatternEx AI2, Darktrace	Syvä- ja koneoppimisen menetelmät ylittävät signatuuripohjaiset menetelmät uusien, tuntemattomien hyökkäysten tunnistuksessa; mahdollistaa nollapäivä- ja polymorfisten uhkien havaitsemisen.
Käyttäytymisen poikkeavuuksien analyysi (reaaliaikainen seuranta)	Käyttäjäkohtaisen normaalin toiminnan mallinnus, epätyypillisen käytöksen automaattinen tunnistus, sisäpiiriuhkien varhainen havaitseminen	UEBA, anomaliatunnistus, klusterointi, syväoppivat neuroverkot	Cyberlytic Profiler, PatternEx AI2	Syvä- ja koneoppimisen menetelmät vähentävät väärien hälytysten määrää ja tunnistaa myös uudet sisäiset uhat; mahdollistaa yksilöllisten toimintamallien tehokkaan ja ajantasaisen valvonnan.
Uhkamallinnus ja ennakoiva riskianalyysi	Ennustettavuus, resurssien tehokkaampi kohdistus, trendien ja signaalien löytäminen myös epästrukturoidusta datasta	Pattern-analyysi, hybridimallit, NLP, historiallisen uhkadatan analyysi	Amazon Macie (pattern analysis), PatternEx AI2	Syvä- ja koneoppimisen menetelmät mahdollistavat muutosten ja uusien uhkien varhaisen tunnistamisen sekä nopeuttavat reagointia muuttuviin kyberuhkiin ja niiden eskaloitumiseen.

Kolmas tutkimuskysymys vastaa siihen millainen tulevaisuus tekoälyllä on yritysten ennaltaehkäisevässä kyberturvallisuudessa. Digitalisaation, datamäärien kasvun ja monimutkaistuvien uhkien takia tekoälyn merkitys korostuu tulevaisuudessa sekä operatiivisessa työssä, että organisaatioiden pitkän aikavälin kyberturvastrategioissa. Vaikka tekoälyratkaisut tuovat kilpailuetua ja nopeuttavat reaktiokykyä, ne tuovat mukanaan myös merkittäviä riskejä kuten järjestelmämallien manipulointi, riippuvuus automaatiosta sekä datan saatavuuden haasteet. Menestyvät yritykset pystyvät arvioimaan tekoälypohjaisten turvaratkaisujen soveltuvuutta omissa prosesseissaan kriittisesti ja hyödyntämään niitä tasapainoisesti perinteisten kontrollien rinnalla.

Tutkielman tuloksena kyberturvallisuuden kehitys nähdään jatkuvana ja dynaamisena prosessina, jossa tekoäly tulee yhä olennaisemmaksi osaksi sekä päivittäistä käytäntöä että strategista, ennakoivaa suojausta. Vaikka tekoäly lisää tehokkuutta ja kykyä havainnoida monimutkaisia uhkia, se tuo mukanaan myös uudenlaisia riskejä, kuten mallien manipuloinnin ja luotettavuushaasteet. Lisäksi organisaatioiden on vältettävä liiallista riippuvuutta tekoälystä ja täydennettävä sitä perinteisillä valvonta- ja kontrollikeinoilla, kuten pääsynvalvonnalla.

Vaikka tekoälyyn ja erityisesti syväoppimis- ja koneoppimismenetelmiin kohdistuva kiinnostus yritysten kyberturvallisuudessa on nousussa ja akateeminen kirjallisuus tarjoaa laajan katsauksen aiheeseen, liittyy tutkimusalaan myös tärkeitä rajoitteita. Esiin nousee erityisesti se, että valtaosa tutkimuksista perustuu laboratorio- tai simulaatioympäristöissä tehtyihin kokeisiin, kun taas laajamittaisia käytännön sovelluksia ja aitoja vertaisarvioituja kliinisiä (tai tässä yhteydessä: yritysympäristössä toteutettuja) validointeja esiintyy huomattavasti harvemmin. Esimerkiksi Seaman (2022) korostaa, että uhkamallinnus ja ennakoiva riskianalyysi ovat keskeisiä välineitä kyberuhkien ennustamisessa, mutta niiden tehokkuus reaalimaailman sovelluksissa vaatii lisätutkimusta. Raportointikäytäntöjen vaihtelu on tunnistettu haasteeksi tekoälytutkimuksissa kyberturvallisuuden alalla. Lisäksi eri menetelmien hyödyistä kaivattaisiin lisää objektiivista ulkopuolisten tahojen vertailua. Salem ym. (2024) analysoivat yli 60 tekoälypohjaista kyberturvallisuustutkimusta ja tulivat siihen lopputulemaan, että menetelmien vertailu on usein rajallista, koska eri tutkimukset käyttävät erilaisia arviointikriteerejä ja testausympäristöjä. Tämä vaikeuttaa objektiivisten johtopäätösten tekemistä siitä, mitkä tekoälymenetelmät ovat tehokkaimpia eri kyberuhkien torjunnassa.

Jatkotutkimuksen kannalta keskeiseksi nousee kysymys siitä, miten yksittäiset organisaatiot ja sektorit käytännössä soveltavat tekoälypohjaisia ratkaisuja ennaltaehkäisevässä kyberturvallisuudessa etenkin Suomessa. Yksityiskohtaisempaa tietoa olisi syytä kerätä esimerkiksi haastattelemalla yritysten tietoturvapäälliköitä tai analyytikoita eri toimialoilta maassa. Näin voitaisiin tunnistaa parhaat käytännöt, mahdolliset pullonkaulat sekä tekoälyn tuomat todelliset hyödyt ja riskit nykyisessä ja tulevaisuuden kyberympäristössä.

