Tekoälyn käyttäminen tietoturvahyökkäyksissä

Oulun yliopisto
Tietojenkäsittelytiede
LuK-tutkielma
Iiro Candelin
2024

Tiivistelmä

Tekoälyn kehitys tuo mukanaan koko ajan uusia uhkia, joista yksi on itse tekoäly. Tekoälyn käyttö tietoturvahyökkäyksissä tekee kyberrikollisuudesta vaarallisempaa, kuin se ennen on ollut. Kyberrikolliset voivat käyttää tekoälyä esimerkiksi automatisoidakseen hyökkäyksiään, tehostaakseen haittaohjelmiaan tai käyttääkseen tekoälyä itseään hyökkäystyökaluna. Kun tekoälyä käytetään esimerkiksi haittaohjelmissa apuna, se pystyy tekemään haittaohjelmasta autonomisemman, kehittyneemmän, nopeamman sekä vaikeammin havaittavan. Tähän uuteen uhkaan on siis varauduttava sekä tutkittava, kuinka kyberrikolliset käyttävät tekoälyä apunaan. Kun tekoälyä käytetään itsestään hyökkäystyökaluna, kyberrikolliset hyökkäävät tekoälyä vastaan syöttöhyökkäyksillä. Hyökkäykset voivat olla myös fyysisiä esimerkiksi stop-merkistä voidaan teipata jokin pieni osa piiloon, jonka seurauksena itseajavien ajoneuvojen tekoäly ei enää tunnista enää stop-merkkiä. Tekoälyyn voidaan vaikuttaa myös datamyrkytyshyökkäyksillä. Tämä hyökkäystapa eroaa syöttöhyökkäyksistä siten, että sen tarkoituksena on syötteen muuttaminen pitkän ajan kuluessa, jotta tekoälyn analysointitiedot siirtyvät ja ovat jo luonnostaan puutteellisia.

Tekoälyssä käytetään kahta eri teknologiaa, jotka ovat koneoppiminen ja syväoppiminen. Koneoppiminen on keskeinen asia tekoälyssä erityisesti tietoturvallisuudesta puhuttaessa. Koneoppiminen perustuu aikaisemmista tapauksista oppiviin algoritmeihin, jotka kykenevät oppimaan toistuvia kaavoja. Syväoppiminen puolestaan perustuu oppimisen mekanismiin. Näitä oppimisen mekanismeja on kolme erilaista: ohjattu oppiminen, ohjaamaton oppiminen sekä vahvistettu oppiminen. Kun tekoälyä käytetään haitallisesti, se lisää hyökkäyksissä onnistumista ja hyökkäyksen ominaisuuksia. Kyberrikolliset ovat oppineet käyttämään tekoälyteknologialla tehostettuja oppimistapoja sekä onnistuneet kääntämään ne omaksi edukseen

Tekoäly tuo mukanaan tietoturvan piiriin haittojen ja uhkien lisäksi myös hyötypuolia. Tekoälyä voidaan myös käyttää tietoturvahyökkäyksiä vastaan. Tekoäly oppii ja pystyy reagoimaan erilaisiin hyökkäyksiin nopeampaa kuin ihminen, minkä johdosta tekoälyn käyttäminen mullistaa kyberrikollisuuden lisäksi myös tietoturvan sekä puolustuksen puolen.

Avainsanat
Tekoäly, tietoturvahyökkäykset, kyberrikollisuus, tekoälyn väärinkäyttö, tekoälyn hyödyntäminen

Ohjaaja
FT, Yliopistonlehtori Elina Annanperä

Sisällysluettelo

Tiivistelmä 2
1. Johdanto 4
1.1 Tutkimuksen rakenne 5
1.2 Motivaatio 5
1.3 Tutkimusmenetelmät 5
2. Tekoälyn käyttö haittaohjelmissa 7
2.1 Tekoälyn käyttäminen hyökkäystyökaluna 8
2.2 Datamyrkytys eli data poisoning 8
3. Tekoälyssä käytetyt teknologiat sekä tekoälyn haitallinen käyttö 10
3.1 Tekoälyn haitallinen käyttäminen 11
4. Tekoälyn käyttäminen tietoturvahyökkäyksiä vastaan 13
5. Pohdinta 16
6. Johtopäätökset 18
Lähteet 20

1. Johdanto

Tekoälyn avulla on jo hetki voitu rakentaa parempaa tietoturvaverkostoa. Tämä kuitenkin on muuttumassa, sillä myös rikolliset ovat myös alkaneet käyttämään tekoälyä pahoihin tarkoituksiin. Tietoturvarikoksien vuotuinen vaikutus maailmantalouteen on noin 400 miljardia dollaria. Rikolliset pystyvät entistä helpommin tekoälyn avulla aiheuttamaan tuhoisia taloudellisia menetyksiä ja voivat vaikuttaa organisaatioihin (Guembe y., 2022). Tekoälyn käyttäminen tietoturvahyökkäyksissä tekee kyberrikollisista entistäkin vaarallisempia, mikä tuo hyvän pohjan asian tutkimiselle. Tutkimusta käsitellään seuraavan tutkimuskysymyksen kautta: Minkälaisiin tietoturvauhkiin tekoälyä käytetään?

Truong ym., (2020) kertovat, että kyberrikolliset käyttävät tekoälyä muun muassa automatisoidakseen mahdollisia hyökkäyksiä sekä tehdäkseen viruksia, jotka ovat tekoälypohjaisia. Tietoturvallisuusuhkat siis kehittyvät ja muuttuvat nopeasti. Vaikka tekoälyä käytetään nykyään myös tietoturvauhkien torjumisessa, se ei vie pois kokonaan kaikkea sitä hyötyä, mitä se tuo rikollisten puolelle (Truong ym., 2020). Tähän uuteen uhkaan on siis tarpeen varautua ja tutkia sitä, kuinka uhkatekijät voivat käyttää tekoälyä haitallisiin tarkoituksiin (Truong Thanh, & Zelinka, 2019).

