TURUN YLIOPISTO
Kauppakorkeakoulu

# Tekoälyn hyödyntäminen yritysten kyberturvallisuudessa

Tietojärjestelmätieteen kandidaatintutkielma

Laatija:
Emmiina Arte

Ohjaaja:
FT Kai Kimppa

8.5.2024
Turku

Kanditutkielma

**Oppiaine**: Tietojärjestelmätiede
**Tekijä**: Emmiina Arte
**Otsikko**: Tekoälyn hyödyntäminen yritysten kyberturvallisuudessa
**Ohjaaja**: FT Kai Kimppa
**Sivumäärä**: 32 sivua
**Päivämäärä**: 8.5.2024

Kyberrikollisuus ja sen aiheuttamat taloudelliset vahingot ovat jatkuvassa kasvussa. Tekoälyn hyödyntäminen kyberrikollisuudessa on lisääntynyt. Tämä on pakottanut myös yritykset ottamaan käyttöön tekoälyratkaisuja järjestelmiensä suojaamiseksi.

Tässä kandidaatintutkielmassa selvitetään, miten tekoälyä voidaan hyödyntää parantamaan yritysten kyberturvallisuutta. Tutkielma toteutetaan kirjallisuuskatsauksena, jossa käytetään aineistona tekoälyä ja kyberturvallisuutta käsitteleviä tieteellisiä artikkeleita. Kirjallisuuskatsauksessa perehdytään yritysten kohtaamiin kyberuhkiin ja kyberhyökkäysten muotoihin ja koneoppimisen sekä syväoppimisen tekoälytekniikoihin. Keskeisinä tarkastelun kohteina ovat yrityksiä uhkaavat kyberhyökkäykset, koneoppimisen ja syväoppimisen teknologiat sekä näiden teknologioiden soveltamismahdollisuudet kyberturvallisuudessa. Tutkielmassa käsitellään tekoälyn hyötyjä ja toiminnallisuutta kyberturvallisuuden toiminnoissa. Tutkielmassa pohditaan hyötyjen lisäksi myös tekoälyn tuomia haasteita, uhkakuvia ja tulevaisuuden näkymiä.

Tekoälytekniikat voivat perinteisten teknologioiden ohella parantaa kyberturvallisuutta merkittävästi. Koneoppimisen ja syväoppimisen menetelmät voivat turvata ja suojella yrityksen tärkeää ja arkaluontoista dataa ja ohjelmistoja. Menetelmät voivat myös nopeuttaa reagointia kyberhyökkäyksiin. Tekoäly voi datan ja ohjelmistojen turvaamisen ohella tehdä kyberturvallisuuden toimintojen suorittamisesta nopeampaa ja varmempaa.

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

# SISÄLLYS

## KUVIOT



## TAULUKOT

# 1 Johdanto

Maailmanlaajuisen kyberrikollisuuden aiheuttamien taloudellisten vahinkojen ennustetaan kasvavan 15 prosentilla vuodessa, nousten 10,5 biljoonaan dollariin vuoteen 2025 mennessä (Morgan, 2023). Vaikka kyberturvallisuuden tilastot ovat hälyttäviä, on organisaatioilla uusi liittolainen taistelussa kyberrikollisuutta vastaan: tekoäly (Skwarczek, 2023).

Kyberhyökkäyksissä pyritään murtautumaan organisaation tietojärjestelmään. Yleisimpiä kyberhyökkäysten tyyppejä ovat muun muassa haittaohjelmat, kiritysohjelmat, palvelunestohyökkäykset, tietojenkalastelu ja sosiaalinen manipulointi. Kyberturvallisuuden ongelmien ratkaisemiseen voidaan hyödyntää tekoälytekniikoita, kuten koneoppimisen ja syväoppimisen menetelmiä. (Sarker ym., 2021.)

Tekoälyn käyttö kyberrikollisuudessa on kasvanut merkittävästi. Tämän seurauksena useat yritykset ovat alkaneet niin ikään hyödyntämään tekoälyä omien järjestelmiensä suojauksessa kyberhyökkäyksiä vastaan (Tao ym., 2021). Tässä tutkielmassa perehdytään tekoälyn hyödyntämiseen yritysten kyberturvallisuudessa.

Tutkielman tutkimuskysymykset ovat seuraavat:

1. Miten yritykset voivat hyödyntää tekoälyä kyberturvallisuudessa?
2. Millaisia kyberuhkia yritykset kohtaavat?
3. Millaisia tekoälytekniikoita yritykset käyttävät kyberuhkien torjuntaan?

Tutkielmassa käsitellään aihetta yrityksen näkökulmasta. Kyberhyökkäyksiä on paljon erilaisia, jonka vuoksi tässä tutkielmassa kyberhyökkäykset rajataan palvelunestohyökkäyksiin ja haittaohjelmiin.

Tutkielman rakenne on seuraava. Tutkielman toisessa luvussa käsitellään toista tutkimuskysymystä. Siinä käsitellään yritysten kohtaamia kyberuhkia, mitä kyberturvallisuus ja kyberhyökkäykset ovat ja miten kyberhyökkäyksiin voidaan reagoida. Tutkielman kolmannessa luvussa käsitellään kolmatta tutkimuskysymystä. Sen aiheena on, mitä koneoppiminen ja syväoppiminen ovat. Luvussa käsitellään miten ja miksi näitä tekoälytekniikoita hyödynnetään kyberturvallisuudessa. Kolmannessa luvussa käsitellään myös tekoälyn hyödyntämisen uhkia. Luvussa neljä on yhteenveto,

jossa vastataan tutkielman ensimmäiseen tutkimuskysymykseen eli päätutkimuskysymykseen, miten yritykset voivat hyödyntää tekoälyä kyberturvallisuudessa, ja työstä tehdyt johtopäätökset.

# 2 Yritysten kohtaamat kyberuhat

## 2.1 Mitä kyberturvallisuus ja kyberhyökkäykset ovat

Kyberturvallisuus on tietokoneiden, palvelimien, mobiililaitteiden, elektronisten järjestelmien, verkkojen ja datan suojaamista haitallisilta hyökkäyksiltä. Kyberturvallisuuden termi on laaja, ja se kattaa kaiken tietoturvasta hyökkäyksistä toipumiseen ja loppukäyttäjien koulutukseen. (Martínez Torres ym., 2019.) International Organization for Standardization (ISO) on määritellyt kyberturvallisuuden luottamuksellisuuden, eheyden ja tiedon saatavuuden säilyttämiseksi kyberavaruudessa. ISO:n määritelmän mukaan kyberturvallisuuteen liittyy läheisesti viisi käsitettä, jotka ovat tietoturva, verkkoturva, internet-turva, kriittisen tietoinfrastruktuurin suojaus ja kyberrikollisuus. Tietoturva huolehtii tietojen luottamuksellisuuden, eheyden ja saatavuuden suojelusta. Verkkoturva huolehtii verkkojen suunnittelusta, toteutuksesta ja toiminnasta saavuttaakseen tietoturvan verkkojen sisällä. Internet-turva huolehtii internetin palveluiden ja järjestelmien suojelusta ja internet-palveluiden saatavuudesta. Kriittisen tietoinfrastruktuurin suojaus varmistaa, että kriittiset järjestelmät ovat suojattuja tieto-, verkko-, kyber- ja internetturvariskeiltä. Kyberrikollisuus on rikollista toimintaa, jossa järjestelmiä hyödynnetään rikoksen tekemiseen tai käytetään rikoksen kohteena. (International Organization for Standardization, 2012.)

Kyberhyökkäykset ovat lisääntyneet ja ovat yhä yleisempiä uhkia yrityksille. Kyberhyökkäys on tahallinen ja ilkivaltainen pyrkimys vahingoittaa tietojärjestelmää (AL-Hawamleh, 2023.) Hyökkäysten motiivina voi olla esimerkiksi taloudellinen hyöty tai poliittinen toiminta (Mijwil ym., 2023). Myös datan- ja tietojenkalastelu tai niiden tuhoaminen voi olla hyökkäyksen motiivina (AL-Hawamleh, 2023). Kyberhyökkäyksiä toteutetaan hyödyntämällä sähköisten järjestelmien ja verkkojen aukkoja (Mijwil ym., 2023).