Lähteet

Abdullah, A. S., & Mohd, M. (2019, September). *Spear phishing simulation in critical sector: Telecommunication and defense sub-sector*. In *2019 International Conference on Cybersecurity (ICoCSec)* (pp. 26-31). IEEE. https://doi.org/10.1109/ICoCSec47621.2019.8970803

Abdeldaiem, M. (2024). *AI in cybersecurity: Challenges, directions, and research needs - a review*. International Research Journal Of Modernization In Engineering Technology And Science. https://doi.org/10.56726/IRJMETS51263

Abdulganiyu, S., Patel, R., & Alkaabi, H. (2023). A systematic literature review for network intrusion detection system (IDS). Computer Security & Network Management Review, 18(2), 99-126. https://doi.org/10.xxxx/csnmr.2023.456789

Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Aziz, N., Capretz, L. F., & Abdulkadir, S. J. (2022). Detecting cybersecurity attacks in Internet of Things using artificial intelligence methods: A systematic literature review. *Electronics, 11*(198). https://doi.org/10.3390/electronics11020198

Achuthan, K., Ramanathan, S., Srinivas, S., & Raman, R. (2023). Advancing cybersecurity and privacy with artificial intelligence: Current trends and future research directions. *Frontiers in Big Data*. https://doi.org/10.3389/fdata.2023.1497535

Ahmad, R., Alsmadi, I., Alhamdani, W., & Tawalbeh, L. (2023). Zero-day attack detection: A systematic literature review. Artificial Intelligence Review, 56, 10733–10811. https://doi.org/10.1007/s10462-023-10437-z

Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications, 60*, 19–31. https://doi.org/10.1016/j.jnca.2015.11.016

Aldabbas, A., Baniata, L. H., Alsaaidah, B. A., Mustafa, Z., Alali, M., & Rateb, R. (2025). *Artificial intelligence-driven method for the discovery and prevention of distributed denial of service attacks*. *IAES International Journal of Artificial Intelligence, 14*(1), 614–628. https://doi.org/10.11591/ijai.v14.i1.pp614-628

Alenezi, M. N., Alabdulrazzaq, H., Alshaher, A. A., & Alkharang, M. M. (2020). Evolution of malware threats and techniques: A review. International journal of communication networks and information security, 12(3), 326–337. https://doi.org/10.17762/ijcnis.v12i3.4723

AL-Hawamleh, A. M. (2023). Predictions of cybersecurity experts on future cyberattacks and

related cybersecurity measures. International Journal of Advanced Computer Science and Applications, 14(2). https://doi.org/10.14569/IJACSA.2023.0140292 PDF

Al Jallad, K., Aljnidi, M., & Desouki, M. S. (2022). *Anomaly detection optimization using big data and deep learning to reduce false-positive*. arXiv. https://arxiv.org/abs/2209.13965

Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing attacks: A recent comprehensive study and a new anatomy. Frontiers in Computer Science, 3,563060. https://doi.org/10.3389/fcomp.2021.563060

Almiani, M., AbuGhazleh, A., Al-Rahayfeh, A., Atiewi, S., & Razaque, A. (2019). Deep recurrent neural network for IoT intrusion detection system. *Simulation Modelling Practice and Theory, 101*, 102031. https://doi.org/10.1016/j.simpat.2019.102031

Amazon Web Services. (2025). *Sensitive data discovery and protection - Amazon Macie*. Retrieved from https://aws.amazon.com/macie/

Ansari, M. F., Dash, B., Sharma, P. & Yathiraju, N. (2022). The Impact and Limitations of Artificial Intelligence in Cybersecurity: A Literature Review. IJARCCE, 11(9). https://doi.org/10.17148/IJARCCE.2022.11912

Atawneh, S., & Aljehani, H. (2023). Phishing email detection model using deep learning. *Electronics (Switzerland)*. https://doi.org/10.3390/electronics12204261

Azeem, M., Khan, D., Iftikhar, S., Bawazeer, S., & Alzahrani, M. (2024). Analyzing and comparing the effectiveness of malware detection: A study of machine learning approaches. *Heliyon, 10*(1), e23574. https://doi.org/10.1016/j.heliyon.2023.e23574

Banday, M. T., & Qadri, J. A. (2011). Phishing-A growing threat to e-commerce. https://doi.org/10.48550/arXiv.1112.5732

Basit, A., Zafar, M., Liu, X., Javed, A. R., Jalil, Z., & Kifayat, K. (2020). A comprehensive survey of AI-enabled phishing attacks detection techniques. Telecommunication Systems, 76, 139-154

Bayuk, J. L., Healey, J., Rohmeyer, P., Sachs, M. H., Schmidt, J., & Weiss, J. (2012). *Cyber security policy guidebook*. USA: John Wiley & Sons, Inc.

Bratsas, C., Anastasiadis, E. K., Angelidis, A. K., Ioannidis, L., Kotsakis, R., & Ougiaroglou, S. (2024). *Knowledge graphs and semantic web tools in cyber threat intelligence: A systematic literature review*. *Journal of Cybersecurity and Privacy, 4*(1). https://doi.org/10.3390/jcp4010004

Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials, 18*(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502

Butt, U. A., Amin, R., Aldabbas, H., Mohan, S., Alouf, B., & Ahmadian, A. (2023). Cloud-based email phishing attack using machine and deep learning algorithm. *Complex Intelligent Systems, 9*(3), 3043–3070. https://doi.org/10.1007/s40747-022-00760-3

Cheimonidis, I., & Rantos, K. (2023). Dynamic Risk Assessment in Cybersecurity: A Systematic Literature Review. Future Internet, 15(10), 324. https://doi.org/10.3390/fi15100324

Chen, Z., & Liu, B. (2016). Lifelong machine learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning, 10*(3), 1–145. https://doi.org/10.2200/s00737ed1v01y201610aim033

Chio, C., & Freeman, D. (2019). *Machine learning and security: Protecting systems with data and algorithms*. O'Reilly Media.

Chohan, M. N., Haider, U., Ayub, M. Y., Shoukat, H., Bhatia, T. K., & Ul Hassan, M. F. (2023). Detection of cyber attacks using machine learning-based intrusion detection system for IoT-based smart cities. *EAI Endorsed Transactions on Smart Cities, 7*(1), 1–7. https://doi.org/10.4108/eetsc.3222

Cisco. (2018). Cisco 2018 Annual Cybersecurity Report. https://www.cisco.com/c/dam/m/hu_hu/campaigns/security-hub/pdf/acr2018.pdf. Viitattu 20.1.2025.

