Eetu Malvela

TEKOÄLYPOHJAISTEN UHKIEN VAIKUTUS TIETOJÄRJESTELMIIN JA NIIDEN TORJUNTA

JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2025

TIIVISTELMÄ

Malvela, Eetu
Tekoälypohjaisten uhkien vaikutus tietojärjestelmiin ja niiden torjunta
Jyväskylä: Jyväskylän yliopisto, 2025, 35 s.
Tietojärjestelmätiede, kandidaatin tutkielma
Ohjaaja(t): Kokko, Tuomas

Tekoälyä integroitaessa kyberturvallisuuteen tuodaan esiin sekä ainutlaatuisia mahdollisuuksia että merkittäviä haasteita tietojärjestelmien suojaamisessa. Tässä tutkielmassa tarkastellaan tekoälypohjaisten uhkien vaikutusta tietojärjestelmien kyberturvallisuuteen sekä torjuntakeinoja näiden riskien lieventämiseksi.

Kirjallisuuskatsauksen avulla analysoidaan tekoälyn kaksoisroolia: kyberturvallisuuden vahvistamisen työkaluna ja toisaalta kehittyneempien kyberhyökkäysten välineenä. Tekoälyohjattujen kyberhyökkäysten havaitaan hyödyntävän tietojärjestelmien eheyttä, saatavuutta ja luottamuksellisuutta. Tekoälyllä varustettujen hyökkäys- ja väistötekniikoiden, generatiivisten vastakkaisverkostojen (GAN) ja älykkäiden haittaohjelmien käyttö tunnistetaan yleistyväksi. Näitä käytetään muun muassa disinformaation luomisessa ja koneoppimismallien harhauttamisessa ja tekoälypohjaisissa kyberhyökkäyksissä. Lisäksi vastakkaisten hyökkäysten havaitaan johtavan taloudellisiin tappioihin, operatiivisiin viivästyksiin ja luottamuksen heikkenemiseen tekoälyteknologioita kohtaan. Tutkimustuloksissa korostuu tekoälyn ja ihmisten yhteistyötä käyttävien, sopeutuvien ja ennakoivien puolustusmekanismien, kuten kehittyneiden poikkeavuustunnistusjärjestelmien ja tekoälytehostetun uhkatiedustelun kehittäminen. Näiden mekanismien kiireellisyyttä painotetaan, sillä niillä pyritään vastaamaan tekoälypohjaisten uhkien monimutkaisuuteen. Lisäksi tekoälypohjaisten uhkien vaikutusta voidaan minimoida sitkeiden tietojärjestelmien (RIS), sekä selitettävän tekoälyn (XAI) käytöllä.

Tekoälyn vaikutuksen tarkastelu uhkakuviin ja puolustusmalleihin antaa ymmärrystä innovatiivisten ratkaisujen kehittämiseksi tietojärjestelmien suojaamiseksi yhä tekoälypainotteisemmassa ympäristössä.

Asiasanat: tekoäly, AI-pohjaiset kyberhyökkäykset, tietojärjestelmien kyberturvallisuus

ABSTRACT

Malvela, Eetu
The Impact of AI-Based Threats on Information Systems and Their Mitigation
Jyväskylä: University of Jyväskylä, 2025, 35 pp.
Information systems science, bachelor's thesis
Supervisor(s): Kokko, Tuomas

The integration of artificial intelligence (AI) into cybersecurity is seen to present both significant opportunities and notable challenges for protecting information systems. This thesis investigates how AI-based threats impact the cybersecurity of information systems and examines countermeasures to mitigate these risks.

A comprehensive literature review is employed to explore AI's dual role in cybersecurity—both as a tool to enhance defensive measures and as a vector for advanced cyberattacks. It is highlighted that vulnerabilities in information systems are exploited through AI-driven cyberattacks, undermining their integrity, availability, and confidentiality. The prevalence of intelligent malware, AI-powered evasion- and attack -techniques, and generative adversarial networks (GANs) is identified. These technologies are utilized for creating disinformation and deceiving machine learning models. Furthermore, adversarial attacks, where AI-generated malicious inputs disrupt machine learning algorithms, are shown to lead to economic losses, operational delays, and diminished trust in AI technologies. The findings highlight the development of adaptive and proactive defense mechanisms that integrate human and AI collaboration, such as advanced anomaly detection systems and AI-enhanced threat intelligence.

The urgency of these mechanisms is emphasized to address the complexity of AI-driven threats effectively. Additionally, the impact of AI-based threats can be mitigated through the implementation of resilient information systems (RIS) and explainable AI (XAI), ensuring a more robust and transparent approach to cybersecurity challenges.

By examining the transformative impact of AI on both threat scenarios and defensive frameworks, the research contributes to the development of innovative solutions for protecting information systems in an AI-dominated environment.

Keywords: artificial intelligence, AI-driven cyberattacks, information systems cybersecurity

KUVIOT

SISÄLLYS

1 JOHDANTO

Tekoäly (AI) on noussut yhdeksi merkittävimmistä teknologioista, joka muuttaa toimintatapoja niin organisaatioissa kuin yhteiskunnassa laajemmin. Tekoäly näyttäytyy älykkäänä: tekoälyn tekninen konsepti on havainnoida ja toimia (Russell, 2020). Tekoälyä hyödynnetään monilla aloilla, kuten terveydenhuollossa diagnostiikan tukena, teollisuudessa tuotannon optimoinnissa sekä kyberturvallisuudessa uhkien tunnistamisessa ja torjumisessa (Liu & Chen, 2024). Teknologian nopea kehitys on kuitenkin tuonut mukanaan myös uudenlaisia riskejä, erityisesti kyberuhkien muodossa. Tekoäly mahdollistaa entistä kehittyneempiä ja sopeutuvampia kyberhyökkäyksiä, jotka uhkaavat tietojärjestelmiä ja organisaatioiden toiminnan jatkuvuutta (Jeong, 2020).