Erilaisista kyberhyökkäyksistä ja niiden toteuttamistavoista ymmärretään vain pieni osa. Samalla hyökkäyksiltä suojautumiseen ja turvatoimien kehittämiseen vaaditaan kattavaa ymmärrystä kyberhyökkäyksistä. (AL-Hawamleh, 2023.) Jotta kyberhyökkäyksiltä voidaan suojautua, on kyberhyökkäysten luokittelun ja ymmärtämisen oltava tietoturvatoimien keskiössä. Monet yritykset ja organisaatiot ovat joutuneet kyberhyökkäysten uhriksi. Ciscon entisen toimitusjohtajan mukaan on olemassa

kahdenlaisia yrityksiä. Ensimmäinen luokka sisältää ne yritykset, jotka ovat joutuneet hyökkäyksen kohteeksi ja toinen luokka sisältää ne, jotka eivät vielä tiedä joutuneensa hyökkäyksen kohteeksi (Cisco, 2018).

## 2.2 Erilaiset kyberhyökkäykset

### 2.2.1 Haittaohjelmahyökkäys

Haittaohjelmahyökkäys (engl. malware) on yleisnimitys erilaisille tunkeileville ohjelmistoille. Kyberrikolliset kehittävät haittaohjelmia tietojen varastamiseen, pääsykontrollien ohittamiseen ja järjestelmien vahingoittamiseen. Haittaohjelmahyökkäykset ovat yleistyneet ja ovat houkuttelevia kyberrikollisille niiden potentiaalisen taloudellisen tuottavuuden vuoksi. Haittaohjelmat keräävät miljardien dollarien voittoja. Haittaohjelmahyökkäysten määrä, tyypit ja kompleksisuus ovat moninkertaistuneet ja ne ovatkin muodostuneet äärimmäisen ongelmallisiksi ohjelmistoiksi torjua. (Alenezi ym., 2020.)

Haittaohjelmat voidaan jakaa kahteen kategoriaan: ensimmäisen sukupolven haittaohjelmiin eli staattisiin haittaohjelmiin ja toisen sukupolven haittaohjelmiin eli dynaamisiin haittaohjelmiin. Ensimmäisen ja toisen sukupolven haittaohjelmat luokitellaan infektiostrategian perusteella. Ensimmäisen sukupolven haittaohjelmien käyttäytyminen pysyy muuttumattomana kohdejärjestelmän tartuttamisen jälkeen. Toisen sukupolven haittaohjelmat muodostavat uuden variantin jokaisen tartuttamisen jälkeen. (Alenezi ym., 2020.)

Haittaohjelmat voidaan jakaa erilaisiin, toisiaan täydentäviin kategorioihin riippuen niiden tarkoituksesta ja leviämisjärjestelmästä. Takaportti on tietokonesovellus, joka on suunniteltu kiertämään järjestelmän turvamekanismeja ja asentumaan tietokoneeseen, mahdollistaen hyökkääjän pääsyn siihen. Botti on ohjelmisto, joka suorittaa automaattisesti tiettyjä toimintoja, kuten hajautettuja palvelunestohyökkäyksiä tai muun haittaohjelman levittämistä, osana bottiverkkoa. Lunnasohjelma eli kiristyshaittaohjelma on ohjelma, joka rajoittaa käyttäjän pääsyä tietokonejärjestelmään salakirjoittamalla tiedostoja tai lukitsemalla järjestelmän ja vaatimalla lunnaita sen vapauttamiseksi. Troijalainen on ohjelma, joka on harmittomaksi naamioitu haitallinen ohjelma, joka huijaa käyttäjää asentamaan ohjelman järjestelmään. Virus on itsestään

leviävä haittaohjelma, joka leviää laitteesta toiseen. Mato on viruksen tyyppi, joka hyödyntää käyttöjärjestelmän haavoittuvuuksia levitäkseen. (Gibert ym., 2020.)

### 2.2.2 Kiristyshaittaohjelmahyökkäys

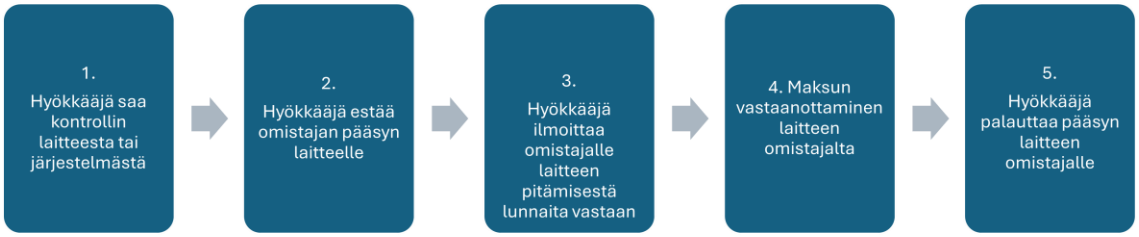
Kiristyshaittaohjelmahyökkäys (engl. ransomware) on yksi kehittyneimmistä kyberhyökkäyksistä. Kiristyshaittaohjelmahyökkäyksessä hyökkääjä tekee sarjan toimia, jossa salataan hyökkäyksen uhrin tiedostoja tai koko tietokonejärjestelmän sisältö. Tämän jälkeen hyökkääjä vaatii maksua vastineeksi siitä, että hän antaa uhrille salauksenpurkuavaimen. Kiristyshaittaohjelmahyökkäys voidaan toteuttaa useiden eri kanavien kautta, kuten tietojenkalastelun kautta. Uhrit joutuvat maksamaan hyökkääjälle saadakseen hyökkäyksen kohteena olleet tiedot takaisin, koska kiristyshaittaohjelmahyökkäys voi aiheuttaa uhrille esimerkiksi huomattavaa taloudellista haittaa. On huomattava, että maksaminen ei kuitenkaan takaa tietojen takaisin saantia. (Mijwil ym., 2023.)

Kiristyshaittaohjelmahyökkäyksen eteneminen voidaan jakaa viiteen eri vaiheeseen (ks. kuva 1). Kiristyshaittaohjelmahyökkäyksen ensimmäinen vaihe on haittaohjelman asentaminen uhrin laitteelle. Ohjelman asentaminen tapahtuu usein ilman, että uhri huomaa tai ymmärtää sen tapahtuneen. Tämä voidaan toteuttaa esimerkiksi ponnahdusikkunan avulla, joka huijaa uhria painamaan ikkunasta. Ponnahdusikkunassa voidaan esimerkiksi kertoa uhrille tietokoneen saneen tartunnan ja painamaan jotakin nappia torjuakseen tartunnan. Haittaohjelman asentamiseen voidaan myös hyödyntää luotettavan näköisiä sähköpostiviestejä, joissa on haitallisia liitteitä. Uhrin valinta voi perustua siihen, kuinka helposti uhria uskotaan voitavan huijata. (Tandon & Nayyar, 2019.)

Toisessa vaiheessa haittaohjelman asentamisen jälkeen muodostetaan yhteys hyökkääjän hallitsemaan keskuspalvelimeen. Tämä tapahtuu taustalla ilman, että uhri huomaa mitään. Haittaohjelma lähettää tiedot uhrin laitteesta hyökkääjän palvelimelle. Tämä prosessi kertoo hyökkääjälle, kuinka arvokas uhri on ja kuinka paljon lunnaita voidaan vaatia. Yhteyden muodostamisen jälkeen haittaohjelma ja hyökkääjän palvelin vaihtavat salausavaimia. Salausavaimia käytetään lukitsemaan uhrin tiedostot, jotta uhri ei pääse niihin käsiksi ilman salausavainta. (Tandon & Nayyar, 2019.)

Kolmannessa vaiheessa haittaohjelma näyttää viestejä uhrin laitteella kertoen uhrille, että tämän tiedostot ovat lukittuja ja niiden lukituksen poistamista vastaan vaaditaan maksua. Jotkin haittaohjelmat poistavat osan tiedostoista uhrin koneelta painostaakseen uhria maksamaan vaaditun summan. (Tandon & Nayyar, 2019.)

Neljännessä vaiheessa laitteen omistajalta vastaanotetaan maksu. Vaikka uhri maksaa lunnaat, ei ole takeita siitä, että pääsy tiedostoihin palautetaan tai että haittaohjelma poistetaan järjestelmästä. Jos viides vaihe käynnistyy, hyökkääjä palauttaa pääsyn tiedostoihin laitteen omistajalle. (Tandon & Nayyar, 2019.)




Kuva 1 Kiristyshaittaohjelmahyökkäyksen eteneminen (muokattu lähteestä Tandon & Nayyar, 2019)

### 2.2.3 Palvelunestohyökkäys

Palvelunestohyökkäykset (engl. denial-of-service, DoS) ovat kyberhyökkäyksiä, joissa ylikuormitetaan järjestelmien, palvelimien tai verkkojen resursseja (AL-Hawamleh, 2023). Vaihtoehtoisesti hyökkääjä voi lähettää haitallisia paketteja uhrin koneelle hämätäkseen koneella käynnissä olevaa sovellusta. Palvelunestohyökkäyksessä hyökkääjään tavoitteena on estää käyttäjien pääsy verkkopalveluun tai verkkopalveluihin. Pääsy estetään häiritsemällä internetiin yhteydessä olevan järjestelmän tarjoamia palveluja. (Kaur Chahal ym., 2019.)