Tekoälyn käyttäminen tietoturvahyökkäyksiä vastaan on muuttamassa pysyvästi tietoturvan kenttää. Tekoälyn yksi osa eli koneoppiminen on se, joka pystyy jatkuvalla syötöllä lukemaan sekä soveltamaan dataa ja sen myötä parantamaan omia toimintojaan sekä strategioitaan. Tekoälyn avulla voidaan siis torjua hyökkäyksiä ja reagoida niihin nopeammin kuin ihmiset, sillä ihminen ei pysty suodattamaan niin paljoa dataa kerrallaan ja niin nopeasti kuin tekoäly. Tekoäly pystyy siis suodattamaan sekä käsittelemään tämän kaiken datan nopeasti sekä sujuvasti koko ajan kaikkina aikoina vuodesta kellonajasta riippumatta. (Lasic, 2019)

Tutkimuksen tavoitteena on antaa kuvaa siitä, kuinka tekoälyä käytetään tietoturvahyökkäyksissä apuvälineenä sekä kuinka tekoälyn avulla voidaan torjua erilaisia tietoturvahyökkäyksiä ja kuinka tekoälyä hyödyntämällä

tietoturvaa voidaan parantaa. Tutkimuksen perusteella pystytään tunnistamaan tekoälyn kanssa tehtyjä tietoturvahyökkäyksiä ja kuinka tekoälyn käyttämisellä on myös erilaisia hyötyjä. Tutkimuksen perusteella voidaan tunnistaa tekoälyä käyttävät tietoturvauhkat ja niihin pystytään reagoimaan tarpeeksi ajoissa.

1.1 Tutkimuksen rakenne

Luvussa kaksi käydään läpi aiempaa tutkimusta, johon kuuluvat tekoälyn käyttäminen haittaohjelmissa, tekoälyn käyttäminen hyökkäystyökaluna sekä datamyrkytykset. Kolmannessa luvussa käydään läpi tekoälyssä käytettyjä teknologioita sekä tekoälyn haitallista käyttämistä. Neljännessä luvussa käydään läpi siitä, kuinka tekoälyä voidaan käyttää tietoturvahyökkäyksiä vastaan. Viidennestä luvusta löytyy pohdinta. Kuudennesta johtopäätökset.

Viimeisestä kappaleesta löytyy tutkimuksessa käytettyjen lähteiden lähdeluettelo.

1.2 Motivaatio

Tämä aihe on minulle mielenkiintoinen, sillä olen aina ollut kiinnostunut tietoturvasta. Tekoälyn yleistymisen myötä mielenkiintoni on kasvanut, sillä minua kiinnostaa mitä kaikkea tekoäly voi tehdä tai mahdollistaa. Yhdistämällä nämä kaksi mielenkiinnon kohdetta päädyin tähän aiheeseen. Tekoäly on myös lähiaikoina tullut kunnolla esille, kun julkisuuteen ja kaikkien käyttöön julkaistiin ChatGPT eli GPT3. Nämä molemmat tulevat vielä varmasti vielä vastaan tietoturvapiirissä, joko hyvällä tai huonolla tavalla.

1.3 Tutkimusmenetelmät

Tämä tutkielma tehdään semi-systemaattisena kirjallisuuskatsauksena. Tutkielma keskittyy tutkimaan tekoälyn käyttöä tietoturvahyökkäyksissä,

sekä sitä, kuinka tekoälyä käytetään haitallisesti tai väärin. Tekstissä käsitellään myös sitä, kuinka tekoälyä voidaan käyttää tietoturvahyökkäyksiä vastaan. Olen hakenut tiedon pääsääntöisesti seuraavista tietokannoista: Google Scholar, ieee, researchgate ja ProQuest. Rajauksina käytin seuraavia hakusanoja ja niiden yhdistelmiä: ’Artificial intelligence’, ’Ai’, ’Cybersecurity’, ’Attacks’, ’malicious use’, ’malware’, ’AI’, ’offensive’, ‘defense’. Hakusanoina käytin siis seuraavia: (Malicious use) OR (Cybersecurity) AND (Artificial intelligence) OR (AI) AND (Malware) AND (Defensive). Lähteet löysin käyttäen näitä hakusanoja. Näiden rajauksien lisäksi käytin tutkielman tiedonhakuun lähteiden omia lähdeluetteloita.

Tämän tutkielman viittauksissa käytetään APA 7. versiota.

2. Tekoälyn käyttö haittaohjelmissa

Tietokonevirukset saivat alkunsa vuonna 1983, kun Fred Cohen esitteli tietokoneohjelman, joka levisi ja tarttui tietokoneisiin. Tästä lähtien haittaohjelmiin on sisällytetty monia toiminnallisuuksia, kuten salaustekniikoita (Truong Thanh, C. & Zelinka, I. 2019). Tekoälyä on viime vuosina alettu käyttämään entistä enemmän osana tietoturvahyökkäysmenetelmiä (Fritsch ym, 2020). Graafisen käyttöliittymän haittaohjelman tavoitteena on tunnistaa, havaita ja paikantaa uhrin käyttöliittymien kuvakkeet esimerkiksi verkkoselaimesta. Tämän avulla se voi haalia itselleen kaikki uhrin tallentamat kirjautumistiedot. Näin haittaohjelmat voivat hyödyntää internetiä uhrin omaisuuden varastamiseen (Yu ym, 2020).

Uuden sukupolven haittaohjelmat tulevat olemaan entistäkin autonomisempia, kehittyneempiä, nopeampia sekä paljon vaikeammin havaittavia. Tekoälyä käyttävät haittaohjelmat pystyvät jatkossa toimimaan itsenäisesti ja ne ovat tämän myötä paljon älykkäämpiä. Tällaiset haittaohjelmat voivat siis itsenäisesti levitä järjestelmässä omien päätöksien perusteella (Truong ym., 2020). On olemassa järjestelmiä, jotka parantavat haittaohjelmien suorituskykyä. Yksi tällainen järjestelmä on IBM kehittämä DeepLocker. DeepLocker esiteltiin Yhdysvalloissa vuonna 2018. DeepLocker:illa tehostetut haittaohjelmat ovat suuri riski, sillä ne pystyvät tartuttamaan monia eri järjestelmiä ilman, että haittaohjelmaa havaitaan (Blauth ym., 2022).