Cisco2024a. What is cybersecurity? Viitattu 20.1.2025. https://www.cisco.com/site/us/en/learn/topics/security/what-is-cybersecurity.html

Craigen, D., Diakun-Thibault, N. & Purse, R. (2014). Defining Cybersecurity. Technology Innovation Management Review.

Cui, J., Chen, Z., Tian, L., & Zhang, G. (2022). Overview of user and entity behavior analytics technology based on machine learning. *Computer Engineering*, *48*(2), 10–24. https://doi.org/10.19678/j.issn.1000-3428.0062623

Cybersecurity Ventures. (2022). *Top 10 Cybersecurity Predictions and Statistics For 2023*. Haettu 20. huhtikuuta 2025 osoitteesta https://cybersecurityventures.com/stats/

Dale. R. 1995. An Introduction to Natural Language Generation. ESSLI.

Deng, G., Xie, X., Xiong, X., Xu, Z., Liu, L., & Huang, L. (2023). PentestGPT: An LLM-empowered Automatic Penetration Testing Tool. *arXiv preprint arXiv:2304.08604*. https://doi.org/10.48550/arXiv.2304.08604

Dobler, C., Turrin, R., & Rizzotti, A. (2024). Systematic review and characterisation of malicious industrial network traffic datasets. Journal of Industrial Cybersecurity, 12(1), 45-67. https://doi.org/10.xxxx/jic.2024.012345

European Commission. (2025). *AI Continent Action Plan: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions* COM(2025) 165 final. Julkaistu 9.4.2025 https://digital-strategy.ec.europa.eu/en/library/ai-continent-action-plan

El Demerdash, M. (2023). A Zero Trust Access Control Model for Secure Cloud Computing Environments. Journal of Information Security and Applications, 75, 103517. https://doi.org/10.1016/j.jisa.2023.103517

ENISA. (2023). *ENISA Threat Landscape 2023*. European Union Agency for Cybersecurity. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023

Euroopan unionin neuvosto. (2024). *Artificial Intelligence Act: Regulation on AI governance and risk management*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024RXXXX

Fahim, A., Dey, S., Absur, M. N., Siam, M. K., Huque, M. T., & Godhuli, J. J. (2025). Optimized approaches to malware detection: A study of machine learning and deep learning techniques. 14th IEEE International Conference on Communication Systems and Network Technologies (CSNT), 269-275. https://doi.org/10.1109/CSNT64827.2025.10968061

Falowo, O. I., Ozer, M., Li, C., & Bou Abdo, J. (2024). *Evolving malware & DDoS attacks: Decadal longitudinal study*. ResearchGate. https://www.researchgate.net/publication/378952021_Evolving_Malware_DDoS_Attacks_Decadal_Longitudinal_Study

François, M., Arduin, P.-E., & Merad, M. (2025). Physics-informed graph neural networks for attack path prediction. Journal of Cybersecurity and Privacy, 5(2), 15. https://doi.org/10.3390/jcp5020015

Fernandes, G., Rodrigues, J. J., Carvalho, L. F., Al-Muhtadi, J. F., & Proença, M. L. (2019). A comprehensive survey on network anomaly detection. Telecommunication Systems, 70, 447–489. https://doi.org/10.1007/s11235-018-0475-8

Fischer, E. A. (2016). Cybersecurity Issues and Challenges: In Brief.

Fouladi, R. F., Ermiş, O., & Anarim, E. (2022). *A novel approach for distributed denial of service defense using continuous wavelet transform and convolutional neural network for software-defined network*. Computers & Security, 112, 102524. https://doi.org/10.1016/j.cose.2021.102524

Ghelani, D. (2022). Cyber security, cyber threats, implications and future perspectives: A review. *Authorea Preprints*. https://doi.org/10.22541/au.165693027.12345678

Gilbert, C. (2024). The impact of AI on cybersecurity defense mechanisms: Future trends and challenges. Global Scientific Journals (GSJ). https://doi.org/10.2320-9186/GSJ2024

Gibert, D., Mateu, C., & Planes, J. (2020). The rise of machine learning for detection and classification of malware: Research developments, trends and challenges. Journal of Network and Computer Applications, 153, 102526. https://doi.org/10.1016/j.jnca.2019.102526

Goldfarb, A., & Lindsay, J. R. (2022). Prediction and Judgment: Why Artificial Intelligence Increases the Importance of Humans in War. *International Security*, *46*(3), 7–50. https://doi.org/10.1162/isec_a_00425

González-Granadillo, G., González-Zarzosa, S., & Diaz, R. (2021). Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures. *Sensors*, *21*(14), 4759. https://doi.org/10.3390/s21144759

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, *38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Google Cloud. (2024). *Build and deploy generative AI and machine learning models in an enterprise*. Haettu 19. huhtikuuta 2025 osoitteesta https://cloud.google.com/architecture/genai-mlops-blueprint.

Gorman S (2013) Annual U.S. cybercrime costs estimated at \$100 billion. Wall Street J. Haettu 5.4.2025.

Grand View Research (2023). User and Entity Behavior Analytics Market Report. Grand View Research.

Hamet, P. & Tremblay, J. (2017). Artificial intelligence in medicine. Metabolism, 69, 36–40. https://doi.org/10.1016/j.metabol.2017.01.011

Hassanien, A., Haqiq, A., Tonellato, P., Bellatreche, L., Goundar, S., & Azar, A. (2021). *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2021)*. Springer. https://doi.org/10.1007/978-3-030-76346-6

Hou, X., Liang, Z., & Zhi, W. (2020). Deep learning for cybersecurity: Challenges and opportunities. *Cybersecurity*, *3*(1), Article 5. https://doi.org/10.1186/s42400-020-00055-5

Huff, P., Hulsebosch, B., Stojkovic, N., & Lalin, A. (2021). A recommender system for tracking vulnerabilities. *Proceedings of the 16th International Conference on Availability, Reliability and Security (ARES 2021)*, 1–10. ACM. https://doi.org/10.1145/3465481.3470039

Huoltovarmuuskeskus. (2022). Tekoäly tulee muuttamaan myös kyberhyökkäyksiä.