Perinteisesti palvelunestohyökkäykset ovat alkaneet yhdestä lähteestä, mutta myöhemmin niitä on tehty hajautetusti (Kaur Chahal ym., 2019). Hajautetussa

palvelunestohyökkäyksessä (engl. distributed denial-of-service, DDoS) ylikuormitus aiheutetaan useiden, yleensä tuhansien tai jopa miljoonien tietokoneiden samanaikaisella toiminnalla (AL-Hawamleh, 2023). Hajautettujen palvelunestohyökkäysten tavoitteena on estää täysin tiettyjen internetin palveluiden saatavuus (Kaur Chahal ym., 2019). Hajautetun palvelunestohyökkäyksen torjuminen ja käsittely on haastavaa hyökkäyslähteen määrittämisen vaikeuden vuoksi. Hyökkäyslähde voi olla vaikeaa määrittää hyökkääjien useiden maailmanlaajuisten IP-osoitteiden vuoksi. (AL-Hawamleh, 2023.) Hajautetut palvelunestohyökkäykset ovat kehittyneet yhdeksi internetin yleisimmistä uhista.

### 2.2.4 Muut kyberhyökkäykset

Tutkielmassa keskitytään haittaohjelmahyökkäyksiin ja palvelunestohyökkäyksiin. Muita kyberhyökkäyksiä ovat muun muassa esineiden internet -hyökkäys, pilvihyökkäys ja tietojenkalasteluhyökkäys.

Esineiden internet (engl. Internet of Things, IoT) on verkko, joka koostuu fyysisistä laitteista, ajoneuvoista, kodinkoneista ja muista sähköisistä esineistä, jotka ovat yhteydessä internetiin. Esineiden internet -hyökkäyksessä hyödynnetään esineiden internet -laitteita osana hyökkäystä. Yhä useammat esineet ovat yhteydessä esineiden internetiin, joka mahdollistaa yhä enemmän hyökkäyksiä. Kaksi yleisintä esineiden internet -hyökkäystä ovat spoofing-hyökkäykset ja palvelunestohyökkäykset. Jotta voidaan suorittaa onnistunut esineiden internet -hyökkäys, on kohteena olevan esineiden internet -laitteen täytettävä kolme ehtoa. Ensinnäkin on oltava mahdollista löytää laitteita, joihin voidaan kohdistaa hyökkäys. Toiseksi kohteena olevassa laitteessa tulee olla haavoittuvuuksia, jotka johtavat laitteen vaarantumiseen. Kolmanneksi laitteen pitää voida muodostaa yhteys suojaamattomien kanavien kautta, jotta haittaohjelma voi kommunikoida laiteen kanssa. (Shafiq ym., 2022.)

Pilvipalveluiden käytön yleistyessä yrityksissä yleistyvät myös pilvihyökkäykset (engl. cloud attacks). Yritykset varmuuskopioivat tiedostoja ja dataa digitaaliseen pilveen. Salauksen ja tunnistautumisen puute sekä asetusten virheellinen määritys ovat yleinen syy vaarantuneelle tietoturvalle pilvessä. Pilvihyökkäyksissä tarkoituksena on löytää haavoittuvuuksia, joiden avulla voitaisiin päästä käsiksi arkaluontoisiin tietoihin. (Mijwil ym., 2023.) Pilvipalveluihin kohdistuvissa hyökkäyksissä toinen yleinen hyökkäyksen toteuttamistapa on palvelunestohyökkäys. Pilvipalveluissa

palvelunestohyökkäysten tarkoituksena on kaataa palvelimet ja verkkoinfrastruktuuri, kuten palomuuri. Palvelunestohyökkäykset voivat aiheuttaa merkittäviä taloudellisia haittoja, jos yritys käyttää pilvipalveluita päivittäisissä toiminnoissa ja menettää pääsyn kyseisiin palveluihin palvelunestohyökkäyksen vuoksi. (Shafiq ym., 2022.)

Tietojenkalasteluhyökkäykset ovat nousseet huolenaiheeksi, koska monet internetin käyttäjät, sekä yksityishenkilöt että yritykset, lankeavat niihin. Tietojenkalasteluhyökkäys on sosiaalista manipulointia, jossa tietojenkalastelija pyrkii saamaan uhrin luovuttamaan arkaluontoisia tietojaan. Tietojenkalastelijat käyttävät hyväkseen kehittyneiden tekniikoiden sijaan ihmisluontoa. Suurin osa tietojenkalasteluhyökkäyksistä alkaa sähköpostiviestillä. Tietojenkalastelijat käyttävät sähköpostiviesteissä laittomasti hyödyksi organisaatioita, joihin uhri luottaa kerätäkseen arkaluontoista tietoa. Sähköpostiviesteihin on upotettu linkki, joka vie tietojenkalasteluhyökkäyksen uhrin haitalliselle verkkosivustolle, jossa tietojenkalastelu tapahtuu. (Alkhalil ym., 2021.) Tietojenkalastelija voi vaihtoehtoisesti esimerkiksi liittää näennäisesti luotettavaan sähköpostiin tiedostoja ja näin saada käsiinsä arkaluontoisia tietoja, kuten kirjautumistietoja tai luottokorttitietoja (AL-Hawamleh, 2023).

## 2.3 Miten kyberuhkiin on perinteisesti reagoitu

Kyberturvallisuuspuolustuksen strategiat suojaavat tietokonejärjestelmiin ja -verkkoihin liittyvää laitteistoa, ohjelmistoa tai dataa vahingoittumiselta. Tarkoituksena on estää tietomurto tai tietoturvapoikkeama. Perinteisesti kyberturvallisuuden mekanismeina on käytetty pääsynvalvontaa, palomuuria, viruksentorjuntaohjelmistoa, hiekkalaatikkoa (engl. sandbox), tietoturvatietojen ja -tapahtumien hallintaa ja salausta. (Sarker ym., 2021.)

Pääsynvalvonnassa säädellään resurssien, kuten tietokoneverkkojen tai datan, käyttöä. Esimerkiksi organisaatiossa sellaisten käyttäjien, jotka eivät tarvitse vastuidensa perusteella tiettyjä resursseja, pääsyä verkkoon voidaan rajoittaa. Näin organisaatiot ja yritykset voivat minimoida tietoturvariskit. Palomuuri seuraa ja säätelee saapuvaa ja lähtevää verkkoliikennettä. Palomuurit ovat verkkopohjaisia järjestelmiä, jotka perustuvat turvallisuussääntöihin. Palomuuri suodattaa liikennettä epäilyttävistä lähteistä välttämiseksi. Viruksentorjuntaohjelmisto on tietokoneohjelma, jolla havaitaan ja estetään tietokoneviruksien ja haittaohjelmien pääsy järjestelmään. Hiekkalaatikko on

tietoturvamekanismi, jolla lievennetään järjestelmähäiriöiden tai -haavoittuvuuksien leviämistä. Esimerkiksi epäluotettavat ohjelmat varmentamattomilta toimittajilta tai epäluotettavilta osapuolilta suoritetaan hiekkalaatikolla. Tietoturvatietojen ja -tapahtumien hallinta on yhdistelmä tietoturvatietojen hallintaa ja tietoturvatapahtumien hallintaa. Hallinta tarjoaa analyysin laitteiston tietoturvahälytyksistä. Salaus on menetelmä tietojen suojaamiseen. Salauksessa käytetään salaisia avaimia salaukseen ja tietojen purkamiseen. (Sarker ym., 2021.)