Tekoälyn vaikutus ulottuu syvälle yhteiskuntaan, aina yksilöiden arkipäivästä kyberturvallisuuteen ja kansainväliseen kauppaan (Murugesan, 2022). Esimerkiksi tekoälypohjaisten kielimallien, kuten ChatGPT:n, avulla ihmiset voivat helposti suorittaa monimutkaisia tehtäviä, kuten ohjelmointia tai tietojen analysointia (Alawida ym., 2024). Samalla tekoälyn kyky käsitellä ja analysoida suuria tietomääriä avaa uusia mahdollisuuksia tietojärjestelmien hallinnassa. Tekoäly on kuitenkin kaksikäyttöinen teknologia, eli sitä voidaan käyttää hyviin sekä pahoihin käyttötarkoituksiin (Jeong, 2020).

Tietojärjestelmät ovat keskeisessä roolissa nyky-yhteiskunnassa. Esimerkiksi ne tukevat turvallisia tunnistautumispalveluja, terveydenhuollon järjestelmiä ja liiketoiminnan toimintoja (Murugesan, 2022). Tekoälypohjaiset uhat kuitenkin haastavat näiden järjestelmien turvallisuuden ja luotettavuuden (Osman & El-Gendy, 2024). Esimerkiksi tekoälyn avulla voidaan toteuttaa automaattisia kyberhyökkäyksiä, jotka pystyvät mukautumaan puolustusmekanismeihin ja kiertämään perinteiset turvajärjestelmät.

Tämän tutkielman tavoitteena on selvittää, miten tekoälypohjaiset uhat vaikuttavat tietojärjestelmiin, ja kuinka niitä voidaan torjua. Tutkielmaa varten on tehty kirjallisuuskatsaus tekoälyyn liittyvien hyötyjen, haasteiden ja riskien kartoittamiseen. Tutkielman toisena tavoitteena on tuottaa kattava kuva tekoälyn roolista sekä hyökkäyksissä että puolustuksessa, ja tarjota käytännön näkökulmia tekoälyn hyödyntämiseen kyberturvallisuuden parantamisessa.

Tutkimuskysymys on: Miten tekoälypohjaiset uhat vaikuttavat tietojärjestelmiin, ja kuinka niitä voidaan torjua? Kysymys on ajankohtainen, sillä tekoälypohjaiset uhkat ovat jatkuvassa kasvussa, mikä vaatii uusia ja innovatiivisia puolustusstrategioita.

1.1 Tutkimusmenetelmä ja lähestymistapa

Tässä tutkielmassa käsitellään ensin pääkäsitteiden teoreettista viitekehystä ja tekoälyn roolia kyberturvallisuudessa. Sen jälkeen esitellään tutkimusmenetelmät ja aineiston analyysitavat. Luvuissa kaksi ja kolme aihetta avataan ja luodaan pohjaa seuraavia lukuja varten. Tuloksia tarkastellaan luvuissa, jotka keskittyvät tekoälypohjaisten uhkien vaikutuksiin sekä niiden torjuntaan: luvut neljä ja viisi. Lopuksi tehdään yhteenveto ja pohditaan jatkotutkimusaiheita. Tutkielmassa keskityttiin tekoälyn rooliin kyberhyökkäysten välineenä aiheen rajauksena. Ihmisnäkökulman tarkastelu kyberturvallisuusriskinä jätettiin tämän takia vain yleiselle tasolle.

Laadullinen tutkimusmenetelmä valittiin, koska se mahdollistaa syvällisen ymmärryksen tekoälyyn liittyvien kyberuhkien luonteesta ja niiden vaikutuksista tietojärjestelmiin. Systemaattisen kirjallisuuskatsauksen avulla tutkielmassa analysoidaan laaja-alaisesti aiempia tutkimuksia aiheesta ja tuotetaan synteesi olemassa olevasta tiedosta.

Tiedonhaku toteutettiin hyödyntämällä akateemista tietokantaa Web of Science ja Jyväskylän Yliopiston kirjastoa JYKDOK. Hakuparametreinä käytin tekoälyä tietojärjestelmien, kyberhyökkäysten ja kyberhyökkäyksiltä suojautumisen konteksteissa. Lähteistä osa kuitenkin on muihin aihepiireihin kohdistuvaa. Muiden aihepiirien, kuten tietojärjestelmiin ja kyberturvallisuuteen liittyviä lähteitä käytettiin täydentämään aukkoja, johtuen siitä, että tutkimuksia tekoälypohjaisten uhkien vaikutuksesta tietojärjestelmiin on hyvin vähän. Lähteiden valintakriteereihin kuuluivat niiden relevanssi, akateeminen luotettavuus ja ajankohtaisuus. Lähteitä oli helppo löytää tekoälyyn liittyen, mutta on huomattava, että tutkimustietoa tekoälypohjaisista kyberuhista tarvitaan paljon lisää.

1.2 Keskeiset käsitteet

Tässä alaluvussa käyn läpi tutkielman keskeisimpiä käsitteitä, jotka lukijan on hyvä ymmärtää. Nämä käsitteet toistuvat kirjallisuuskatsauksessa useita kertoja ja niitä käytetään useissa tutkielman luvuissa. Käsitteet ovat olennaisia ymmärtää tutkielman kannalta, ja ne selkeyttävät tekstiä.

Tekoäly (AI): Viittaa tietokoneisiin tai tietokoneohjelmiin, jotka pyrkivät jäljittelemään ihmisen älyllistä toimintaa, kuten päätöksentekoa, oppimista ja ongelmanratkaisua (Gignac & Szodorai, 2024). Tekoälyn osa-alueisiin kuuluvat

esimerkiksi koneoppiminen (ML) ja syväoppiminen (DL). Tekoäly on kaksikäyttöinen teknologia, eli sitä voidaan käyttää niin hyviin kuin pahoihinkin tarkoitusperiin (Jeong, 2020). Tekoälyn käyttö kyberturvallisuuden parantamisessa on esimerkki tekoälyn hyvästä tarkoitusperästä, ja vastaavasti käyttö kyberhyökkäyksissä edustaa esimerkkiä tekoälyn valjastamisesta pahaan tarkoitukseen (Jeong, 2020).