Haittaohjelmien tekijöiden täytyy omaksua uusi ympäristö ja käyttää hyväksi aiempaa tekoälyn esiintymistä ja siitä hankittua tietoa uusien älykkäiden virusten sekä haittaohjelmien luomiseen. Tämän seurauksena haittaohjelmista tulisi itsenäisiä, ympäristöön integroituvia, kyvykkäitä toimia tietoturvatyökaluja vastaan, sekä ne voisivat hyödyntää aiemmin hankittuja tietoja hyökkäyksessä kohdejärjestelmään (Truong ym., 2020). Perinteisesti haittaohjelmien havaitsemiseen käytettävät menetelmät perustuvat allekirjoituspohjaiseen menetelmään. Tämä menetelmä ei kuitenkaan huomaa tekoälypohjaisia haittaohjelmia, minkä vuoksi ne voivat kiertää nämä menetelmät helposti. (Truong ym., 2020).

2.1 Tekoälyn käyttäminen hyökkäystyökaluna

Niin sanotut syöttöhyökkäykset laukaisevat tekoälyjärjestelmässä toimintahäiriön, kun järjestelmään syötetään väärää dataa: syöttöhyökkäykset eivät siis edellytä sitä, että hyökkääjä olisi valmiiksi jo päässyt saastuttamaan tekoälyjärjestelmän. Syöttöhyökkäykset ovat vaarallisia, sillä niiden hyökkäysmallit eivät välttämättä ole havaittavissa tai ne saattavat olla kokonaan huomaamattomia (Comiter, 2019). Monesti hyökkäyksen huomaa vasta sitten, kun tekoälyalgoritmi antaa väärän vastauksen (Kuzlu ym, 2021). Tämän kaltaisissa hyökkäyksissä hyökkääjä muuttaa vain pienen osan syötteestä, mikä rikkoo järjestelmän aiemmin oppimia malleja. Hyökkäyksiin, jotka on kohdistettu fyysisiin esineisiin, tarvitaan kamera tai jokin anturi, jotta hyökkääjä pystyy syöttämään syötteen tekoälyjärjestelmään. Hyökkäykset, joka tehdään digitaaliseen objektiin eli syötetään suoraan tekoälyjärjestelmään, voivat olla ihmissilmälle huomaamattomia hyökkäysmalleja, sillä muutokset saattavat tapahtua vain pikselitasolla. (Comiter, 2019).

Syöttöhyökkäykset voivat olla mitä tahansa fyysisen stop-merkin teippaamisesta itseajavien autojen hämmentämiseen tai pienen määrän melun lisäämistä (Kuzlu ym, 2021). Syöttöhyökkäyksiä mitataan kahdella eri mittarilla, joista toinen on havaittavuus, joka mittaa sitä, kuinka havaittavissa itse hyökkäys on ihmissilmälle. Toinen on muoto, joka mittaa sitä, kuinka digitaalinen hyökkäys on. (Kuzlu ym, 2021).

2.2 Datamyrkytys eli data poisoning

Datamyrkytys tarkoittaa väärien tietojen injektointia. Tämä hyökkäys on hyvin samankaltainen kuin syöttöhyökkäys, vaikka Syöttöhyökkäyksissä tarkoituksena on muuttaa yksinkertaisesti syöttöä siten, että tekoälyn algoritmi menee sekaisin. Datamyrkytyksen tavoitteena on se, että syötettä muutetaan pitkän ajan kuluessa, jolloin tekoälyn analysointitiedot ovat siirtyneet ja ovat jo luonnostaan puutteellisia. Datamyrkytys tehdään

kyseiseen tekoälyyn, kun tekoäly on päällä ja ennen tekoälyn käyttöönottoa (Kuzlu ym, 2021). Datamyrkytyksiä on kolmea erilaista, jotka ovat tietojoukon myrkytys, algoritmin myrkytyshyökkäykset sekä mallin myrkytys (Kuzlu ym, 2021).

Tietojoukon myrkytyksellä tarkoitetaan sellaista hyökkäystä, jossa tekoälyn tietojoukkoja myrkytetään. Tämä tapa on kaikista näistä kolmesta tavasta suorin tapa toteuttaa hyökkäys. Tekoäly saa kaikki omat tietonsa opetustietosarjojen kautta, joten tällaisessa tapauksessa tietojoukoissa olevat virheet tulevat ajan kanssa heikentämään tekoälyn tietämystä merkittävästi. Tietojoukot ovat monesti todella suuria, joten saastuneiden tietojoukkojen löytäminen voi olla todella vaikeaa (Kuzlu ym, 2021).

Algoritmin myrkytyshyökkäykset ovat hyökkäyksiä, jotka käyttävät hyödykseen tekoälyn omaa oppimisalgoritmia ja sen heikkouksia. Algoritmi, joka on myrkytetty, on siis muihin algoritmeihin yhdistetty ja se tämän myötä se voi myrkyttää lopullisen mallin (Kuzlu ym, 2021). Tämän kaltaisissa hyökkäyksissä yhdistetty oppiminen kerää mahdollisimman paljon arkaluontoista tietoa käyttäjiltä ja muodostaa niistä yhden tietojoukon, minkä kautta kehittää pieniä malleja suoraan hyökkäämiinsä laitteisiin ja yhdistää lopuksi nämä mallit lopulliseen muotoon (Kuzlu ym, 2021). Yhdistetty oppiminen tuo potentiaalisen ratkaisun moniin haastaviin ongelmiin julkisessa politiikassa, sillä käyttäjien tietoja ja yksityisyyttä voidaan sitä käyttäen edelleen hyödyntää ja analysoida varsinaisesti keräämättä näitä tietoja (Comiter, 2019).