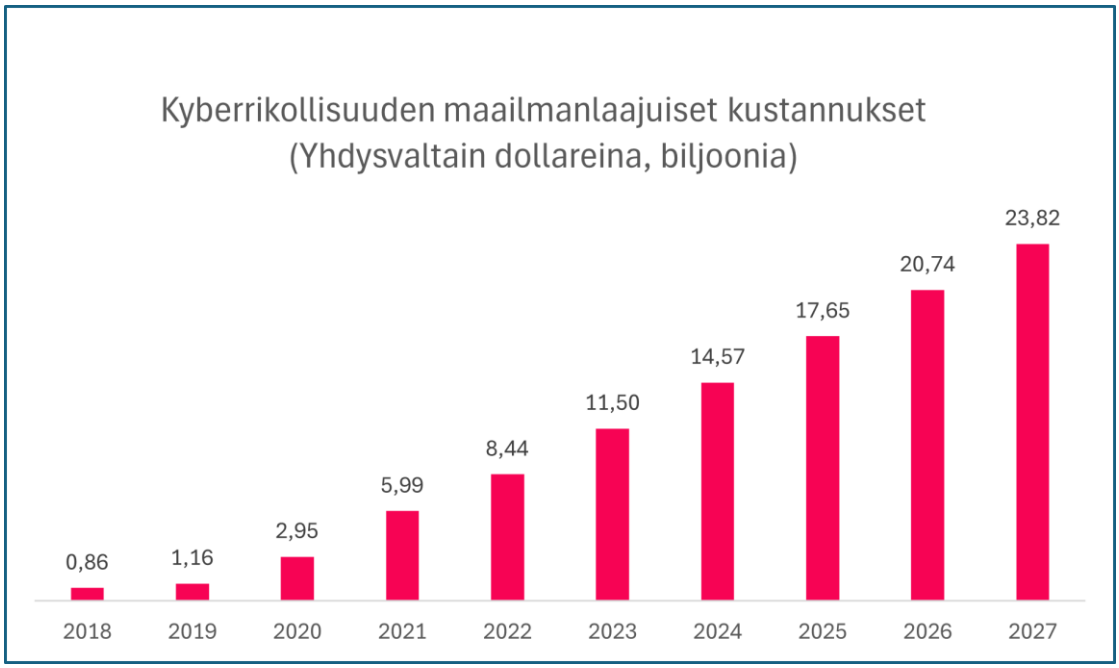
## 2.4 Kyberhyökkäykset tulevaisuudessa

Tulevaisuudessa odotetaan uusia strategioita kyberhyökkäyksiin. Kyberhyökkäykset tulevat kehittymään entistä monimutkaisemmiksi ja voimakkaimmiksi, joka tekee niiden torjunnasta entistä vaikeampaa.

Yritysten on ryhdyttävä varotoimiin arkaluonteisen liiketoimintansa suojelemiseksi, koska kyberhyökkäysten määrä kasvaa jatkuvasti. Etenkin yritysten, jotka toimivat alalla, jossa tallennetaan erityisen arkaluontoista dataa, kuten potilastietoja, on varauduttava kehittyviin kyberuhkiin. (Fadziso ym., 2023.) Kyberuhkia voidaan lieventää ja torjua rakentamalla yrityksen kulttuurista kyberturvallisuuskeskeinen. Työntekijöiden tulisi jokaisella yrityksen tasolla ymmärtää kyberturvallisuuden olevan keskeistä, jotta jokainen työntekijä voi tehdä panoksensa parantaakseen yrityksen kyberturvallisuutta. (Fisher ym., 2021.)

Kyberhyökkäysten havaitseminen on luonteeltaan ollut reaktiivista. Vain tiettyihin havaittuihin kaavoihin ja poikkeamiin on pystytty reagoimaan. Kyberhyökkäysten yleistyessä reaktiivisen lähestymistavan sijasta ennakoiva lähestymistapa on tarpeellinen. Ennakoivassa lähestymistavassa kyberhyökkäyksiin reagoidaan ennen kuin ne ehtivät aiheuttaa vahinkoa. (AL-Hawamleh, 2023.)

Kyberrikollisuuden maailmanlaajuisten kustannusten ennustetaan jatkavan jyrkkää kasvua noin kolmella biljoonalla Yhdysvaltain dollarilla vuodessa (ks. kuva 2 alla), kun yhä useammat ihmiset siirtyvät verkkoon. Näin myös kyberrikollisilla on enemmän mahdollisuuksia, joita käyttää hyödyksi. Etenkin koronaviruspandemian aikana kyberhyökkäyksissä tapahtui muutos etätyöskentelyn vuoksi. (Fleck, A, 2022).




Kuva 2 Kyberrikollisuuden maailmanlaajuiset kustannukset (muokattu lähteestä Fleck, A, 2022)

# 3 Tekoälytekniikat yritysten kohtaamien kyberuhkien torjumiseen

## 3.1 Tekoälytekniikat

### 3.1.1 Koneoppiminen

Koneoppiminen (engl. machine learning, ML) on tekoälyn haara, joka liittyy läheisesti laskennallisiin tilastoihin. Sekä koneoppimisessa että laskennallisissa tilastoissa keskitytään ennustamaan tietokoneiden avulla. Koneoppimisessa tietokoneilla on kyky oppia ilman erillistä ohjelmointia. Koneoppiminen keskittyy ensisijaisesti luokitteluun ja regressioon, joka pohjautuu aiemmin opittuihin piirteisiin, jotka on opittu opetusdatasta. (Xin ym., 2018.) Tyypillinen koneoppimisalgoritmi sisältää kaksi rinnakkaista toimintaa tai kaksi erillistä algoritmia, jotka ovat luokittelija ja oppimismalli. Luokittelijat ottavat syötteen eli joukon piirteitä ja tuottavat tulostekategorian. Oppimismalleiksi kutsuttavat algoritmit on ensin koulutettava testidatalla. Tämä testidata on valmiiksi luokiteltu ja tulostekategoria merkitty. (Burrell, 2016.)

Ohjaamaton oppiminen (engl. unsupervised learning) on koneoppimismalli, jota käytetään niin sanotussa tutkivassa aineiston analyysissa, kun luonnollisia etsittäviä ryhmiä ei tiedetä etukäteen ja analysoitavana on suuri aineisto. Ohjaamatonta oppimista käytetään myös, kun luokat ovat tiedossa etukäteen ja halutaan validoida koulutusprosessi ja muuttujien joukot. Ohjaamattomalla oppimisella on erilaisia tavoitteita, kuten ryhmittely, hierarkioiden luominen, ulottuvuuksien vähentäminen tai tulkinta ja visualisointi. Ohjattu oppiminen (engl. supervised learning) on koneoppimismalli, jossa käytetty algoritmi saa joukon esimerkkejä niiden vastausten kanssa, eli malli luodaan käyttämällä sen antamaa tulosta. (Martínez Torres ym., 2019.)

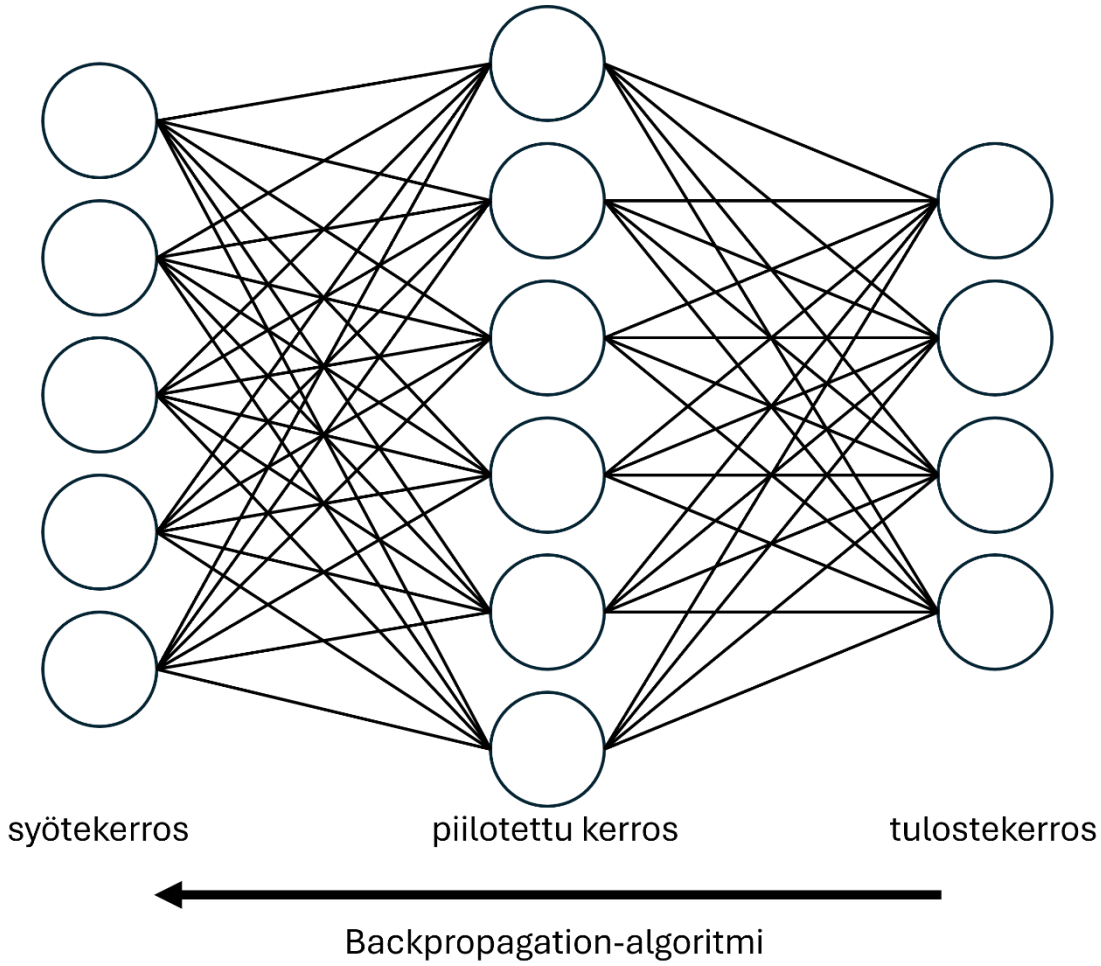
Koneoppimista käytetään esimerkiksi tunnistamaan esineitä kuvista, muuntamaan puhe tekstimuotoon ja sovittamaan sisältöä käyttäjien kiinnostuksenkohteisiin. Koneoppimista käytetään hyväksi monilla yhteiskunnan osa-alueilla. Koneoppiminen mahdollistaa esimerkiksi verkkohaut, sisällönsuodatuksen sosiaalisessa mediassa ja suositusten antamisen verkkosivustoilla. Koneoppiminen on läsnä useissa kuluttajatuotteissa, kuten älypuhelimissa. (LeCun ym., 2015.)