Koneoppiminen (ML) ja Syväoppiminen (DL): Koneoppiminen on tekoälyn osa-alue, jossa algoritmit oppivat analysoimaan dataa, tunnistamaan kaavoja ja tekemään päätöksiä ilman ennalta määriteltyjä sääntöjä (Jakhar & Kaur, 2020). Syväoppiminen puolestaan on koneoppimisen alalaji, joka hyödyntää monikerroksisia keinotekoisia neuroverkkoja. Syväoppiminen on erityisen tehokas monimutkaisissa datan analyysitehtävissä, kuten kuvien tai puheen käsittelyssä, mutta se vaatii enemmän laskentatehoa ja suurempia tietomääriä kuin perinteiset koneoppimisen menetelmät (Jakhar & Kaur, 2020).

Tekoälypohjaiset kyberuhat: Hyödyntävät tekoälyn ja koneoppimisen algoritmeja, mikä mahdollistaa monivaiheisten ja sopeutuvien hyökkäysten toteuttamisen hyödyntäen suurta volyymiä ja mukautuvuutta (Mirky ym., 2022). Nämä uhat edustavat uudenlaisia haastetta tietojärjestelmille ja perinteisille kyberturvallisuuden strategioille (Mirky ym., 2022).

Kyberturvallisuus: On tieto- ja viestintäjärjestelmien, verkkojen ja datan suojaamista luvattomalta pääsyltä, väärinkäytöltä ja vahingoittamiselta. Se kattaa sekä tekniset että organisatoriset toimenpiteet, joiden tavoitteena on estää, havaita ja reagoida tietoturvauhkiin (Kaur, Gabrijelcic & Klobucar, 2023). Kyberturvallisuus suojaa yksityisyyttä, omaisuutta ja yhteiskunnan kriittisiä toimintoja digitaalisten uhkien varalta. Sen merkitys korostuu jatkuvasti teknologian kehityksen ja kyberuhkien monimuotoistumisen myötä (Kaur, Gabrijelcic & Klobucar, 2023).

Tietojärjestelmä: Ovat järjestelmiä, jotka koostuvat ihmisistä, teknologioista, prosesseista ja tiedosta (Palko ym., 2023). Niiden tarkoituksena on kerätä, tallentaa, käsitellä ja jakaa tietoa organisaation tai muun toiminnan tehostamiseksi. Tietojärjestelmät sisältävät ohjelmistoja, laitteistoja, tietokantoja ja viestintäverkkoja (Palko ym., 2023).

Sitkeät tietojärjestelmät: Ovat kehittyneitä tietojärjestelmiä, jotka voivat käyttää tekoälyä tietojärjestelmän hallinnoimiseen. Tällä saadaan aikaan parempi kyky kestää kyberhyökkäyksiä, sillä sitkeä tietojärjestelmä kestää hyökkäysten paineen paremmin sulkemalla haavoittuvat- ja priorisoimalla kriittiset osat (Gupta, Mogdil, Meissonier & Dwivedi, 2024).

2 TEKOÄLY

Tässä luvussa tarkastellaan tekoälyn kehittymistä, sen käyttöalueita sekä siihen liittyviä keskeisiä käsitteitä ja haasteita. Aluksi käsitellään tekoälyn määritelmää ja sen sovellusmahdollisuuksia eri aloilla, kuten kyberturvallisuudessa. Sen jälkeen esitellään tekoälyn kaksikäyttöisyyden vaikutuksia ja tekoälyteknologioiden, kuten koneoppimisen ja syväoppimisen, eroja ja vaatimuksia. Lopuksi keskitytään tekoälyyn liittyviin eettisiin ja teknisiin haasteisiin, kuten niin sanottujen mustan laatikon mallien selitettävyysongelmaan, sekä esitellään ratkaisuja, kuten ymmärrettävä ja selitettävä tekoäly. Näiden näkökulmien avulla luodaan pohjaa tekoälyn mahdollisuuksien ja riskien ymmärtämiselle erityisesti kyberturvallisuuden kontekstissa.

Tekoäly (AI) on kehittynyt viime vuosina harppauksin ja sitä otetaan käyttöön yhä useammassa eri käyttöympäristössä. Russelin (2021) mukaan tekoälystä on tullut aihe suurvaltojen välisen kilpailun keskiössä, ja monet asiantuntijat pitävät sitä tulevan taloudellisen nousun teknologisena perustana. Liu ja Chen, (2024) toteavat tutkimuksessaan, että tekoäly on monitieteinen aihe, joka on noussut 2000-luvulta alkaen, ja sen tutkimus kattaa laajan määrän eri aihealueita. Tekoälypohjaiset-sovellukset siirtyvät vähitellen tutkimuslaboratorioiden turvallisten seinien ulkopuolelle ja tulevat osaksi arkea (Schramm, Wehner & Schmid, 2023). Russelin (2021) mukaan älykäs tai rationaalinen toiminta on sellaista, jonka voidaan odottaa saavuttavan asetetut tavoitteet. Tekoäly on tietokonesimulaatio ihmisenkaltaisesta älyllisestä prosessista, jota voidaan käyttää monenlaisissa toiminnoissa (Gignac & Szodorai, 2024). Tekoälyn käyttöön kuuluvat esimerkiksi kuvantunnistus, puheteknologia, kielimallit, teolliset prosessit, robotiikka, analytiikka, terveydenhuolto, talous ja kyberturvallisuus (Jeong, 2020).

2.1 Tekoälymenetelmät, haasteet ja ratkaisut

Gignacin ja Szodorain (2024) mukaan tekoälyn suurimpana valttina pidetään sen nopeutta ja kykyä käsitellä suurta datamäärää verrattuna ihmiseen. Tutkimuksessa kuitenkin todetaan, että koska tietojärjestelmien nopeus on usein keskeinen

ominaisuus, voidaan spekuloida, että tekoälyjärjestelmien nopeuden vaihtelu saattaa vaikuttaa tekoälyn älykkyyden tiedostamiseen (Gignac ja Szodorai, 2024).