Mallin myrkytyksessä hyökkääjät pystyvät korvaamaan toimivan mallin jo valmiiksi myrkytetyllä mallilla. Hyökkääjän täytyy vain päästä järjestelmään käsiksi ja vaihtaa tiedosto (Kuzlu ym, 2021). Tämä tapa on vaarallinen, sillä vaikka malli on valmiiksi koulutettu ja tarkistettu niin ei pysty olemaan varma, että hyökkääjä ei pystyisi muuttamaan mallia sen jakelun eri kohdissa ei ole (Kuzlu ym, 2021).

3. Tekoälyssä käytetyt teknologiat sekä tekoälyn haitallinen käyttö

Tekoälyssä käytetään kahta eri teknologiaa, jotka ovat koneoppiminen ja syväoppiminen. Koneoppiminen on keskeinen keino tekoälyn käyttöön tietoturvallisuudessa. Se perustuu aikaisemmista tapauksista oppiviin algoritmeihin, jotka kykenevät oppimaan toistuvia kaavoja. Sitä voidaan käyttää sekä hyökkäys- että puolustustarkoituksiin. Koska maailma muuttuu koko ajan digitaalisemmaksi, eri tekniikoiden välillä on paljon samankaltaisuuksia. Tässä piilee riski, että kyberrikolliset sabotoivat järjestelmiä omaksi edukseen (Yamin ym., 2021). Koneoppiminen perustuu vahvasti matemaattisiin tekniikoihin (Truong ym., 2020). Koneoppimisen lisäksi käytetään massadataa, joka yhdessä tekoälyn kanssa mahdollistaa puheentunnistuksen, virtuaaliassistentit sekä luonnollisen kielenluomisen, mikä helpottaa ja nopeuttaa ihmisten elämää. Tekoäly on tuonut mukanaan laajamittaiset, autonomiset ja aktiiviset hyökkäykset. Nämä hyökkäykset tuovat uusia haasteita tietoturvan piiriin. Tekoälyn kehittyminen tekee sosiaalisen manipuloinnin hyökkäyksistä yhä älykkäämpiä. (Zeng, 2022).

Syväoppiminen eli DeepLearning. Syväoppimisen luokittelu perustuu oppimisen mekanismiin. Näitä mekanismeja on kolme: ohjattu oppiminen, ohjaamaton oppiminen ja vahvistettu oppiminen (Li, 2018). Ohjattu oppiminen tarvitsee aina selkeästi merkittyä dataa. Ohjattua oppimista käytetään regressiomekanismin luokituksena. Hyvänä esimerkkinä ohjattuun oppimiseen toimii se, että haittaohjelmien tunnistus on binääriskenaarioluokitus eli onko hyvänlaatuinen vai haitallinen eli toisin kuin regressio oppiminen tekee vain ennustearvion (Li, 2018). Ohjaamaton oppiminen eroaa ohjatusta oppimisesta, sillä että tämä oppimistapa on merkitsemätöntä. Tätä oppimistapaa käytetään yleensä tiheyden arvioimiseen, ulottuvuuden vähentämiseen tai tiedon klusterointiin (Li, 2018). Vahvistettu oppiminen perustuu palkitsemiseen, eli se on ohjaamattoman ja ohjatun oppimisen fuusio. Vahvistettu oppiminen siis sopii parhaiten sellaisiin tehtäviin, mitkä ovat pitkäaikaisia (Li, 2018).

3.1 Tekoälyn haitallinen käyttäminen

Kun teknologia kehittyy, se tuo mukanaan myös erilaisia mahdollisuuksia rikollisuuteen. Kun tekoälyä käytetään haitallisesti, se lisää hyökkäyksissä onnistumista ja hyökkäyksen ominaisuuksia. Tekoälyn käyttö siis lisää mahdollisuuksia rikoksen teossa sekä se luo itsestään uuden uhkan. Rikolliset ovat oppineet käyttämään tekoälyteknologialla tehostettuja oppimistapoja sekä onnistuneet kääntämään ne omaksi edukseen. Hyökkääjät voivat esimerkiksi automatisoida hyökkäysprosessejaan (Kaloudi & Jingyue, 2020). Rikolliset käyttävät tekoälyä haitallisesti esimerkiksi sosiaalisen manipuloinnin hyökkäyksissä, joissa on tarkoituksena manipuloida ihmisiä antamaan erilaisia tietoja. Tekoälyn avulla hyökkääjät voivat siis luoda erilaisia manipulointitapoja, joiden avulla he voivat lisätä mahdollisuutta onnistumiseen (Blauth ym., 2022). Tekoälyä käyttämällä hyökkääjä pystyy louhimaan todella isoja määriä dataa sosiaalisesta verkosta ja dataa pystytään käyttämään käyttäjien vaarantamiseen (Truong ym., 2020).

Tekoälyä voidaan käyttää monissa eri tietoturvahyökkäyksissä. Tästä löytyy kuitenkin vain pari esimerkkiä mihin tekoälyä voidaan käyttää ja mitä nämä hyökkäykset ovat. Syväväärennökset ovat kuvia sekä videoita, jotka ovat todella realistisia (Blauth ym., 2022). Syväväärennöksiä tuottavat koneoppimistyökalut ovat tekniikoita, jotka tuottavat uusia tuotteita korvaamalla osia alkuperäisestä tuotteesta. Nämä ovat saaneet myös laajaa mediahuomiota (Peters, 2019). Tekoälyn avulla voidaan luoda henkilö tekemään asioita tai sanomaan jotakin, mitä ei ole oikeasti tapahtunut. Syväväärennökset eivät kuitenkaan ole mikään uusi asia, sillä ennen tekoälyä näitä tehtiin PhotoShopin sekä muiden kuvien ja videoiden muokkausohjelmien avulla. Tekoäly kuitenkin mahdollistaa sen, että kuvista sekä videoista voidaan tehdä niin aidonnäköisiä, että on todella vaikeaa erottaa mikä on totta ja mikä ei. Nykyään videoita ja kuvia tehdään jo valmiiksi tunnetuista henkilöistä. Nämä väärennökset ovat kuitenkin monesti luotuja pahoihin tarkoituksiin, kuten esimerkiksi kiusaamiseen, propagandaan tai kiristykseen. Syväväärennöksillä voi olla vaikutus suoraan kansainvälisiin suhteisiin tai politiikkaan. Syväväärennöksien uhriksi

joutumisen sekä niiden huonoja seurauksia voitaisiin kuitenkin ehkäistä, jos näistä kerrottaisiin ihmisille avoimesti (Blauth ym., 2022).