Perinteiset koneoppimistekniikat ovat rajoittuneita kyvyssään käsitellä raakaa dataa. Koneoppiminen vaatii erityistä osaamista ja asiantuntemusta suunnitella piirteen erotin, joka muuttaa raakadatan sisäiseksi esitykseksi, josta oppimisalaosa pystyy havaitsemaan kuvion syötteestä. Ratkaisuna perinteisen koneoppimisen ongelmiin on kehitetty menetelmiä, jotka mahdollistavat tarvittavien edustusten havaitsemisen tai luokittelun automaattisesti. (LeCun ym., 2015.)

### 3.1.2 Syväoppiminen

Syväoppiminen (engl. deep learning, DL) on dataoppimisen karakterisointiin perustuva koneoppimismenetelmä (Xin ym., 2018). Syväoppiminen mahdollistaa sen, että useista käsittelykerroksista koostuvat laskennalliset mallit oppivat tietojen esityksiä useilla abstraktiotasoilla. Nämä menetelmät ovat parantaneet muun muassa puheentunnistusta ja visuaalisen objektin tunnistusta. (LeCun ym., 2015.)

Syväoppiminen käyttää backpropagation-algoritmia löytääkseen monimutkaisen rakenteen suurista tietojoukoista. Syväoppimismenetelmät ovat oppimismenetelmiä, joissa on useita edustustasojen tasoja, jotka on saatu yhdistämällä yksinkertaisia moduuleja monikerroksisiksi verkoiksi. Nämä moduulit muuttavat edustusta yhdellä kerroksella hieman abstraktimmalle tasolle. (LeCun ym., 2015.) Neuroverkossa joukko syötteen (engl. input) solmuja yhdistyvät toiseen sarjaan solmuja, joita kutsutaan piilotetuiksi kerroksiksi. Jokainen syötteen solmu on yhteydessä piilotetun kerroksen solmuun, ja jokainen piilotetun kerroksen solmu on yhteydessä tulosteeseen (engl. output). (Burrell, 2016.) Kuvassa 3 esitetään, miten backpropagation-algoritmin solmut ja kerrokset yhdistyvät toisiinsa. Kun suoritetaan tarpeeksi tällaisia muutoksia, voivat menetelmät oppia hyvinkin monimutkaisia toimintoja. Korkeammat edustustasot korostavat syötteen tärkeitä osia erottelussa ja tukahduttavat epäolennaiset vaihtelut. Syväoppimisen keskeinen näkökohta on, että piirteiden tasoja eivät suunnittele insinöörit, vaan järjestelmä oppii ne datasta. (LeCun ym., 2015.) On huomionarvoista, että neuroverkko ei käsittele dataa tavalla, joka olisi helposti ymmärrettävissä ihmisille. Tämä ei-kaavamainen painotus johtuu laskennallisen oppimisen käsitteestä. Syväoppimista sovelletaan sellaisiin ongelmiin, joissa eksplisiittisen päätöksentekologiikan ohjelmointi toimii huonosti. (Burrell, 2016.)




Kuva 3 Backpropagation-algoritmi (muokattu lähteestä Burrell, 2016)

Syväoppiminen on mahdollistanut suuren edistysaskeleen ottamisen monien sellaisten ongelmien ratkaisemisessa, joita tekoäly-yhteisö ei ole yrityksistä huolimatta saanut aiemmin ratkaistua. Syväoppiminen on sovellettavissa monille tieteenaloille ja liike-elämän alueille, koska sillä on kyky löytää monimutkaisia rakenteita korkeaulotteisesta datasta. (LeCun ym., 2015.)

### 3.1.3 Erot koneoppimisen ja syväoppimisen välillä

Riippuvuus datasta on koneoppimisen ja syväoppimisen suurin ero. Koneoppimisen ja syväoppimisen suorituskykyihin vaikuttaa datan määrä. Syväoppimisalgoritmit vaativat enemmän dataa kuin koneoppimisalgoritmit. Syväoppimisalgoritmit eivät toimi yhtä hyvin kuin koneoppimisalgoritmit, kun dataa on vähän. Tämä johtuu siitä, että syväoppimisalgoritmit vaativat suuren määrän dataa ymmärtääkseen datan täydellisesti. (Xin ym., 2018.)

Sekä koneoppimisalgoritmit että syväoppimisalgoritmit ovat riippuvaisia laitteistosta. Syväoppimisalgoritmit vaativat useita matriisioperaatioita, joissa hyödynnetään grafiikkasuoritinta (engl. graphics processing unit, GPU). Tämän vuoksi syväoppiminen tarvitsee grafiikkasuorittimen toimiakseen tehokkaasti. Syväoppiminen on riippuvainen suorituskykyisemmistä koneista, joissa on grafiikkasuoritin, kuin koneoppiminen. (Xin ym., 2018.)

Monien koneoppimisalgoritmien suorituskyky riippuu piirteiden tarkkuudesta. Piirteiden käsittely on aikaa vievää ja vaatii alan asiantuntijan erikoistietämystä. Piirteen käsittely on prosessi, jossa piirteen erottimeen syötetään arvoalueen tuntemus datan monimutkaisuuden vähenemiseksi ja oppimisalgoritmien toimintaa parantavien kuvioiden luomiseksi. Piirteet voivat olla esimerkiksi pikseliarvoja, muotoja, tekstuureja tai sijainteja. Toisin kuin koneoppimisessa, syväoppimisessa piirteiden hankkiminen tapahtuu suoraan datasta. Syväoppiminen säästää vaivaa ja aikaa piirteen erottimeen suunnittelussa. Tämä luo merkittävän eron koneoppimisen ja syväoppimisen välillä. (Xin ym., 2018.)

Koneoppimisen ja syväoppimisen koulutus- ja testiajoissa on huomattava ero. Syväoppimisalgoritmin koulutus kestää kauemmin kuin koneoppimisen koulutus. Syväoppimisen koulutus saattaa kestää jopa viikkoja koneoppimisen koulutuksen kestäessä vain sekunneista tunteihin. Testiaika taas on täysin päinvastainen. Syväoppimisalgoritmien testaaminen vaatii vain vähän aikaa, kun taas jotkin koneoppimisalgoritmit vaativat sitä enemmän aikaa mitä enemmän dataa on. Tämä ei kuitenkaan päde jokaiseen koneoppimisalgoritmiin, vaan osan koneoppimisalgoritmien testaus on nopeaa, kuten syväoppimisalgoritmien testaus. (Xin ym., 2018.)

## 3.2 Mitä tekoälytekniikoita käytetään minkäkin kyberuhan torjumiseen

Keskeinen ongelma perinteisten järjestelmien hyödyntämisessä kyberuhkien torjumisessa ja minimoinnissa on, että niistä vastaavat tietoturva-asiantuntijat, joiden voi olla vaikeaa vastata yrityksen kyberturvallisuustarpeisiin tehokkaasti, älykkäästi ja järjestelmällisesti. Tekoälyllä on mahdollisuus laskentatehonsa ja kykyjensä ansiosta tarjota automatisoituja, tehokkaita ja älykkäitä kyberturvallisuuspalveluita. (Sarker ym., 2021.)