On myös hyvä ottaa huomioon tekoälyn kaksikäyttöisyys. Alawida ym., (2024) kertovat tutkimuksessaan, että tekoäly voi helpottaa kyberrikollisten toimintaa, mutta sillä on myös potentiaalia vahvistaa kyberturvallisuustoimia. Tekoälyn kouluttaminen on usein pitkä ja monimutkainen prosessi, joka vaatii aikaa, suurta laskentatehoa ja datamäärää (Alawida ym., 2024). Tekoälyn laskennalliset algoritmit voivat vaihdella yksinkertaisista sääntöpohjaisista ohjeista monimutkaisiin prosesseihin, kuten koneoppimisen ja syväoppimisen menetelmiin (Gignac & Szodorai, 2024). Tutkimuksessaan Russell (2021) kertoo että, 1980-luvulla tekoälytutkijat alkoivat käsitellä todellisiin havaintoihin ja ihmisiltä tai koneoppimisen kautta hankittuun tietoon liittyvää epävarmuutta.

Koneoppiminen on tekoälyn menetelmä, jossa algoritmit oppivat havaitsemaan tunnistettavia kaavoja ja tekemään ratkaisuja niiden pohjalta, ilman ohjelmoituja sääntöjä. **Syväoppiminen** on koneoppimisen osa-alue, mikä perustuu keinotekoisiin syviin neuroverkkoihin. Keskeisimmät erot koneoppimisen ja syväoppimisen välillä ovat, että syväoppiminen tarvitsee enemmän dataa ja laskentatehoa oppimiseen ja toimintaan (Jakhar & Kaur, 2020).

Russel (2021) kertoo tutkimuksessaan, että syväoppimista käytetään usein monimutkaisen ja muuttuvan datan käsittelyyn, kuten kuvioiden tai puheen käsittelyyn ja koneoppimista käytetään analytiikkaan ja erilaisiin ennustusmalleihin. Russel (2021) kuitenkin huomauttaa, että viimeaikaiset tutkimukset ovat osoittaneet, että syväoppimisjärjestelmät epäonnistuvat usein yleistämisessä luotettavasti ja ovat alttiita virheellisille säännönmukaisuuksille koulutusdatassa. Tekoälyoppimisella voidaan tarkoittaa havaittavaa muutosta tekoälyjärjestelmän tietyn vasteen tai päätöksenteon todennäköisyydessä tai intensiteetissä, jota tukevat laskennalliset algoritmit ja data (Gignac & Szodorai, 2024). Koneoppimiseen perustuvat päätöksentekomallit perustuvat algoritmeihin, jotka oppivat tekemään päätöksiä analysoimalla dataa ja tunnistamalla kaavoja, mutta koneoppiminen hyödyntää myös muita tekoälyn päätöksentekomalleja (Gicnac & Szodorai, 2024).

Muut tekoälyn päätöksentekomallit

Sääntöpohjainen päätöksentekomalli perustuu määriteltyihin sääntöihin, ehtoihin ja raja-arvoihin, sitä käytetään yksinkertaisissa järjestelmissä, mutta myös laajoissa järjestelmissä tuhansilla säännöillä, kuten koneoppimismalleissa (Kliegr, Bahnik & Furnkranz, 2021).

Päätöspuut toimivat haarautuvalla rakenteella, jossa haarat edustavat eri päätösvaihtoehtoja. Päätöspuu-päätöksentekomallia käytetään esimerkiksi lääketieteessä (Tanya, Nguyen, Buchanan & Jackman, 2023).

Bayesilainen päättely perustuu todennäköisyyksiin ja tilastollisiin malleihin. Tekoäly käyttää Bayeslaista päättelyä tunnistaakseen samankaltaisia rakenteita ja tehdäkseen niistä itsenäisiä johtopäätöksiä. Bayesilainen päättely on myös erittäin hyödyllistä koneoppimismalleille (Voskoglou, 2020).

Optimointimallit perustuvat tavoitteisiin kuten haittojen minimointiin ja hyötyjen maksimointiin, ja optimointimallit voivat hyödyntää Bayeslaista päättelyä toiminnassaan. Tekoälyn optimointipäätösmallin kykyjä voidaan käyttää esimerkiksi tietokoneiden, syväoppimismallien ja virrankäytön optimointiin (Surianarayanan, Lawrence, Chelliah, Prakash, & Hewage, 2023).

Ymmärrettävä tekoäly ja mustan laatikon malli

Useimmat tekoälymenetelmät, kuten syvät neuroverkot (syväoppiminen), ovat liian monimutkaisia ja vaikeasti ymmärrettäviä käyttäjälle ja ovat niin sanottuja mustan laatikon malleja (Schramm ym., 2023), jotka ovat tärkeä osa tekoälykeskustelua. Yksi tekoälyn keskustelluimmista aiheista mustan ovat laatikon mallit, yleisimmin ne ovat syväoppimismalleja ja niiden oppimiseen käytetystä suuresta ja monimutkaisesta datamäärästä johtuen, käyttäjä saa tietää vain lopputuloksen, mutta ei voida ymmärtää, miten tai miksi se on saavutettu. Tätä ongelmaa ratkaistakseen on kehitetty ymmärrettävää tekoälyä (Comprehensible AI tai CAI) ja selitettävää tekoälyä (Explainable AI tai XAI) (Schramm ym., 2023). Ymmärrettävä tekoäly tekee mustan laatikon tekoälymallista läpinäkyvän ilman, että sen suorituskyky heikkenee (Schramm ym., 2023).