Kyberrikolliset pystyvät käyttämään tekoälyä botin tekemiseen ja sen kehittämiseen. Tällainen tekoälyllä tehostettu botti tekee rikollisen puolesta mitä he haluavat, eli botti manipuloi kohdetta pyynnöstä. Tällaiset botit ovat algoritmeja, joiden tarkoitus on pystyä jäljittämään ihmisen käyttäytymistä ja nämä botit keräävät sisältöä olemalla tekemisissä ihmisten kanssa internetissä. Hyvänä esimerkkinä toimii ”CyberLover” niminen treffichattibotti. Tämä treffichattibotti julkaistiin vuonna 2007 houkuttelemaan käyttäjiä chat-huoneissa jakamaan käyttäjien henkilökohtaisia tietoja tai painamaan vaarallisia linkkejä. Tämä kyseinen chat-botti käytti NLP:tä eli luonnollisen kielen käsittelyä, jotta se voisi tarjota uhreilleen mukautetun keskustelun. Kun tämä tuli julkisuuteen, esiin nousi huoli siitä, minkälaista teknologiaa rikollisilla on käytössä (Blauth ym., 2022).

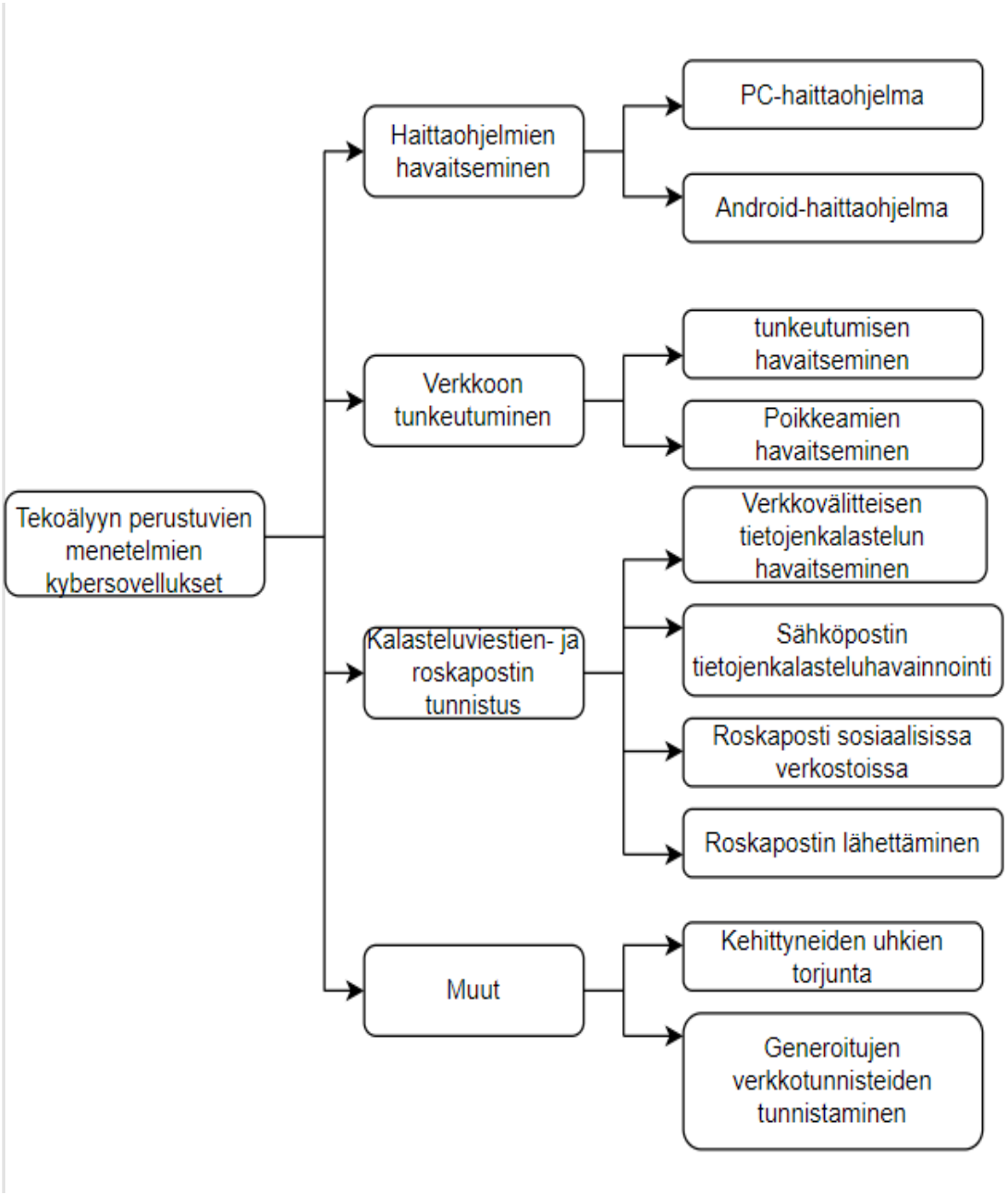
Hyökkäyssuunnitelma on toimintasarja, jossa tietokone tai henkilö suorittaa erilaisia toimia, joilla voisi vahingoittaa kohdettaan. Automaattisesti suunniteltu tietoturvahyökkäys on yksi tekoälyn haara, jota voidaan pitää aikaa vievänä prosessina. Tekoälyä tukeva automatisoitu hyökkäyssuunnittelija on keinotekoinen menetelmä, joka tukee suunnittelun päätös- ja harkintaprosessia. Tämän kaltainen hyökkäyksensuunnittelija tuo mukanaan lisää kattavuutta sekä tarkkuutta arviointiin (Zouave ym., 2020). Kyberrikollisten hyökkäyssuunnitelmat edellyttävät aina kohteen huolellista arviointia, suunnittelua sekä valmistelua (Mathew, 2020). Salasanahyökkäykset voivat olla yksi tekoälyä hyödyntävä tietoturvahyökkäys. Tutkijat ovat testanneet kolmea erityyppistä tekoälypohjaista salasanahyökkäystä, jotka ovat salasanan arvailu, salasanan varastaminen sekä salasanan raakapakottaminen (Zouave ym., 2020).