### 3.2.1 Koneoppimisen soveltaminen haittaohjelmien ja palvelunestohyökkäysten estoon

Koneoppiminen on tärkeä osa tekoälyä, jota voidaan hyödyntää tietoturvamallinnuksen rakentamiseen hyödyntäen tietoturvadataa. Koneoppimisen tietoturvamallissa käytetään yleensä eri lähteistä koottua dataa ja algoritmeja, jotka toimivat kyseisellä datalla. Koottuun dataan sisältyy dataa verkkokäyttäytymisestä, käyttäjäaktiivisuudesta ja sovellusaktiivisuudesta. (Sarker ym., 2021.) Koneoppimiseen pohjautuvat tietoturvajärjestelmät ovat hyödyllisiä DoS-hyökkäysten ja haittaohjelmahyökkäysten havaitsemisessa (Alashhab ym., 2022).

Koneoppimista hyödynnetään haittaohjelmien havaitsemisessa useilla eri tavoilla. Koneoppimismallit hyödyntävät erilaisia ominaisuuksia, kuten tavusekvenssejä (engl. byte sequences), API-kutsuja, N-grammeja, järjestelmäkutsuja, operaatiokoodeja (engl. opcodes) ja muita haittaohjelmien käyttäytymiseen liittyviä tietoja. Näitä ominaisuuksia kerätään analysoimalla haittaohjelmanäytteitä joko staattisesti eli ilman suorittamista tai dynaamisesti eli suorittamisen aikana. Koneoppimisalgoritmeja, kuten päätöspuita, satunnaismetsiä ja tukivektoreita voidaan kouluttaa tunnistamaan haittaohjelmia tunnetuista näytteistä. Koulutuksen jälkeen näitä malleja käytetään ennustamaan uusien näytteiden luokitusta reaaliajassa, jolloin haittaohjelmia voidaan tunnistaa ja estää leviämästä. Haittaohjelmien havaitsemisessa koneoppimismallien avulla käytetään vaihtelevia oppimisstrategioita. Haittaohjelmien havaitsemisessa hyödynnetään sekä ohjattua että ohjaamatonta oppimista. Joissakin tapauksissa myös puoliohjattua oppimista voidaan hyödyntää. Haittaohjelmien toimintaan liittyvät poikkeamat, kuten epätavalliset verkkoaktiviteetit tai järjestelmätason muutokset, voidaan tunnistaa ja käsitellä ennakkoon koneoppimisalgoritmien avulla. Koneoppimisalgoritmit voivat havaita poikkeamat ennen kuin ne aiheuttavat häiriötä. (Ucci ym., 2019.)

Haittaohjelmat eivät ole uhka ainoastaan tietokonejärjestelmille vaan myös älylaitteille, kuten älypuhelimille, etenkin niille, jotka käyttävät Android-käyttöjärjestelmää. Kolmannen osapuolen sovellukset voivat tartuttaa älypuhelimia ja vaikuttaa niihin samankaltaisesti kuin tietokoneisiin. Älylaitteiden haittaohjelmien torjumiseen on kehitetty useita koneoppimiseen perustuvia malleja. Päätöspuita on testattu useissa tutkimuksissa hyvin tuloksin. Naive Bayes -menetelmä suoriutui paremmin kuin muut luokittelumenetelmät, kuten päätöspuut, Baeysian verkot tai SVM. Parhaat tulokset on

kuitenkin saavutettu käyttäessä monipiirreyhteispäätössovitusmenetelmää (engl. multifeature collaborative decision fusion), johon sisältyivät SVM, Naive Bayes ja päätöspuut. (Martínez Torres ym., 2019.)

Palvelunestohyökkäysten lisääntyessä tehokkaiden puolustusmekanismien kysyntä kasvaa. Poikkeamien tunnistaminen verkossa suoritetaan usein verkkopohjaisten tunkeutumisen havainnointi- ja estojärjestelmien (engl. network-based intrusion detection and prevention systems, NIDPS) avulla. NIDPS:t mahdollistavat tunnettujen hyökkäysten pysäyttämisen, mutta NIDPS:t eivät kuitenkaan ole kestäviä ratkaisuja palvelunestohyökkäysten aiheuttamien poikkeamien jatkuvan vaihtelun vuoksi. Koneoppisparadigma tarjoaa algoritmeja, jotka voivat tehokkaasti vähentää käsitevirran muutoksia kyberuhkien tietomallien kehittyessä. Koneoppimisen avulla voidaan kehittää järjestelmiä, jotka oppivat tunnistamaan normaalin verkkoliikenteen ja havaitsemaan poikkeavuudet ajoissa ennen häiriöitä. (Coscia ym., 2024.) Koneoppimisalgoritmit voivat olla ratkaisevassa asemassa palvelunestohyökkäysten havaitsemisessa. On kuitenkin tarve yhdistää koneoppimiseen perustuvat palvelunestohyökkäysten havaitsemisjärjestelmät yhteen kyberturvallisuusalustan, kun NIDPS:n kanssa. (Khalaf ym., 2019.) Tämä voidaan saavuttaa selittämällä koneoppimisalgoritmien tuottamat tulokset ja lähtötiedot käyttämällä joukkoa menetelmiä, jotka kuuluvat selittävän tekoälyn (engl. explainable AI, XAI) alaan (Gilpin ym., 2018). Koneoppimismallin selittäminen pyrkii tarjoamaan mallin itsensä päättelemät loogiset päätössäännöt ihmiselle ymmärrettävässä muodossa. Yleisesti ottaen, mitä korkeampi tulkittavuus mallilla on, sitä parempi luotettavuus ennustejärjestelmällä on. Näin ollen korkea yhteensopivuus tarkoittaa kykyä ymmärtää, miten koneoppimismalli tekee ennusteita ja miten kukin tieto-ominaisuus vaikuttaa ennusteeseen. (Coscia ym., 2024.) Selittävää koneoppimismallia valittaessa suositaan mahdollisimman yksinkertaista mallia (Bargagli Stoffi ym., 2022).

### 3.2.2 Syväoppimisen soveltaminen haittaohjelmien ja palvelunestohyökkäysten estoon

Haittaohjelmahyökkäyksissä on havaittu kaavoja, joita voidaan hyödyntää tekoälyn kouluttamisessa. Perinteisten koneoppimisalgoritmien toiminta ja tehokkuus ovat rajoittuneita koulutussyvyyden ollessa rajoittunut. Syväoppimisen myötä haittaohjelmahyökkäysten havaitseminen ja tunnistustarkkuus ovat parantuneet.

Syväoppimismallien syntyminen on luonut uudenlaisia koulutusmahdollisuuksia. Tämä mahdollistaa tekoälyn hyödyntämisen haittaohjelmahyökkäysten tehokkaassa torjumisessa. Haittaohjelmien torjumisessa hyödynnettyjä syväoppimismenetelmiä ovat haittaohjelmien ennustamiseen käytetty konvoluutioneuroverkko (engl. convolutional neural network, CNN), haittaohjelmien havaitsemiseen käytetyt takaisinkytkeytyvä neuroverkko (engl. recurrent neural network, RNN) ja long short-term memory (LSTM) ja poikkeamien havaitsemiseen käytetyt automaattiset koodittajat (engl. auto encoders). (Majid ym., 2023.)

Syväoppimista voidaan hyödyntää palvelunestohyökkäysten havaitsemisessa useilla tavoilla, jotka perustuvat syväoppimistekniikoiden kykyyn mallintaa ja tunnistaa monimutkaisia malleja suurista määristä dataa. Monikerroksista perseptroniverkkoa (engl. multilayer perceptron, MLP) ja geneettistä algoritmia (engl. genetic algorithm, GA) käytetään luokittelumalleina optimoimaan luokitteluprosessia. Menetelmä oppii erottelemaan hyökkäysten ja normaalin liikenteen piirteitä. Automaattiset koodittajat ovat valvomattomia ja generatiivisia malleja, jotka oppivat tunnistamaan tehokkaasti poikkeavuuksia verkkoliikenteessä. Takaisinkytkeytyviä neuroverkkoja ja LSTM:ää käytetään sekvenssidatan, kuten verkkoliikenteen, käsittelyyn. Nämä mallit oppivat ja muistavat tietoja pitkään, joka on hyödyllistä erityisesti palvelunestohyökkäysten dynaamisuuden vuoksi. Konvoluutioneuroverkkoa voidaan käyttää havaitsemaan ja luokittelemaan rakenteellisia piirteitä verkkoliikenteestä. Syvä uskomusverkko (engl. deep belief network, DBN) on tehokas tunnistamaan monimutkaisia malleja sovelluskerroksen hyökkäysten havaitsemisessa. (Malliga ym., 2022.)