Ymmärrettävät koneoppimismallit ovat tulkittavia, mikä saa käsitteen päällekkäiseksi selitettävän tekoälyn käytön kanssa. Epäselvä raja ja ylikuormitus johtuvat samankaltaisista korkeista vaateista, joita tulkittavat koneoppimismallit ja selitettävä tekoäly täyttävät (Schramm ym., 2023). Kun tulkittava koneoppiminen pyrkii paljastamaan koko koneoppimismallin päätöksentekoprosessin, selitettävä tekoäly keskittyy mustan laatikon mallin syötteen ja sen tuloksen välisen yhteyden kartoittamiseen (Schramm ym., 2023). Tekoälyn selitettävyys on tärkeää eettisesti aroilla aloilla, kuten henkilötietojen käsittelyssä tai kyberturvallisuudessa. Tekoälyn selitettävyys parantaa luottamusta, läpinäkyvyyttä, lisäksi tekoälymallia on helpompi kehittää, kun kehittäjät ymmärtävät miten ja miksi tekoäly luo päätöksiä. Myös lainsäädännön näkökulmasta selitettävyys ja toiminnan ymmärtäminen on tärkeää. Schramm ym., (2023) tutkimuksessaan kertovat, että selitettävä tekoäly keskittyy mustan laatikon mallin syötteen ja sen tuloksen välisen yhteyden kartoittamiseen. Tällöin selitettävät tekoälymenetelmät tarjoavat selityksiä koneoppimismallin käyttäytymiselle. Koska selitysmekanismi on erillinen itse mallista, täydellistä uskollisuutta ei voida taata. Selitettävän tekoälyn menetelmän laatua mitataan sillä, kuinka hyvin sen tarjoamat selitykset vastaavat mallin todellista käyttäytymistä. Selitettävän tekoälyn mallit eivät kuitenkaan tee mustan laatikon mallin sisäistä päättelyä ja prosessointia läpinäkyväksi.

3 TIETOJÄRJESTELMIEN KYBERTURVALLISUUS

Tietojärjestelmien kyberturvallisuus on keskeinen osa nykyaikaisten organisaatioiden riskienhallintaa ja tietojen suojausta. Tietojärjestelmät koostuvat usein hajautetuista komponenteista, jotka mahdollistavat tehokkaan tiedonkäsittelyn, mutta samalla lisäävät altistumista kyberuhille (Palko ym., 2023). Tämä luku tarkastelee tietojärjestelmien arkkitehtuuria, keskeisiä uhkatekijöitä sekä ratkaisuja, joita voidaan soveltaa niiden turvallisuuden varmistamiseksi.

Gupta ym., (2024) mukaan digitaalinen tieto on arvokas resurssi organisaatioille ja yksilöille, eikä niitä ole varaa menettää. Tietojärjestelmät ovat yhä monimutkaisempia ja sisältävät suuren määrän hajautettuja komponentteja, kuten tietokantoja, verkkopalvelimia ja sovellusliittymiä (Gupta ym., 2024). Palko ym., (2023) toteavat, että tietojärjestelmien hajautettu arkkitehtuuri koostuu lukuisista komponenteista, jotka ovat vuorovaikutuksessa toistensa kanssa ja kytketty toisiinsa tiedonsiirtokanavien avulla. Tämä hajautettu rakenne parantaa järjestelmien suorituskykyä, mutta tekee niistä samalla alttiita hyökkäyksille. Palko ym., (2023) mukaan on uskottavaa, että kaikki internetiin yhdistetyt laitteet ja ohjelmistot ovat lähtökohtaisesti haavoittuvia hakkeroinnille. Tutkijat ovat tunnistaneet digitaalisen intensiteetin merkittäväksi teknologiseksi tekijäksi, joka määrää tietomurtojen esiintymisen. Digitaalinen intensiteetti viittaa digitaaliteknologian käytön laajuuteen eri tasoilla organisaation operatiivisissa ja strategisissa toiminnoissa. Tutkimuksessa havaittiin, että sairaaloissa, joissa käytetään yhä enemmän sähköisiä potilastietojärjestelmiä, myös kyberhyökkäysten riski kasvaa (Mukhopadhyay & Jain 2024). Tekoälyn käytöllä voidaan kuitenkin vähentää ihmisten työtä digitaalisten järjestelmien parissa, jolloin ihmisen antamalla syötteellä tekoäly voi suorittaa esimerkiksi monimutkaisia kyberturvallisuustoimia (Murugesan, 2020).

Pilvipalveluiden käyttö lisää tietojärjestelmien tehokkuutta, mutta tuo mukanaan monimutkaisia tietoturvariskejä. Pilvipohjainen tietojenkäsittely, hajautetun tietojenkäsittely, mobiilitietojenkäsittely ja muiden teknologioiden nopea kehitys ja edistys ovat tuoneet mukanaan potentiaalisia hyökkäyksiä (Wang, Xue & Zhang, 2023). Pilvipalvelut käyttävät kolmea palvelumallia tarjotakseen erilaisia palveluja käyttäjille. Pilvipalveluiden SaaS (Software as a Service), PaaS

(Platform as a Service) ja IaaS (Infrastructure as a Service) -palvelumallit tarjoavat asiakkaille infrastruktuuriresursseja, sovellusalustoja ja ohjelmistoratkaisuja palveluna (Abdullayeva, 2023). Pilvipalvelukonseptin muodostavien elementtien, kuten verkon, arkkitehtuurin, sovellusohjelmointirajapinnan ja laitteiston moninaisuus lisää turvallisuusongelmien monimutkaisuutta (Abdullayeva, 2023). SaaS-mallin turvallisuuteen liittyy useita keskeisiä kysymyksiä, kuten tietoturva, verkkoturva, pääsynhallinta ja virtualisoinnin haavoittuvuudet. (Abdullayeva, 2023) PaaS-mallissa palveluntarjoaja vastaa alimpien kerrosten turvallisuudesta, mutta kehittäjät voivat rakentaa sovelluksia itse. Tietojen erottelu eri sovellusten välillä on tärkeää (Abdullayeva, 2023). IaaS-mallissa käyttäjä hallitsee itse IT-järjestelmän turvallisuutta, kun taas palveluntarjoaja vastaa fyysisistä ja virtualisoinnin turvatoimista (Abdullayeva, 2023).