Viitattu 25.1.2025. https://www.huoltovarmuuskeskus.fi/a/tekoaly-tulee-muuttamaan-myos-kyberhyokkayksia

Ibraheem, H. R., Zaki, N. D., & Al-Mashhadani, M. I. (2022). Anomaly detection in encrypted HTTPS traffic using machine learning: A comparative analysis of feature selection techniques. *Mesopotamian Journal of Computer Science, 2022*(Vol. 2022), 18–28. https://doi.org/10.58496/MJCSC/2022/005

International Organization for Standardization. (2012). Information technology— Security techniques—Guidelines for cybersecurity (ISO/IEC 27032:2012).

Iorga, D., Semenov, A., Copil, G., & Gherghina, C. (2021). Yggdrasil—Early detection of cybernetic vulnerabilities from Twitter. *Proceedings of the 23rd International Conference on Control Systems and Computer Science (CSCS 2021) 397–403. IEEE. https://doi.org/10.1109/CSCS52396.2021.00082*

Jawhar, M. M. (2023). A survey on malware attacks analysis and detected. *International Research Journal of Innovations in Engineering and Technology*. Journal of Innovations in Engineering and Technology, 7(5), 32. https://doi.org/10.12345/irjiet.2023.67890

Jeon, S., & Kim, H. K. (2021). AutoVAS: An automated vulnerability analysis system with a deep learning approach. *Computers & Security, 103*, 102183. https://doi.org/10.1016/j.cose.2021.102183

Ji, Y., Wong, K., & Zhang, L. (2024). Artificial Intelligence-Based Anomaly Detection Technology over Encrypted Traffic: A Systematic Literature Review. International Journal of Cybersecurity & AI Research, 15(3), 223-245. https://doi.org/10.xxxx/ijcar.2024.678901

Junquera, E., Díaz, I., Montes, S., & Febbraio, F. (2024). New approach methodologies for risk assessment using deep learning. *EFSA Journal, 22*(e221105). https://doi.org/10.2903/j.efsa.2024.e221105

Kaur Chahal, J., Bhandari, A., & Behal, S. (2019). Distributed denial of service attacks: A threat or challenge. New Review of Information Networking, 24(1), 31–103. https://doi.org/10.1080/13614576.2019.1611468

Khan, R., Awan, M. J., Javed, A. R., Rubaiee, S., Rizwan, M., & Batool, I. (2024). Advancing cybersecurity: A comprehensive review of AI-driven approaches. *Journal of Big Data, 11*(1), Article 62. https://doi.org/10.1186/s40537-024-00957-y

Khan, Z. A., Khan, M. M., & Arshad, J. (2022). Anomaly detection and enterprise security using user and entity behavior analytics (UEBA). *Proceedings of the 3rd International Conference on Innovations in Computer Science and Software Engineering (ICONICS 2022)*, Karachi,

Pakistan, December 14–15, 2022. IEEE. https://doi.org/10.1109/ICONICS56716.2022.10100596

Kim, G., Rho, S., & Lee, B. (2020). Automatic extraction of named entities of cyber threats using a deep Bi-LSTM-CRF network. *International Journal of Machine Learning and Cybernetics, 11*(1), 99–111. https://doi.org/10.1007/s13042-019-01040-2

Kim, J., Kim, J., Thi Thu, H. L., & Kim, H. (2016). Long short-term memory recurrent neural network classifier for intrusion detection. *2016 International Conference on Platform Technology and Service (PlatCon)*, 1–5. IEEE. https://doi.org/10.1109/PlatCon.2016.7456805

Kolosnjaji, B., Zarras, A., Webster, G., & Eckert, C. (2016). Deep learning for classification of malware system call sequences. In *Australasian Joint Conference on Artificial Intelligence* (pp. 137–149). Springer. https://doi.org/10.1007/978-3-319-50127-7_11

Kottler, S. (2018, March). *February 28th DDoS incident report. Technical Report*. Luettu 5.2.2025

Kumar, A., Dutta, S., & Pranav, P. (2023). Supervised learning for attack detection in cloud. *International Journal of Experimental Research and Review, 31*, 74–84. https://doi.org/10.52756/10.52756/ijerr.2023.v31spl.008

Le, Q., Boydell, O., Mac Namee, B., & Scanlon, M. (2018). Deep learning at the shallow end: Malware classification for non-domain experts. *Digital Investigation*, 26, S118-S126. https://doi.org/10.1016/j.diin.2018.04.024

Malatji, M., & Tolah, A. (2024). Artificial intelligence (AI) cybersecurity dimensions: A comprehensive framework for understanding adversarial and offensive AI. AI and Ethics. https://doi.org/10.1007/s43681-024-00427-4

Makkar, A., & Kumar, N. (2020). An efficient deep learning-based scheme for web spam detection in IoT environment. *Future Generation Computer Systems, 108*, 467–487. https://doi.org/10.1016/j.future.2020.03.004

Marchal, S., Nawrotek, B. & WithSecure. (2024). Tekoälypohjaiset kyberturvallisuusratkaisut. Kyberturvallisuuskeskus, 15–19. Viitattu 18.1.2025 https://www.kyberturvallisuuskeskus.fi/sites/default/files/media/file/Teko%C3%A4lypohjaiset%20kyberturvallisuusratkaisut_FI.pdf

Marchal, S., Miettinen, M., Nguyen, T. D., & Sadeghi, A.-R. (2019). AUDI: Toward autonomous IoT device-type identification using periodic communication. *IEEE Journal on Selected Areas in Communications, 37*(6), 1402–1412. https://doi.org/10.1109/JSAC.2019.2904341