## 3.3 Tekoälyn hyödyntämisen uhat kyberturvallisuudelle

Luottamuksessa tekoälyyn kyberturvallisuustehtävien suorittamisessa on riskinsä. Tekoälypohjaisten kyberturvallisuusmallien käyttöönotto voi parantaa merkittävästi kyberturvallisuutta ja se käytäntöjä, mutta samalla se voi myös edesauttaa luomaan uusia hyökkäysmuotoja, jotka kohdistuvat tekoälysovelluksiin. Tämä on vakava uhka kyberturvallisuudelle. (Taddeo ym., 2019.)

Perinteisesti kyberhyökkäysten tarkoituksena on ollut tietojen varastaminen ja järjestelmien häirintä ja rikkominen. Tekoälyjärjestelmiin kohdistuvien uusien hyökkäysten tarkoituksena voi olla järjestelmän haltuunotto ja sen käyttäytymisen muuttaminen. Järjestelmän haltuunottoon on kolme päätapaa, jotka ovat datamyrkytys

(engl. data poisoning), luokittelumallien temperointi (engl. tempering of categorization models) ja takaportit (engl. backdoors). Kaikki näistä tavoista käyttävät hyödykseen tekoälyjärjestelmien oppimiskykyä muuttaakseen tekoälyjärjestelmän käyttäytymistä. Hyökkääjät esimerkiksi tuovat järjestelmän aidon koulutusdatan sekaan huolellisesti muotoiltua virheellistä dataa, jotta järjestelmän käyttäytyminen muuttuisi. Luokittelumallien manipuloinnilla tutkijat ovat muun muassa 3D-kuvia käyttämällä opettaneet tekoälyjärjestelmää luokittelemaan kilpikonnat kivääreiksi. Hyökkääjät voivat siis esimerkiksi opettaa järjestelmää luokittelemaan tietyt kuvat aivan eri luokkiin kuin oli alkuperäisen datan pohjalta järjestelmää opetettu. Takaporttiin perustuvat hyökkäykset hyödyntävät piilotettuja yhteyksiä, jotka niin ikään lisätään tekoälyjärjestelmään luokittelun muuttamiseksi. Näin saadaan järjestelmä suorittamaan hyökkääjän haluamia toimintoja. (Taddeo ym., 2019.) Tutkimuksessa pysähtymismerkkeihin lisättiin erityinen tarra, jonka jälkeen pysähtymismerkit lisättiin tekoälyjärjestelmän koulutusdataan nopeusrajoitusmerkkien joukkoon. Tekoälyjärjestelmä luokitteli kaikki pysähtymismerkit, jotka sisälsivät kyseisen tarran, nopeutusrajoitusmerkeiksi. Autonomisten ajoneuvojen toiminnan kannalta tämä aiheutti vakavia turvallisuusriskejä autojen ajaessa suoraan risteysten läpi pysähtymisen sijaan. (Eykholt ym., 2018.)

Tekoälyä vastaan käynnistetty hyökkäys on vaikea havaita, koska tekoälyjärjestelmät ovat luonteeltaan verkottuneita, dynaamisia ja mukautuvia. Tällainen luonne tekee sisäisten prosessien selittämisestä ongelmallista, koska läpinäkyvyys puuttuu. Lisäksi käyttäytymistä ja sitä, mikä on johtanut tiettyyn lopputulokseen, on vaikea ymmärtää. Tekoälyyn kohdistuvat hyökkäykset voivat myös johtaa harhaan. Tekoälyjärjestelmät saattavat jatkaa odotettua käyttäytymistä manipuloinnin jälkeen, kunnes hyökkääjät laukaisevat järjestelmän käytöksen muutoksen. Vaikka hyökkäys havaittaisiin, voi olla vaikeaa huomata, mikä tekoälyjärjestelmän käyttäytymisessä on muuttunut, koska monet hyökkäykset ovat suunniteltu tarkasti ja taitavasti aiheuttamaan vain pienen, mutta tavoitteiden saavuttamisen kannalta riittävän poikkeaman käyttäytymisessä. (Taddeo ym., 2019.)

Tekoälyjärjestelmän luotettavuuden varmistaminen häirintätilanteissa on käytännössä mahdotonta, koska mahdollisten häiriöiden määrä on suuri. Kaikkien mahdollisten häiriöiden ennakoiminen ja poikkeamien havaitseminen on laskennallisesti ratkaisematon ongelma. Niin kauan kuin luotettavuutta kyberturvallisuuden

tekoälyjärjestelmiin ei voida arvioida, ei tekoälyjärjestelmiin kyberturvallisuudessa perustellusti voida täysin luottaa. Tämä ei tarkoita sitä, etteikö osaa kyberturvallisuuteen liittyvistä tehtävistä voitaisi delegoida tekoälylle sen osoittautuessa kykeneväksi suorittamaan ne tehokkaasti. Tekoäly ei kuitenkaan voi yksin vastata kyberturvallisuuteen liittyvistä tehtävistä, vaan muita valvontamuotoja tarvitaan lieventämään riskejä, jotka liittyvät tekoälyn ongelmiin. (Taddeo ym., 2019.)

# 4 Yhteenveto ja johtopäätökset

Tässä tutkielmassa tutkittiin tekoälyn sovellusmahdollisuuksia kyberturvallisuuden alalla. Näkökulma rajattiin kyberuhkien laajuuden vuoksi haittaohjelmahyökkäyksiin ja palvelunestohyökkäyksiin. Kyberrikollisuuden kasvaessa ja kyberuhkien ja -hyökkäysten monipuolistuessa kyberturvallisuuden alalla on tarve kehitykseen ja innovaatioihin. Kyberhyökkäykset kehittyvät jatkuvasti ja perinteiset toimenpiteet niiden havaitsemiseen ja torjumiseen eivät enää ole riittäviä. Tutkielmassa selvitettiin, millaisia tekoälytekniikoita voidaan hyödyntää haittaohjelmahyökkäysten ja palvelunestohyökkäysten havaitsemiseen ja torjumiseen. Tutkielmassa arvioitiin tekoälyn hyötyjen lisäksi myös mahdollisia uhkia, joita tekoälyn hyödyntämisestä voi seurata.

Tutkielman toisessa luvussa käsiteltiin kyberuhkia ja -hyökkäyksiä, joita yritykset kohtaavat. Toisessa luvussa vastattiin tutkielman toiseen tutkimuskysymykseen: millaisia kyberuhkia yritykset kohtaavat? Yleisiä kyberuhkia ovat haittaohjelmahyökkäykset ja palvelunestohyökkäykset. Tutkielmassa kuvattiin, miten nämä kyberhyökkäykset toteutetaan ja mitä hyökkääjä hyökkäyksellä tavoittelee. Hyökkäysten taustalla on usein taloudelliset tai poliittiset motiivit. Haittaohjelmahyökkäysten ja palvelunestohyökkäysten lisäksi yleisiä kyberhyökkäyksiä ovat esineiden internet -hyökkäykset, pilvihyökkäykset ja tietojenkalasteluhyökkäykset. Kyberhyökkäyksiin on perinteisesti reagoitu vasta, kun kyberhyökkäys on havaittu. Perinteisesti kyberhyökkäysten havaitsemiseen ja torjumiseen on käytetty useita ohjelmia, kuten pääsynvalvontaa, palomuuria, viruksentorjuntaohjelmistoa, hiekkalaatikkoa, tietoturvatietojen ja -tapahtumien hallintaa ja salausta. Kyberhyökkäyksiin odotetaan uusia strategioita tulevaisuudessa. Kyberhyökkäysten torjunta muuttuu entistä vaikeammaksi niiden kehittyessä monimutkaisemmiksi. Kyberhyökkäysten kehittyessä on keskeistä muuttaa lähestymistapa reaktiivisesta ennakoivaksi, jolloin kyberhyökkäyksiin reagoidaan ennen vahingon syntymistä.