4. Tekoälyn käyttäminen tietoturvahyökkäyksiä vastaan

Tekoälyn käytöllä on tulevaisuudessa paljon potentiaalia tietoturvan puolustuksessa. Tekoälyn käyttö on suositeltavaa, sillä mitä enemmän tietoa tekoäly saa, sitä paremmin se tulevaisuudessa pystyy torjumaan tietoturvahyökkäyksiä. Tekoälyn käyttäminen on esimerkiksi siitä hyvä, että se oppii koko ajan ja pystyy tämän kautta torjumaan reaaliajassa vihollisen hyökkäyksiä esimerkiksi luomalla suojausta käyttäen aiemmin oppimaansa dataa. Tekoälyn käyttäminen on siitä suositeltavaa, koska se voi koko ajan luoda uusia tapoja, miten suojata järjestelmää. (Talwar & Koury, 2017).

Tietenkin tärkeimpänä kysymyksenä toimii se, että viekö tekoälyn käyttäminen työt kaikilta tietoturva-asiantuntijoilta, mutta näin ei kuitenkaan ole ainakaan vielä. Tekoäly ei kuitenkaan vielä osaa kaikkea ja siksi se tarvitsee ihmisiä opettamaan itseään. Jossakin vaiheessa tulee todennäköisesti vaihe, jossa ihmisiä ei enää tarvita opettamaan tekoälyä. (Talwar & Koury, 2017).

Tekoälyä voidaan siis käyttää myös hyvään tarkoitukseen, kuten erilaisten tietoturvauhkien estämiseen. Tutkijat ovat löytäneet monia erilaisia tekniikoita, jotka käyttävät tekoälyä erilaisten haittaohjelmien löytämiseen sekä niiden luokitteluun. Tekoälyn avulla pystytään myös havaitsemaan erilaisia verkkotunkeutumisia, roskapostihyökkäyksiä sekä tietojenkalastelua (Truong ym., 2020). Nämä tekoälyä käyttävät tietoturvasovellukset näytetään kuvassa 1.



Kuva 1. Tekoälyä käyttävien tietoturvasovellusten eri päähaarat.

4.1 Tekoälyn käyttämisen hyödyt sekä haitat

Tekoälyn käyttäminen tietoturvallisuuden puolustamisen osalta on kannattavaa sekä suositeltavaa. Siihen on monia erilaisia syitä: Tekoäly pystyy parantamaan tietoturvallisuutta, sillä tekoäly kykenee havaitsemaan pienetkin tietoturvahyökkäykset, sekä tämän myötä sen avulla voidaan parantaa tietoturvaa ja nopeuttaa tietoturvahyökkäyksiin reagointia. Puolustustarkoitetukseen käytettävästä tekoälystä voidaan myös hyötyä sillä,

että tekoälyn käytöllä voidaan parantaa sekä nopeuttaa havaitsemisaikaa, sillä tekoälyllä on kykyä huomata sekä mitata erilaisia riskejä. Kun tekoälyä käytetään hyvään eli puolustamistarkoitukseen se vähentää myös erilaisten organisaatioiden kustannuksia. Esimerkiksi nopean reagoimisen myötä tietoturvahyökkäyksien vaikutus voidaan pienentää. Tekoälyn käytöllä voidaan myös vähentää sellaisia virheitä, joita ihmisen tekemien manuaalisten prosessien käyttäminen tuo mukanaan (Murugesan, 2022).

Tekoälyä käyttävät puolustusjärjestelmät voivat käyttää myös vähemmän aikaa hälytyksien sekä vääränlaisien positiivisten tuloksien etsimiseen, jonka seurauksena tietoturvan parissa työskentelevät ihmiset vapautuvat näistä tehtävistä ja pystyvät keskittymään enemmän uhkiin, jotka ovat sillä hetkellä kriittisiä. Organisaatioiden tietoturvaosastojen työtyytyväisyyttä voidaan myös parantaa tekoälyn avulla, sillä päivittäiset aikaa vievät toimenpiteet on siirretty tekoälylle, jonka myötä tietoturvan ammattilaiset voivat keskittyä isomman prioriteetin tehtäviin. Viimeisenä, mutta ei kuitenkaan vähäisempänä, organisaation asiakastyytyväisyys nousee, sillä tietoturvallisuuden luottamus kasvaa, koska tekoäly hoitaa tehtävät, jotka ovat ennen täytyneet tehdä manuaalisesti. Kuten aiemmin mainittua, organisaation reaktioaika nopeutuu mahdollisia hyökkäyksiä vastaan, minkä ansiosta tietoturvahyökkäykset ovat mahdollisesti lievempiä tai ne saadaan suoraan estettyä (Murugesan, 2022).

Tekoäly säästää rahaa sekä aikaa organisaatioilta. Tekoälyn käyttäminen tietoturvapuolustuksessa mahdollistaa nopean reagoinnin tietoturvahyökkäyksiä vastaan. Tämä on tärkeää, sillä nopea reagointi on välttämätöntä organisaation turvallisuuden kannalta. Suurimmassa osassa organisaatioita, jotka ovat ottaneet tekoälyn käyttöönsä, tekoälyn avulla voidaan vähentää tietoturvauhkien reagointiaikaa 12 prosentilla. Joissakin organisaatioissa voidaan vähentää reagointiaikaa jopa 15 prosentilla (Lasic, 2019).