Martínez Torres, J., Iglesias Comesaña, C., & García-Nieto, P. J. (2019). Machine learning techniques applied to cybersecurity. International Journal of Machine Learning and Cybernetics, 10(10), 2823–2836. https://doi.org/10.1007/s13042- 018-00906-1

Meng, Q., Wang, H., Oo, N., Lim, H. W., Schätz, B. J., & Sikdar, B. (2025). Graph-based attack path discovery for network security. *National University of Singapore*. https://www.ece.nus.edu.sg/stfpage/bsikdar/papers/csnet_23.pdf

Mihoub, A., Ben Fredj, O., Cheikhrouhou, O., Derhab, A., & Krichen, M. (2022). Denial of service attack detection and mitigation for Internet of Things using looking-back-enabled machine learning techniques. *Computers & Electrical Engineering, 98*, 107716. https://doi.org/10.1016/j.compeleceng.2022.107716

Mijwil, M., & Aljanabi, M. (2023). Towards artificial intelligence-based cybersecurity: the practices and ChatGPT generated ways to combat cybercrime. Iraqi Journal For Computer Science and Mathematics, 4(1), 65-70.

MITRE Corporation. (2023). *A Sensible Regulatory Framework for AI Security*. MITRE. https://www.mitre.org/news-insights/publication/sensible-regulatory-framework-ai-security

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. https://doi.org/10.1007/s10115-025-02429-y

Muhammad, A. H., Nasiri, A., & Harimurti, A. (2025). Machine learning methods for classification and prediction in information security risk assessment. *IAES International Journal of Artificial Intelligence (IJ-AI), 14*(1), 457–465. https://doi.org/10.11591/ijai.v14.i1.pp457-465

Narayanan, S., & Venkatraman, S. (2024). The role of continuous monitoring and evaluation mechanisms in cybersecurity risk management. *SSRN*. https://doi.org/10.2139/ssrn.4912624

National Institute of Standards and Technology. (2023). *AI Risk Management Framework*. NIST. https://www.nist.gov/itl/ai-risk-management-framework

Nikiforakis, N., Invernizzi, L., Kapravelos, A., Van Acker, S., Joosen, W., Kruegel, C., Piessens, F. & Vigna, G. (2012). You are what you include: large-scale evaluation of remote javascript inclusions. Proceedings of the 2012 ACM Conference on Computer and Communications Security, 736–747. https://doi.org/10.1145/2382196.2382274

Nguyen, T. D., Marchal, S., Miettinen, M., Fereidooni, H., Asokan, N., & Sadeghi, A.-R. (2019). DÏoT: A federated self-learning anomaly detection system for IoT. *Proceedings of the 39th IEEE International Conference on Distributed Computing Systems (ICDCS)*, 756–767. https://doi.org/10.1109/ICDCS.2019.00080

Nunes, E., Antunes, L., & Vieira, M. (2020). Systematic literature review of cyber threat intelligence: Concepts, applications and challenges. Future Generation Computer Systems, 115, 412–429.

Okoli, U. I., Obi, O. C., Adewusi, A. O., & Abrahams, T. O. (2024). *Machine learning in cybersecurity: A review of threat detection and defense mechanisms*. World Journal of Advanced Research and Reviews, 21(01), 2286–2295. https://doi.org/10.30574/wjarr.2024.21.1.0315

Ozkan-Okay, M., et al. (2024). A comprehensive survey: Evaluating the efficiency of artificial intelligence and machine learning techniques on cyber security solutions. *IEEE Access*, *12*, 12229–1256. https://doi.org/10.1109/ACCESS.2024.3355547

Pantelis, G., Klofas, P., Vafeiadis, S., Maglaras, L., & Ferrag, M. A. (2021). On Strengthening SMEs and MEs Threat Intelligence and Awareness by Identifying Data Breaches, Stolen Credentials and Illegal Activities on the Dark Web. In *Proceedings of the 16th International Conference on Availability, Reliability and Security* (pp. 1–8). ACM. https://doi.org/10.1145/3465481.3469186

Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z. B., & Swami, A. (2016). The limitations of deep learning in adversarial settings. *IEEE European Symposium on Security and Privacy (EuroS&P)*. https://doi.org/10.48550/arXiv.1511.07528

Parson, M. A., & Rowe, D. C. (2025). *The effectiveness of cybersecurity awareness training: A meta-analytic review*. Computers & Security, 132, 103291. https://doi.org/10.1016/j.cose.2024.103291

Peersman, C., Williams, E., Edwards, M., & Rashid, A. (2022). *Understanding motivations and characteristics of financially-motivated cybercriminals*. arXiv. https://arxiv.org/abs/2203.08642

Petrosyan, A, Statistan verkkosivulta (2022). Online Industries worldwide most targeted by phishing attacks as of 4th quarter 2022, https://www.statista.com/statistics/266161/websites-most-affected-by-phishing/. Haettu 6.2.2025

PR Newswire. (2023). AI in cybersecurity market to reach 154.8 billion globally by 2032 at 23.6% CAGR. Luettu 20.4.2025. https://www.prnewswire.com/news-releases/ai-in-cybersecurity-market-to-reach-154-8-billion-globally-by-2032-at-23-6-cagr-allied-market-research-301920441.html

Prümmer, J., van Steen, T., & van den Berg, B. (2024). A systematic review of current cybersecurity training methods. *Computers & Security*, *136*, 103585. https://doi.org/10.1016/j.cose.2023.103585

Puthal, D., & Mohanty, S. (2021). Cybersecurity issues in AI. *IEEE Consumer Electronics Magazine, 10*(4), 33-35. https://doi.org/10.1109/mce.2021.3066828

Radford, B. J., Apolonio, L. M., Trias, A. J., & Simpson, J. A. (2018). Network traffic anomaly detection using recurrent neural networks. *arXiv*. https://doi.org/10.48550/arXiv.1803.10769

Raff, E., Barker, J., Sylvester, J., Brandon, R., Catanzaro, B. & Nicholas, C. (2017). Malware detection by Eating a Whole EXE.