Reaktiivisen lähestymistavan mahdollistaa tekoäly. Kolmannessa luvussa käsiteltiin tekoälytekniikoiden hyödyntämistä kyberhyökkäysten torjumiseen. Kolmannessa luvussa vastattiin tutkielman kolmanteen tutkimuskysymykseen: millaisia tekoälytekniikoita yritykset käyttävät kyberuhkien torjuntaan? Kirjallisuuden perusteella kyberturvallisuudessa tekoälytekniikoista hyödynnetään erityisesti koneoppimista ja

syväoppimista. Luvussa pohjustettiin, miten koneoppiminen ja syväoppiminen toimivat ja miten ne eroavat toisistaan. Tämän jälkeen käsiteltiin, mitä tekoälytekniikoita käytetään minkäkin kyberuhan torjumiseen ja miten. Sekä koneoppimista että syväoppimista voidaan tieteellisen kirjallisuuden ja tutkimusten perusteella hyödyntää sekä haittaohjelmahyökkäysten että palvelunestohyökkäysten torjumiseen. Koneoppimisen menetelmistä haittaohjelmahyökkäysten havaitsemiseen ja torjumiseen käytetään tavusekvenssejä, API-kutsuja, N-grammeja, järjestelmäkutsuja ja operaatiokoodeja. Koneoppimisen menetelmistä palvelunestohyökkäysten havaitsemiseen ja torjumiseen käytetään selittävän tekoälyn ja verkkopohjaisten tunkeutumisen havainnointi- ja estojärjestelmien (NIDPS) yhdistelmää. Syväoppimisen metodeja CNN, RNN, LSTM ja automaattisia koodittajia käytetään haittaohjelmahyökkäysten tunnistamiseen. Syväoppimisen metodeja MLP, GA, LSTM, RNN ja DBN käytetään palvelunestohyökkäysten monimutkaisten mallien ja kaavojen tunnistamiseen. Taulukossa 1 vedetään yhteen, mitä tekoälytekniikkaa ja menetelmiä käytetään minkäkin kyberuhan torjumiseen.

Taulukko 1 Tekoälyn hyödyntäminen ja menetelmät

| Tekoälytekniikka | Kyberuhka | Sovellus | Menetelmät |
|---|---|---|---|
| Koneoppiminen | Haittaohjelmahyökkäys | Haittaohjelmien tunnistaminen ja käsittely | Tavusekvenssit, API-kutsut, N-grammit, järjestelmäkutsut, operaatiokoodit |
| Koneoppiminen | Palvelunestohyökkäys | Poikkeamien tunnistaminen verkossa | Selittävän tekoälyn ja NIDPS:n yhdistelmät |
| Syväoppiminen | Haittaohjelmahyökkäys | Haittaohjelmien tunnistustarkkuuden parantaminen | CNN, RNN, LSTM, automaattiset koodittajat |
| Syväoppiminen | Palvelunestohyökkäys | Monimutkaisten mallien tunnistus suuresta datan määrästä | MLP, GA, LSTM, RNN, DBN |

Tutkielman perusteella tekoälyn rooli kyberturvallisuudessa on merkittävä. Tekoälytekniikat, joita hyödynnetään kyberturvallisuudessa tulevat kehittymään, mutta luottamuksessa tekoälyyn kyberturvallisuustehtävien suorittamisessa on riskinsä. Samalla kun tekoäly parantaa kyberturvallisuutta, voi se myös edesauttaa kyberhyökkäysten kehittymistä. Osa kyberturvallisuuden tehtävistä voidaan delegoida tekoälyjärjestelmille, mutta kyberturvallisuutta ei tulisi jättää täysin tekoälyjärjestelmien

vastuulle, vaan myös muita valvontamuotoja tarvitaan kyberturvallisuuden takaamiseksi. Tulevaisuudessa olisi mahdollisesti hyödyllistä tutkia sitä, miten tekoälyn käyttäminen kyberturvallisuuden toiminnoissa vaikuttaa kyberhyökkäysten kehittymiseen ja monimutkaistumiseen, kun kyberrikolliset saatavat kyberturvallisuuteen käytettäviä tekoälymenetelmiä ja -ohjelmia käyttöönsä.

## Lähteet

Alashhab, A. A., Zahid, M. S. M., Azim, M. A., Daha, M. Y., Isyaku, B., & Ali, S. (2022). A survey of low rate ddos detection techniques based on machine learning in software-defined networks. *Symmetry*, *14*(8), 1563. https://doi.org/10.3390/sym14081563

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

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

Bargagli Stoffi, F. J., Cevolani, G., & Gnecco, G. (2022). Simple models in complex worlds: Occam's razor and statistical learning theory. *Minds and Machines*, *32*(1), 13–42. https://doi.org/10.1007/s11023-022-09592-z

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big data & society*, *3*(1), 2053951715622512. https://doi.org/10.1177/2053951715622512

Cisco. (2018). *Cisco 2018 Annual Cybersecurity Report*. https://www.cisco.com/c/dam/m/hu_hu/campaigns/security-hub/pdf/acr-2018.pdf. Haettu 25.04.2024.

Coscia, A., Dentamaro, V., Galantucci, S., Maci, A., & Pirlo, G. (2024). Automatic decision tree-based NIDPS ruleset generation for DoS/DDoS attacks. *Journal of Information Security and Applications*, *82*, 103736. https://doi.org/10.1016/j.jisa.2024.103736

Eykholt, K., Evtimov, I., Fernandes, E., Li, B., Rahmati, A., Xiao, C., Prakash, A., Kohno, T., & Song, D. (2018). Robust Physical-World Attacks on Deep Learning Visual Classification. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1625–1634. https://doi.org/10.1109/CVPR.2018.00175

Fadziso, T., Thaduri, U., Dekkati, S., Ballamudi, V., & Desamsetti, H. (2023). Evolution of the cyber security threat: An overview of the scale of cyber threat. *Digitalization & Sustainability Review*, *3*(1), 1–12. https://doi.org/10.6084/m9.figshare.24189921.v1

Fisher, R., Porod, C., & Peterson, S. (2021). Motivating Employees and Organizations to Adopt a Cybersecurity-Focused Culture. *Journal of Organizational Psychology*, *21*(1). https://doi.org/10.33423/jop.v21i1.4030

Fleck, A. (2022). *Cybercrime Expected To Skyrocket in Coming Years [Digital image]*. https://www.statista.com/chart/28878/expected-cost-of-cybercrime-until-2027/. Haettu 25.04.2024.

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

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining Explanations: An Overview of Interpretability of Machine Learning. *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 80–89. https://doi.org/10.1109/DSAA.2018.00018

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

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

Khalaf, B. A., Mostafa, S. A., Mustapha, A., Mohammed, M. A., & Abduallah, W. M. (2019). Comprehensive Review of Artificial Intelligence and Statistical Approaches in Distributed Denial of Service Attack and Defense Methods. *IEEE Access*, *7*, 51691–51713. https://doi.org/10.1109/ACCESS.2019.2908998

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, *521*, 436–444. https://doi.org/10.1038/nature14539

Majid, A.-A. M., Alshaibi, A. J., Kostyuchenko, E., & Shelupanov, A. (2023). A review of artificial intelligence based malware detection using deep learning. *Materials Today: Proceedings*, *80*, 2678–2683. https://doi.org/10.1016/j.matpr.2021.07.012

Malliga, S., Nandhini, P., & Kogilavani, S. (2022). A Comprehensive Review of Deep Learning Techniques for the Detection of (Distributed) Denial of Service Attacks. *Information Technology and Control*, *51*, 180–215. https://doi.org/10.5755/j01.itc.51.1.29595

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

Mijwil, M., Unogwu, O., Filali, Y., Bala, Indu, & Al-Shahwani, Humam. (2023). Exploring the Top Five Evolving Threats in Cybersecurity: An In-Depth Overview. *Mesopotamian journal of Cybersecurity*, *2023*, 57–63. https://doi.org/10.58496/MJCS/2023/010

Morgan, S. (2023, heinäkuuta 28). Top 10 Cybersecurity Predictions And Statistics For 2023. *Cybercrime Magazine*. https://cybersecurityventures.com/top-5-cybersecurity-facts-figures-predictions-and-statistics-for-2021-to-2025/. Haettu 25.04.2024.

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

Shafiq, M., Gu, Z., Cheikhrouhou, O., Alhakami, W., & Hamam, H. (2022). The Rise of "Internet of Things" Review and Open Research Issues Related to Detection and Prevention of IoT-Based Security Attacks. *Wireless Communications and Mobile Computing*, *2022*, 12. https://doi.org/10.1155/2022/8669348

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-in-cybersecurity-exploring-the-advantages-and-risks/?sh=69beb78429c7. Haettu 25.04.2024.

Taddeo, M., McCutcheon, T., & Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nature Machine Intelligence*, *1*(12), 557–560. https://doi.org/10.1038/s42256-019-0109-1

Tandon, A., & Nayyar, A. (2019). A Comprehensive Survey on Ransomware Attack: A Growing Havoc Cyberthreat: Proceedings of ICDMAI 2018, Volume 2. Teoksessa *Advances in Intelligent Systems and Computing* (ss. 403–420).

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

Ucci, D., Aniello, L., & Baldoni, R. (2019). Survey of machine learning techniques for malware analysis. *Computers & Security*, *81*, 123–147. https://doi.org/10.1016/j.cose.2018.11.001

Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., Mingcheng, G., Hou, H., & Wang, C. (2018). Machine Learning and Deep Learning Methods for Cybersecurity. *IEEE Access*, *PP*, 1–1. https://doi.org/10.1109/ACCESS.2018.2836950