3.1 Yleisimmät kyberuhat

Tietojärjestelmien yleisimmät uhat voidaan jakaa haittaohjelmiin, palvelunestohyökkäyksiin ja järjestelmän sisäisiin haavoittuvuuksiin. Käyn seuraavaksi läpi yleisimpiä kyberuhkia, ja kuinka ne toimivat.

SQL-injection -hyökkäys: Hyökkääjä lisää haitallisia SQL-komentoja SQL-tietokantaan suoritettavaksi, jos niitä ei ole tarpeeksi hyvin suojattu. Haitalliset komennot voivat aiheuttaa esimerkiksi erilaisten taulujen poistamisen tai tietojen muokkaamiseksi käyttökelvottomaksi (Alawida ym., 2024).

Zombiehyökkäys: Hyökkääjä lähettää pyyntöjä verkkoon liitetyistä harmittomista isäntäkoneista (zombeista), mikä aiheuttaa pilvipalvelun ylikuormituksen ja häiritsee sen suorituskykyä. Tämä voi johtaa DoS- tai DDoS-palvelunestohyökkäyksiin palvelimille, kun pilvialustalle kohdistuu liiallinen määrä pyyntöjä, jotka kuluttavat resursseja (Abdullayeva, 2023).

Service injection -hyökkäys: Tässä hyökkäysmuodossa hyökkääjä lisää pilvipalvelujärjestelmään haitallisen palvelun tai virtuaalikoneen, joka tarjoaa käyttäjille vahingollisia palveluja. Hyökkääjä luo esimerkiksi huijaripalvelun (kuten SaaS, PaaS tai IaaS) ja integroi sen osaksi pilvijärjestelmää. Tämän seurauksena käyttäjien normaalit pyynnöt ohjautuvat automaattisesti näille haitallisille palveluille, jolloin käyttäjät saavat vääristettyjä tai haitallisia palveluja, jotka voivat vaikuttaa pilvijärjestelmän toimintaan (Abdullayeva, 2023).

Virtuaalikoneen (VM) ohittaminen: Tässä hyökkäyksessä hyökkääjän ohjelma virtuaalikoneessa rikkoo eristyskerroksen ja saa sekä virtuaalikoneen että hypervisorin käyttöoikeudet. Tämä mahdollistaa suoran yhteyden hypervisoriin. VM Escape -hyökkäyksellä hyökkääjä saa pääsyn isäntäkoneen käyttöjärjestelmään ja muihin virtuaalikoneisiin samalla fyysisellä koneella (Abdullayeva, 2023).

Rootkit hypervisorissa: VM-pohjaiset rootkitit pakottavat hypervisorin tartuttamaan isäntäkoneen käyttöjärjestelmän ja luomaan piilotetun kanavan luvattoman koodin suorittamiseen. Tämä antaa hyökkääjälle hallinnan kaikkiin isäntäkoneella ajettaviin virtuaalikoneisiin ja mahdollisuuden manipuloida järjestelmän toimintoja (Abdullayeva, 2023).

Man in the middle -hyökkäys: Jos SSL (Secure Socket Layer) on väärin konfiguroitu, hyökkääjä voi päästä käsiksi kahden osapuolen väliseen tietojen jakamiseen. Pilvipalveluissa hyökkääjä voi saada pääsyn datakeskusten välisten viestintöjen tietoihin (Abdullayeva, 2023).

Metadata spoofing: Hyökkääjä muokkaa tiedostoa, joka sisältää palvelun tiedot, luodakseen väärää tietoa (Abdullayeva, 2023).

Phishing-hyökkäys: Hyökkääjä manipuloi verkkoyhteyksiä saadakseen arkaluonteisia tietoja ja ohjatakseen käyttäjän väärälle sivustolle, usein kaappaamalla käyttäjien tilejä tai palveluja pilvipalvelussa (Abdullayeva, 2023).

Backdoor-kanavahyökkäys: Hyökkääjä käyttää takaovia saadakseen etäyhteyden saastuttaneensa järjestelmään, hallitsee sen resursseja ja muuttaa sen zombiksi DDoS-hyökkäyksen toteuttamiseksi (Abdullayeva, 2023).

Petos-hyökkäykset: Pilvipalveluiden tehokkuus tallennus- ja laskentatehokkuuden osalta houkuttelee teollisuusorganisaatioita siirtämään hallintajärjestelmänsä pilveen. Petos-hyökkäykset pyrkivät vaarantamaan ohjaussignaalien eheyden muokkaamalla siirrettyä tietoa (Abdullayeva, 2023).

Palvelunestohyökkäykset (DoS): Tämä hyökkäys pyrkii ylikuormittamaan verkon, järjestelmän tai sovelluksen liiallisella liikenteellä, yhteyksillä tai pyynnöillä (Abdullayeva, 2023).

Wangin ym., (2023) mukaan tietokannan tietojärjestelmän turvallisuusuhkatiedonkeruu parantaa nopeasti verkossa leviävien Troijan hevosten, virusten ja hakkeri-iskujen havaitsemista. Nämä virukset osoittavat älykkyyden piirteitä. Wang ym., (2023) jatkavat, että niillä on pidempi piilevä aika, nopeampi leviämisnopeus, laajempi tartuntakohde ja ne ovat vaikeampia havaita haittaohjelmien torjuntaohjelmilla (RA) ja turvallisuusteknologioilla.

3.2 Tietojärjestelmien riskienhallinta

Nykyiset riskienhallintatekniikat perustuvat uhkien, vahinkojen ja haavoittuvuuksien todennäköisyyksien käyttöön. Kuitenkin useimmissa tapauksissa tietoturva-asiantuntijat tekevät arvioinnin sanallisten muotoilujen avulla ja liittävät ne sitten numeerisiin arvoihin käyttäen omaa kokemustaan (Palko ym., 2023).