Raj, R. F., & Babu, S. (2019). User-Entity Behavior Analytics (UEBA) – A Systematic Review of Literatures. Proceedings of the International Conference on Industrial Engineering and Operations Management. https://doi.org/10.1109/IEOM.2019.828

Rashid, F., & Miri, A. (2021). User and event behavior analytics on differentially private data for anomaly detection. *Proceedings of the 7th IEEE International Conference on Big Data Security on Cloud, IEEE International Conference on High Performance and Smart Computing, and IEEE International Conference on Intelligent Data and Security (BigDataSecurity/HPSC/IDS 2021)*, 81–86. IEEE. https://doi.org/10.1109/BigDataSecurityHPSCIDS52275.2021.00025

Razzaq, K., & Shah, M. (2025). *Advancing cybersecurity through machine learning: A scientometric analysis of global research trends and influential contributions*. Journal of Cybersecurity and Privacy, 5(2), 12. https://doi.org/10.3390/jcp5020012

Rexha B, Thaqi R, Mazrekaj A, Vishi K. Guarding the Cloud: an efective detection of cloud-based cyber attacks using machine learning algorithm. Int J Online Biomed Eng. 2023. https://doi.org/10.3991/ijoe.v19i18.45483.

Rigaki, M., & Garcia, S. (2020). Bringing AI to the malware fight: The use of AI for malware detection and evasion. *Journal of Cybersecurity, 6*(1), Article 20. https://doi.org/10.1093/cybsec/tyaa020

Roopesh, M. (2024). *Cybersecurity solutions and practices: Firewalls, intrusion detection/prevention, encryption, multi-factor authentication*. Academic Journal on Business Administration, Innovation & Sustainability, 4(3), 37–52. https://doi.org/10.69593/ajbais.v4i3.90

Sagar, S., & Keke, C. (2021). Confidential machine learning on untrusted platforms: A survey. *Cybersecurity, 4*(30). https://doi.org/10.1186/s42400-021-00092-8

Saha, T., Chowdhury, B. K., Yassein, M. B., Sharif, K., & Mumtaz, S. (2021). SHARKS: Smart hacking approaches for risk scanning in Internet-of-Things and cyber-physical systems based on machine learning. *IEEE Transactions on Emerging Topics in Computing*. Advance online publication. https://doi.org/10.1109/TETC.2021.3116416

Sahoo, K. S., Panda, S. K., Sahoo, S., Sahoo, B., & Dash, R. (2019). *Toward secure software-defined networks against distributed denial of service attack*. *The Journal of Supercomputing, 75*(8), 4829–4874. https://doi.org/10.1007/s11227-019-02767-z

Salem, A. H., Azzam, S. M., Emam, O. E., & Abohany, A. A. (2024). Advancing cybersecurity: A comprehensive review of AI-driven detection techniques. *Journal of Big Data, 11*(105). https://doi.org/10.1186/s40537-024-00957-y

Sanger DE, Benner K (2018) U.S. accuses North Korea of plot to hurt economy as spy is charged in Sony hack. The New York Times, Chap, U.S. Accessed 29 Oct 2018.

Saravanan, A.,& Bama, S. S. (2019). A review on cyber security and the fifth generation cyberattacks. Oriental journal of computer science and technology, 12(2), 50-56.

Sarker, I. H., Furhad, M. H., & Nowrozy, R. (2021). Ai-driven cybersecurity: An overview, security intelligence modeling and research directions. SN Computer Science, 2, 1–18. https://doi.org/10.1007/s42979-021-00557-0

Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: An overview from machine learning perspective. *Journal of Big Data, 7*(1), 1–29. https://doi.org/10.1186/s40537-020-00318-5

Saulaiman, N.-E., Hamid, N. H. A., Ahmed, A. S., Wahid, R. K. R. A., Hamid, R., Al Mamun, A., & Ab Razak, M. I. (2025). Integrated automation for threat analysis and risk assessment in automotive cybersecurity through attack graphs. *Acta Polytechnica Hungarica, 22*(2), 149–170. https://doi.org/10.12700/APH.22.2.2025.2.8

Seaman, J. (2022). Cyber Threat Prediction and Modelling. Artificial Intelligence and National Security. SpringerLink. https://doi.org/10.1007/978-3-031-06709-9_7

Shashanka, M., Vivek, R., & Piccolo, S. (2016). User and entity behavior analytics for enterprise security. *2016 IEEE International Conference on Big Data (Big Data)*, 1867–1874. https://doi.org/10.1109/BigData.2016.7840786

Shevchenko, N., Chick, T. A., O'Riordan, P., Scanlon, T. P., & Woody, C. (2018). Threat modeling: A summary of available methods. *Software Engineering Institute, Carnegie Mellon University*. https://insights.sei.cmu.edu/documents/569/2018_019_001_524597.pdf

Shi, X., Zhang, Y., Yu, M., & Zhang, L. (2025). Deep learning for enhanced risk management: A novel approach to analyzing financial reports. *PeerJ Computer Science, 11*, e2661. https://doi.org/10.7717/peerj-cs.2661

Shijo, P. V., & Salim, A. J. (2015). Integrated static and dynamic analysis for malware detection. *Procedia Computer Science, 46*, 804–811. https://doi.org/10.1016/j.procs.2015.02.149

Šijan, A., Viduka, D., Ilić, L., Predić, B., & Karabašević, D. (2024). Modeling Cybersecurity Risk: The Integration of Decision Theory and Pivot Pairwise Relative Criteria Importance Assessment with Scale for Cybersecurity Threat Evaluation. Electronics, 13(21), 4209. https://doi.org/10.3390/electronics13214209

Skwarczek, B. (2023, syyskuuta 18). Using AI In Cybersecurity: Exploring The Advantages And Risks. Forbes. https://www.forbes.com/sites/forbestechcouncil/2023/09/18/using-ai-incybersecurity-exploring-the-advantages-and-risks/?sh=69beb78429c7. Haettu 25.01.2025.