5. Pohdinta

Tavoitteena oli tehdä kirjallisuuskatsaus siitä, kuinka tekoälyä voidaan hyödyntää tietoturvahyökkäyksissä, eli antaa yleiskuvaa siitä, miten tekoälyä voidaan käyttää hyödyksi erilaisissa tietoturvahyökkäyksissä sekä kuinka tekoälyä voidaan käyttää tietoturvahyökkäyksiä vastaan. Kirjallisuuskatsauksessa käytiin läpi joitakin osa- alueita, mitä tekoälyn käyttämisestä tietoturvahyökkäyksissä on. Kuten esimerkiksi sitä, kuinka tekoälyä käytetään väärin haittaohjelmissa, sosiaalisen manipuloinnin hyökkäyksissä ja/tai esimerkiksi salasanahyökkäyksissä.

Tämä kirjallisuuskatsaus tuo hyvin esille sen, mitä tekoälyn käyttäminen on tietoturvahyökkäyksissä, millaisia erilaisia hyökkäysmenetelmiä on, joissa käytetään apuna tekoälyä ja mitä erilaisia teknologioita ylipäätään tekoäly käyttää voidakseen toimia tietoturvahyökkäyksissä. Tästä aiheesta ei kovin paljoa ainakaan ole tutkimusta suomenkielisenä.

Aiheena tekoälyn käyttäminen tietoturvahyökkäyksissä ja niiden puolustamisessa on suhteellisen uusi, joten tämän aiheen tutkimus on vasta alussa. Tulevaisuus tällä kyseisellä aiheella on valoisa ja tutkittavaa riittää tulevaisuudessa. Vaikka tekoäly sekä tietoturva ovat olleet olemassa jo pitkän aikaa, tekoälyn käyttäminen tietoturvahyökkäyksissä ja niiden puolustamisessa on suhteellisen uutta. Tätä aihetta tullaan varmasti vielä tulevaisuudessa käsittelemään eri näkökulmista, koska kuitenkin aiheena tämä on suhteellisen ajankohtainen. Uskon että tämä aihe pysyy pitkään kuitenkin ajankohtaisena, sillä kuten Guembe (2022) kertoo, tietoturvarikoksien vaikutus maailmantalouteen on noin 400 miljardia dollaria vuodessa. Ongelmana pitkälti tämän kirjallisuuskatsauksen teossa oli, se kuinka löytää lähteitä tästä aiheesta, sillä tutkimusta ei tosiaan kauhean paljoa löydy. Kirjallisuushaun aikana huomasin, että suurin osa tutkimuksista on tehty lähempänä vuotta 2020 kuin sitä aikaisemmin, mikä kertoo siitä, että tutkimukset aiheesta ovat suhteellisen uusia. Tekoälyn käyttäminen erilaisissa tietoturvahyökkäyksissä itsestään tuo jo monia erilaisia ongelmia tietoturvan piiriin. Tekoäly tuo erilaisiin tietoturvahyökkäyksiin erilaisia vaaroja: esimerkiksi tekoälyn avulla voidaan vaikeuttaa tietoturvahyökkäyksien havaitsemista. Tietoturvan puolustukseen

käytettävässä tekoälyssä on monia hyötyjä kuten esimerkiksi se, että tekoälyn käyttäminen pienentää huomattavasti reagointi aikaa tietoturvahyökkäyksien havaitsemiseen sekä torjumiseen. Lasicin (2019) mukaan ne organisaatiot, jotka ovat ottaneet käyttöönsä tekoälyn, ovat parantaneet reaktioaikaansa tietoturvahyökkäyksiin noin 12 prosentilla. Vaikka tekoälyä käytetään pahoihin tarkoituksiin eli tietoturvahyökkäyksiin, niin sen käyttäminen puolustustarkoituksiin tuo samalla tasapainoa tietoturvan piiriin.

6. Johtopäätökset

Tekoälyn käyttäminen haittaohjelmissa sekä muissa tietoturvahyökkäyksissä on yleistymässä, sillä tekoälyn avulla voidaan niin helposti automatisoida työtä ja myös itse kyberrikollisen kiinnijäämisen riski pienenee. Tekoälyä käytetään paljon erilaisissa haittaohjelmissa, salasanahyökkäyksissä, sekä hyökkäyssuunnitelman teossa. Kyberrikolliset saattavat monesti myös käyttää itse tekoälyä hyökkäystyökaluna. Hyökkääjät, jotka käyttävät tekoälyä itse hyökkäystyökaluna, hyökkäävät ensin tekoälyä vastaan esimerkiksi syöttöhyökkäyksellä tai myrkyttävät sen dataa.

Tekoälyn käyttäminen tietoturvahyökkäyksissä on selvästi kasvamassa koko ajan, joten se tuo sellaista uhkaa tietoturvalle, jota ennen ei ole ollut olemassa. Tämän takia tietoturvan puolustuksen laatimisessa täytyy käyttää myös tekoälyä, jotta puolustuksesta saataisiin sopiva vastus hyökkäyksille. Tekoälyn käyttäminen erilaisissa hyökkäyksissä tekee hyökkäyksistä vaikeammin huomattavia sekä tekee niistä paljon vaarallisempia. Tekoälyn väärinkäyttämisen takia kyberrikollisista on tulossa koko ajan vaarallisempia.

Tekoälyä käyttäminen tietoturvauhkien puolustuksessa on myös yleistymässä. Tekoälyn avulla voidaan lyhentää tietoturvahyökkäyksiin reagoimisen nopeutta huomattavasti verrattuna ihmisen reagointiin. Tekoälyä käyttämällä tietoturvan ammattilaiset voivat keskittyä enemmän kriittisiin uhkiin sen sijaan, kun he kävisivät päivittäin läpi pieniä mahdollisia uhkia, Tämän asian tekoäly voi esimerkiksi hoitaa heidän puolestaan. Tekoälyn käyttäminen tuo siis monia erilaisia hyviä puolia tietoturvan piiriin.