Enterprise-tietojärjestelmän turvallisuuden riskinarviointiprosessi (RA) voidaan karkeasti jakaa kolmeen vaiheeseen: suunnitelman valmistelu, kenttäarviointi ja analyysiraportti (Wang ym., 2023). Tämä riskiarviointimenetelmä rajoittaa merkittävästi metodologian mahdollisuuksia yleisesti, sillä riskitekijöitä (uhkat, haavoittuvuudet ja vahingot) analysoidaan heuristisilla menetelmillä, mikä johtaa erilaiseen dataan, jos arviointi tehdään eri asiantuntijoiden toimesta (Wang ym., 2023). Sen vuoksi asiantuntijan arvioon liittyvä luotettavuus voi olla kyseenalainen. Tietojen yhdistelyanalyysi ja sääntöjen kaivaminen tulisi suorittaa automaattisella, systemaattisella ja älykkäällä käsittelyllä, jotta perinteisten riskin tunnistamismenetelmien puutteet voidaan korvata (Wang ym., 2023).

Palko ym., (2023) mukaan nykyiset tekniikat ovat hitaita eivätkä tuota haluttuja tuloksia. Arviointiprosessi vie paljon aikaa, mikä johtaa siihen, että tulosten relevanssi häviää nopeasti. Nykyään kaikkien organisaatioiden, joilla on

käytössä tietoturvapolitiikka, on arvioitava säännöllisesti niille vastuullisten tietojen riskit ja uhat (Palko ym., 2023). Monet yritykset suosivat ennaltaehkäisyn rahoittamista samalla, kun ne laiminlyövät riskien arvioinnin, käsittelyn, reagointisuunnitelmien laatimisen ja muut riskienhallinnan osa-alueet (Palko ym., 2023). Kvantitatiivinen riskianalyysi on numeerinen arvio riskin kokonaistavasta vaikutuksesta projektin tavoitteisiin, kuten kustannuksiin ja aikarajoihin (Palko ym., 2023). Kvalitatiivinen riskianalyysi priorisoi tunnistetut projektiriskit ennalta määritellyn arviointiasteikon avulla ja sisältää uhkien tunnistamisen sekä mahdollisten vaikutusten arvioinnin (Palko ym., 2023).

Palko ym., (2023) jatkavat että, tässä suhteessa kvalitatiivista lähestymistapaa käytetään nykyään riskianalyysissä. Se tarjoaa yksinkertaisen uhkien ja niihin liittyvien riskien luokittelun niiden vakavuuden mukaan. Nykyiset riskienhallinnan haasteet Palko ym., (2023) kertovat tutkimuksessaan hyvin: Arvioinnin riittämätön tarkkuus ja luotettavuus, vaikeudet aineettomien varojen (esim. maine, tiedon salassapito, ideat, liiketoimintasuunnitelmat, henkilöstön terveys) vahinkojen arvioinnissa; pitkän aikavälin kvantitatiivisen riskiarvioinnin tulosten arvon aleneminen jatkuvan automaattisen järjestelmän muokkauksen ja uudelleenkonfiguroinnin vuoksi; vaikeudet epäsuorien menetysten arvioinnissa; ja luotettavien tilastojen puute nopeasti muuttuvassa IT-maailmassa.

4 TEKOÄLY KYBERHYÖKKÄYKSISSÄ

Tässä luvussa käsitellään tekoälyn (AI) roolia kyberhyökkäyksissä ja sen vaikutusta kyberturvallisuuteen. Luvun alussa esitellään perinteiset kyberhyökkäysmenetelmät, kuten tietomurrot, phishing-hyökkäykset ja haittaohjelmat, sekä tarkastellaan, kuinka tekoäly voi tehostaa näitä menetelmiä ja tehdä niistä entistä tehokkaampia. Tekoälyn rooli hyökkäyksissä perustuu sen kykyyn analysoida suuria tietomääriä nopeasti, mukautua muuttuvaan ympäristöön ja suorittaa hyökkäyksiä automatisoidusti ilman jatkuvaa ihmisen valvontaa (Kaloudi & Li, 2020). Seuraavaksi tarkastellaan tekoälyn käyttöä kyberrikollisuudessa, erityisesti talouspetoksissa, kyberterrorismissa ja kyberkiristyksessä. Luvussa käydään myös läpi lyhyesti tekoälyn kyvykkyyttä luoda realistisia huijauksia, kuten äänenkalastelua (voice phishing), sekä tekoälyn roolia massahäirinnässä, kuten sosiaalisen manipulaation tekniikoissa. Luvun loppupuolella syvennytään tekoälyn rooliin kyberhyökkäyksen eri vaiheissa, mukaan lukien hyökkäyksen valmistelu, toteutus ja jälkiseuraukset. Erityistä huomiota kiinnitetään siihen, kuinka tekoäly voi parantaa hyökkäyksen tarkkuutta, optimoida resurssien käyttöä ja sopeutua puolustusmekanismeihin. Tämä luku tarjoaa kattavan katsauksen tekoälyn hyödyntämiseen kyberhyökkäyksissä ja sen aiheuttamiin turvallisuusuhkiin.

4.1 Tekoäly kyberhyökkäyksissä

Chen ym., (2024) mukaan viimeisten vuosikymmenien aikana kyberturvallisuuskenttä on todistanut kasvavaa eskalaatiota kyberhyökkäysten määrän kasvussa ja intensiteetissä. Illiashenko ym., (2023) puolestaan kertovat tutkimuksessaan, että kyberhyökkäykset ovat nousseet merkittäviksi uhiksi, jotka voivat vakavasti vaarantaa kriittisten infrastruktuurien, teollisuuden ohjausjärjestelmien ja liikenteen turvallisuutta. Toisin kuin perinteiset kyberhyökkäykset, jotka ovat tyypillisesti manuaalisia tai skriptejä, tekoälypohjaiset kyberhyökkäykset voivat itsenäisesti oppia ja kehittää taktiikoitaan, tekniikoitaan ja menettelytapojaan

reaaliaikaisen palautteen ja ympäristön muutosten perusteella (Osman & El-Gendy, 2024).