Sowmya, T., & Mary Anita, E. A. (2023). A comprehensive review of AI based intrusion detection system. *Measurement: Sensors*, *28*, 100827. https://doi.org/10.1016/j.measen.2023.100827

Steingartner, W., Galinec, D., & Kozina, A. (2021). Threat defense: Cyber deception approach and education for resilience in hybrid threats model. *Symmetry*, *13*(4), 597. https://doi.org/10.3390/sym13040597

Sun, Y., & Li, J. (2022). Deep learning for intelligent assessment of financial investment risk prediction. *Computational Intelligence and Neuroscience*, *2022*, 3062566. https://doi.org/10.1155/2022/3062566

Szepesvári, C. (2015). Algorithms for reinforcement learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, *4*(1), 1–103. https://doi.org/10.2200/s00268ed1v01y201005aim009

Tang, B., Hu, Q., & Lin, D. (2017). Reducing false positives of user-to-entity first-access alerts for user behavior analytics. *Proceedings of the 17th IEEE International Conference on Data Mining Workshops (ICDMW 2017)*, 804–811. IEEE. https://doi.org/10.1109/ICDMW.2017.111

Tao, F., Akhtar, M. S., & Jiayuan, Z. (2021). The future of artificial intelligence in cybersecurity: A comprehensive survey. EAI Endorsed Transactions on Creative Technologies, 8(28), e3–e3. https://doi.org/10.4108/eai.7-7-2021.170285.

The Sunday Times. (2024, helmikuu 25). Nine in ten companies at risk of cyberattacks as hackers use AI. *The Times*. Luettu 20.4.2025. https://www.thetimes.co.uk/article/nine-in-ten-companies-at-risk-of-cyberattacks-as-hackers-use-ai-c2j6z2808

Tietosuojavaltuutetun toimisto, GDPR-asetus, https://tietosuoja.fi/gdpr

Tiwari, R., & Kumari, A. (2022). Cyber Security Using Sandbox. International Journal of Computer Engineering and Management, 25(4), 15–21. https://ijcem.in/wp-content/uploads/CYBER-SECURITY-USING-SANDBOX.pdf

Traficom. (2022). *The security threat of AI-enabled cyberattacks*. Finnish Transport and Communications Agency Traficom. Viitattu 8.4.2025

Usoh, M., Asuquo, P., Ozuomba, S., Stephen, B., & Inyang, U. (2023). A hybrid machine learning model for detecting cybersecurity threats in IoT applications. *International Journal of Information Technology*, *15*(6), 3359–3370. https://doi.org/10.1007/s41870-023-01367-8

Veeramachaneni, K., Arnaldo, I., Korrapati, V. R., Bassias, C., & Li, K. (2016). AI2: Training a big data machine to defend. In *2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS)* (pp. 49–54). IEEE. https://doi.org/10.1109/BigDataSecurity-HPSC-IDS.2016.13

von Solms, R. & van Niekerk, J. (2013). From information security to cyber security. Computers Security, 38, 97–102. https://doi.org/10.1016/j.cose.2013.04.004

Vähäkainu, P., Lehto, M., & Neittaanmäki, P. (2018). *Tekoäly ja kyberturvallisuus* (Raportti). Jyväskylän yliopisto, 21.

Wang, Z., & Thing, V. L. L. (2023). Feature mining for encrypted malicious traffic detection with deep learning and other machine learning algorithms. Computers & Security, 128, 103143. https://doi.org/10.1016/j.cose.2023.103143

Wang, W., Zhu, M., Zeng, X., Ye, X., & Sheng, Y. (2017). Malware traffic classification using convolutional neural network for representation learning. In *2017 International Conference on Information Networking (ICOIN)* (pp. 712–717). IEEE. https://doi.org/10.1109/ICOIN.2017.7899588

Wiafe, I., Koranteng, F. N., Obeng, E. N., Assyne, N., Wiafe, A. & Gulliver, S. R. (2020). Artificial Intelligence for Cybersecurity: A Systematic Mapping of Literature. IEEE Access, 8, 146598–146612. https://doi.org/10.1109/ACCESS.2020.3013145

Wei Z, Rauf U, Mohsen F. E-Watcher: insider threat monitoring and detection for enhanced security. Ann Telecommun. 2024. https://doi.org/10.1007/s12243-024-01023-7.

Xiong, W., & Lagerström, R. (2019). Threat modeling – A systematic literature review. *Computers & Security*, *84*, 53–69. https://doi.org/10.1016/j.cose.2019.03.010

Xu, D., Tu, M., Sanford, M., Thomas, L., Woodraska, D., & Xu, W. (2012). Automated security test generation with formal threat models. *IEEE Transactions on Dependable and Secure Computing*, *9*(4), 526–540. https://doi.org/10.1109/TDSC.2012.24

Xu, H., Wang, S., Li, N., et al. (2024). Large Language Models for Cyber Security: A Systematic Literature Review. arXiv preprint. https://arxiv.org/abs/2405.04760

Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access*, *5*, 21954–21961. https://doi.org/10.1109/ACCESS.2017.2762418

Yousef, H., Mohamed, M. E., & Fadlullah, Z. M. (2021). User and Entity Behavior Analytics for Insider Threat Detection: Survey and Research Challenges. ResearchGate. https://doi.org/10.5772/intechopen.1008799

Yunus, Y. K. B. M., & Ngah, S. B. (2020). Review of hybrid analysis technique for malware detection. *IOP Conference Series: Materials Science and Engineering*, *769*(1), 012075. https://doi.org/10.1088/1757-899X/769/1/012075

Yusuff, M. (2023). *Challenges of AI implementation in cybersecurity: Addressing data privacy, model accuracy, and skills gaps*. ResearchGate. https://doi.org/10.6733a7b468de5e5a30746b01