Tutkielma rajautuu tietoturvaan, jossa käytetään apuna tekoälyä. Pääosin tutkielmassa on tietoa tekoälyn avulla tehdyissä tietoturvahyökkäyksissä, mutta myös hieman siitä, kuinka tekoälyn avulla voidaan tehdä tietoturvasta turvallisempaa ja kuinka sitä voidaan käyttää myös hyviin tarkoituksiin. Tutkielman aineisto rajautuu vain tiettyihin tietoturvauhkiin sekä myös tiettyihin hyötyihin tekoälyn käyttämisessä tietoturva puolustuksessa, sillä tutkielmassa ei voida käydä kaikkia mahdollisuuksia läpi.

Työ ei myöskään vastaa yleisiin tietoturvauhkiin, joissa ei käytetä tekoälyä. Seuraavaksi tulisi jatkaa tutkimusta, sitä kuinka tekoälyä käyttäviltä tietoturvauhkilta voitaisiin suojautua käyttäen apuna tekoälyä. Eli toisin sanoen kuinka tekoälyä voidaan käyttää tekoälyä vastaan. Seuraavaksi voisi olla myös hyvä tutkia lisää tätä samaa aihetta yleisesti, eli kuinka tekoälyä hyödynnetään tietoturvahyökkäyksissä tai sitten yksittäisesti tutkien vain jotakin tiettyä tekoälyä käyttävää tietoturvahyökkäystä. Jatkotutkimusta tästä aiheesta tulisi tehdä kuitenkin muulla tavalla kuin kirjallisuuskatsauksena, eli jatkotutkimus tulisi tehdä mielestäni tutkimuksena, jossa etsittäisiin uutta tutkimusmateriaalia.

Lähteet

Blauth, T. F., Gstrein, O. J., & Zwitter, A. (2022). *IEEE Xplore Full-Text PDF:* https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9831441

Comiter, M. (2019). *Attacking Artificial Intelligence AI's Security Vulnerability and What Policymakers Can Do About It*. www.belfercenter.org

Fritsch, L., Jaber, A., & Yazidi, A. (2022). *An Overview of Artificial Intelligence Used in Malware*. 41–51. https://doi.org/10.1007/978-3-031-17030-0_4

Guembe, B., Azeta, A., Misra, S., Osamor, V. C., Fernandez-Sanz, L., & Pospelova, V. (2022). The Emerging Threat of Ai-driven Cyber Attacks: A Review. *36*(1). https://doi.org/10.1080/08839514.2022.2037254

Kaloudi, N., & Jingyue, L. I. (2020). The AI-based cyber threat landscape: A survey. *ACM Computing Surveys*, *53*(1). https://doi.org/10.1145/3372823

Kuzlu, M., Fair, · Corinne, & Guler, · Ozgur. (2021). Role of Artificial Intelligence in the Internet of Things (IoT) cybersecurity. *Discover Internet of Things 2021 1:1*, *1*(1), 1–14. https://doi.org/10.1007/S43926-020-00001-4

Lazic, L. (2019). Benefit from AI in Cybersecurity. The 11th International Conference on Business Information Security (BISEC-2019), 18th October 2019, Belgrade, Serbia, 2019. https://www.researchgate.net/publication/336826190_BENEFIT_FROM_AI_IN_CYBERSECURITY

Li, J. Hua. (2018). Cyber security meets artificial intelligence: A survey. *Frontiers of* Information *Technology and Electronic Engineering*, *19*(12),

1462–1474. https://doi.org/10.1631/FITEE.1800573

Mathew, A. (2021). *Artificial Intelligence for Offence and Defense—The Future of Cybersecurity | Request PDF*.

https://www.researchgate.net/publication/351607098_Artificial_Intelligence_for_Offence_and_Defense_-_The_Future_of_Cybersecurity

Murugesan, S. (2022). The AI-Cybersecurity Nexus: The Good and the Evil. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9967400

Peters, K. (2019). 21st Century Crime: How Malicious Artificial Intelligence Will Impact Homeland Security. *Homeland Security Affairs,*

https://www.proquest.com/scholarly-journals/21st-century-crime-how-malicious-artificial/docview/2266265939/se-2artificial/docview/2266265939/se-2

Talwar, K., & Koury, A. (2017). Artificial Intelligence – the Next Frontier in IT Security? *Network Security, 2017(4), 14-17.* DOI:10.1016/S1353-4858(17)30039-9

Truong, T. C., Diep, Q. B., & Zelinka, I. (2020). Artificial Intelligence in the Cyber Domain: Offense and Defense. *Symmetry 2020, Vol. 12, Page 410*, *12*(3), 410–410. https://doi.org/10.3390/SYM12030410

Truong Thanh, C., & Zelinka, I. (2019). *View of A Survey on Artificial Intelligence in Malware as NextGeneration Threats*. https://mendel-journal.org/index.php/mendel/article/view/105/125journal.org/index.php/mendel/article/view/105/125

Truong, T. C., Zelinka, I., Plucar, J., Čandík, M,. Šulc, V. (2020). *Artificial Intelligence and Evolutionary Computations in Engineering Systems*. *668*.

https://doi.org/10.1007/978-981-10-7868-2

Yamin, M. M., Ullah, M., Ullah, H., & Katt, B. (2021). Weaponized AI for cyber-attacks. *Journal of Information Security and Applications*, 57, 102722– 102722.

https://doi.org/10.1016/J.JISA.2020.102722

Yu, N., Tuttle, Z., Jake Thurnau, C., & Mireku, E. (2020). *AI-Powered GUI Attack and Its Defensive Methods*. https://doi.org/10.1145/3374135.3385270

Zeng, Y. (2022). AI Empowers Security Threats and Strategies for Cyber Attacks. *Procedia Computer Science*, *208*, 170–175.

https://doi.org/10.1016/J.PROCS.2022.10.025

Zouave, E., Bruce, M., Colde, K., Jaitner, M., Rodhe, I., & Gustafsson, T. (2020). (Artificially intelligent cyberattacks. *Swedish Defence Research Agency, FOI, Tech. Rep. FOI.*)