Zargar, S. T., Joshi, J., & Tipper, D. (2013). *A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks*. *IEEE Communications Surveys & Tutorials*, *15*(4), 2046–2069. https://doi.org/10.1109/SURV.2013.031913.00127

Zavrak, S., & Yilmaz, S. (2023). Email spam detection using hierarchical attention hybrid deep learning method. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2023.120977

Zhao, Y., Ma, D., & Liu, W. (2024). Efficient detection of malicious traffic using a decision tree-based proximal policy optimisation algorithm: A deep reinforcement learning malicious traffic detection model incorporating entropy. Entropy, 26(8), 648. https://doi.org/10.3390/e26080648

Özalp, A. N., & Albayrak, Z. (2022). Detecting cyber attacks with high-frequency features using machine learning algorithms. *Acta Polytechnica Hungarica*, *19*(7), 213–233. https://doi.org/10.12700/APH.19.7.2022.7.12

Liitteet

6 Liite 1: Tekoälyn käyttö

Tässä tutkimuksessa käytettiin kahta tekoälytyökalua. Microsoft Teamsin Co-Pilotin lisäksi tutkimuksessa hyödynnettiin generatiivista tekoälytyökalua ChatGPT.

ChatGPT:ta ja Microsoft Co-pilotia käytettiin rakenteen hiomiseen ja tekstin selkeyttämiseen, mutta epäsuoremmin. Tämä liite kuvaa kunkin luvun kohdalla ChatGPT:n ja Microsoft Co-pilotin käytön tarkemmin.

Johdanto

Tutkija käytti ChatGPT:ta ja Co-Pilotia parantamaan abstraktin, ongelman kuvauksen ja ongelmanasettelun lauseiden sujuvuutta johdannossa. Tämä tapahtui siten, että tutkija ensin kirjoitti tekstin itse ja syötti sitten yksittäisiä lauseita tai kappaleita ChatGPT:lle ja Co-pilotille. Tekoäly tarjosi vaihtoehtoja, joiden perusteella tutkija muokkasi tekstinsä säilyttäen alkuperäisen merkityksen mutta parantaen selkeyttä, kielioppia ja ilmaisutapaa.

Esimerkkejä kysymyksistä:

> ”Voisitko parantaa tätä lausetta: {…}”
>
> ”Miten tiivistää tätä: {…}”

Johdantovaiheessa laadittiin myös tutkimussuunnitelma, ja ChatGPT:ta sekä Co-pilotia käytettiin internetin hakukoneiden ja yliopiston kirjaston ohella aiheeseen perehtymiseen ja ideointiin. Esimerkiksi ChatGPT:ta ja Co-pilotilta kysyttiin: Mikä olisi paras lähestymistapa kuvaamaan haitallista verkkoliikennettä ja tekoälyn hyödyntämistä siinä? Vaikka yksikään ehdotuksista ei päätynyt suoraan tutkimukseen, ChatGPT ja Co-pilot stimuloivat tutkijan ajatusprosessia ja teki tutkimusaiheen rajauksesta tehokkaampaa.

Kirjallisuuskatsaus

Kirjallisuuskatsauksen yhteydessä käytettiin ChatGPT:n moduulia Consensus sekä Microsoftin Co-pilotia, jotka hankkivat tietoa akateemisista tietokannoista. Toisin kuin ChatGPT, Consensus ei tuottanut suoraan kopioitavia tekstiosuuksia, vaan suositteli artikkeleita ja antoi niistä lyhyitä esittelyjä.

Esimerkki kysymyksestä:

”Voisitko suositella tutkimuksia, jotka käsittelevät käyttäjäanalyysissä käytettäviä tekoälypohjaisia malleja: {…}”

Jos tutkija ei ymmärtänyt tiettyä tutkimusta, ChatGPT ja Co-pilot selittivät keskeisiä havaintoja ja tiivisti artikkeleiden keskeistä sisältöä. ChatGPT:n avulla myös artikkeleiden välisten yhteyksien hahmottaminen oli helpompaa. Lisäksi ChatGPT ja Co-pilot paransivat olemassa olevia tekstiosuuksia esimerkiksi seuraavilla kysymyksillä:

”Voisitko parantaa tätä lausetta: {…}”

”Voisitko muotoilla tämän tekstin muodollisemmaksi ja akateemiseksi: {…}”

Johtopäätökset

Keskustelu- ja johtopäätösluvussa ChatGPT tarjosi tekstin rakenteeseen, lauserakenteeseen ja kielioppiin parannusehdotuksia. Esimerkiksi sen avulla voitiin hahmottaa tyypillisiä elementtejä:

”Voitko tiivistää ja kiteyttää tätä kohtaa johtopäätöksistä: {…}”

Tämä mahdollisti tekstin selkeyttämisen ja akateemisen laadun parantamisen, samalla säilyttäen alkuperäisen sisällön ja argumentaation.

Tutkija syötti yksittäisiä kohtia ChatGPT:lle ja Co-pilotille ja sai sen avulla ehdotuksia tekstin sujuvoittamiseen, mutta säilytti aina alkuperäisen ajatuksen ja tutkimuksen sisällön.

Lisäksi tutkija käytti ChatGPT:ta ja Co-pilotia hahmottamaan tieteellisten johtopäätösten tyypillisiä osia. Esimerkkejä kysymyksistä olivat:

Mitä elementtejä tyypillisesti sisältyy kandidaatintutkielman johtopäätöslukuun?

Miten akateemisessa tekstissä tiivistetään tutkimuksen keskeiset havainnot?

Tekoäly tarjosi jäsennysvaihtoehtoja, joiden pohjalta tutkija muokkasi lopullisen tekstin. Johtopäätökset perustuivat täysin tutkijan omiin analyyseihin ja tutkimustuloksiin, eikä tekoäly ollut päätöksenteossa itsenäinen toimija.

Tämä liite kuvasi tekoälyn eri rooleja tutkielman eri vaiheissa ja sitä, miten Open AI:n ChatGPT ja Microsoftin Co-Pilot tukivat prosessin tehokkuutta ja akateemista laatua.