Jeongin (2020) mukaan kyberrikollisuutta pidetään kyberavaruuden pimeänä puolena. Hän jatkaa, että rikokset jaetaan kahteen tyyppiin: tietokone rikoksen kohteena ja tietokone rikoksen työkaluna. Kun tieto on digitalisoitu ja yhdistetty verkkoon, on syntynyt uudenlaisia rikoksia, kuten kyberterrori, kyberkiristys ja kybersodankäynti. Näitä rikoksia kutsutaan tietokone rikoksen kohteena -tyyppisiksi rikoksiksi (Jeong, 2020). Tietokone rikoksen kohteena -tyyppisten rikosten tavoitteena on häiritä tai tuhota tietokonejärjestelmiä käyttämällä hyökkäystyökaluja, kuten viruksia, matoja, troijalaisia ja vakoiluohjelmia. Samaan aikaan päivittäisen elämän tiedot ovat digitalisoituneet yksityiselämästä liiketoimintaan, mikä on siirtänyt perinteisiä rikoksia, kuten petoksia, uhkauksia, ja jopa lasten hyväksikäyttö- ja vainoamisyrityksiä verkkoon. Näitä puolestaan kutsutaan tietokone työkaluna -tyyppisiksi rikoksiksi (Jeong, 2020).

Kyberrikollisuus liittyy läheisesti kyberturvallisuuteen, sillä suurin osa kyberrikollisuuden hyökkäystekniikoista perustuu potentiaalisten kohteiden haavoittuvuuksien hyödyntämiseen (Jeong, 2020). Osman ja El-Gendy (2024) kertovat että, tekoälyn kehitys on merkittävästi muuttanut kyberuhkien laajuutta, aloittaen uuden aikakauden, jolle ovat ominaisia erittäin kehittyneet ja mukautuvat hyökkäykset. Viime aikojen tapahtumat piirtävät synkän kuvan tästä kehittyvästä todellisuudesta. Vuoden 2020 SolarWinds-hyökkäys, jossa käytettiin kehittyneitä tekoälyalgoritmeja, vaaransi 18 000 organisaation tietoverkot, mukaan lukien teknologiayrityksiä (kuten Microsoft ja FireEye) sekä useita valtion virastoja. Tämä korostaa tekoälyyn pohjautuvien kyberuhkien laajuutta ja kehittyneisyyttä (Osman & El-Gendy, 2024). CPS-järjestelmät (cyber physical systems) yhdistävät laskennallisten fyysisten järjestelmien, kuten tallennuksen, antureiden ja toimilaitteiden, integroinnin kriittisiin tehtäviin, jotta viestintäteknologioiden tehokkuus paranee (Abdullahi ym., 2024). Tekoälyyn pohjautuvat kyberhyökkäykset voivat lamauttaa toimitusketjuja, mikä johtaa saatavuusongelmiin, hintojen nousuun ja yritysten sulkemisiin. Ne voivat heikentää kuluttajien luottamusta, vahingoittaa yritysten mainetta, lisätä työttömyyttä ja hidastaa talouskasvua. Äärimmäisissä tapauksissa ne voivat jopa laukaista maailmanlaajuisia taantumia (Osman & El-Gendy, (2024). Falco, Viswanathan, Caldera ja Shrobe, (2018) kertovat tutkimuksessaan, että esimerkiksi asioiden internet (IoT) jää tulevaisuudessakin hakkeroinnille alttiiksi, sillä sitä on hankala turvata niiden laajuuden ja komponenttien määrän takia.

Tekoälyyn pohjautuvat kyberhyökkäykset tarkoittavat kyberhyökkäyksiä, jotka hyödyntävät tekoälyä ja koneoppimisalgoritmeja parantaakseen hyökkäysten tehokkuutta, huomaamattomuutta ja sopeutumiskykyä (Osman & El-Gendy, 2024). Nämä tekoälypohjaiset kyberhyökkäykset hyödyntävät koneoppimisalgoritmeja ja automaatiota monivaiheisten hyökkäysten suorittamiseen kriittiseen infrastruktuuriin, haavoittuvuuksien itsenäiseen tunnistamiseen, kohdennettujen hyökkäysten käynnistämiseen ja nopeaan kehittymiseen puolustusmekanismeja vastaan (Osman & El-Gendy, 2024). Nämä hyökkäykset hyödyntävät tekoälyalgoritmeja luodakseen harhaanjohtavia syötteitä, automaattisia sosiaalisen

manipuloinnin tekniikoita, parantaakseen haittaohjelmien kykyjä ja järjestääkseen laajamittaisia häiriöitä (Osman & El-Gendy, 2024). Illiashenko ym., (2023) mukaan yksittäisistä hakkereista kehittyneisiin hakkerikeskuksiin, vastustajat etsivät jatkuvasti haavoittuvuuksia saadakseen esimerkiksi luvattoman pääsyn, varastaakseen arkaluonteista tietoa, häiritäkseen toimintoja tai jopa aiheuttaakseen fyysistä vahinkoa.

4.1.1 Tekoälyn väärinkäyttö rikoksissa

Tekoälyrikos-termi esiteltiin alun perin humanististen tieteiden alalla, sillä rikostermi liittyy lakiin ja etiikkaan (Jeong, 2020). Jeong jatkaa tutkimuksessaan, kuinka jotkut tutkijat ovat varoittaneet, että hakkerit ovat jo alkaneet aseistaa tekoälyä parantaakseen murtamistaitojaan ja kehittääkseen uusia kyberhyökkäysten tyyppejä. Tekoälyn haitallinen käyttö lisää hyökkäysten nopeutta ja onnistumisprosenttia sekä vahvistaa hyökkäysten kykyjä. Tieto- ja viestintäteknologiat (ICT) ja tekoäly laajentavat mahdollisuuksia rikosten tekemiselle ja luovat uudenlaisen uhkakuvan, jossa voi toteutua uusia rikollisia taktiikoita (Kaloudi & Li, 2020). Kyberrikolliset ovat alkaneet parantaa tekniikoitaan sisällyttämällä esineiden internetin (IoT), haittaohjelmia, kiristysohjelmia ja tekoälyä lanseeratakseen entistä voimakkaampia hyökkäyksiä. Tällaisia hyökkäyksiä toteuttamalla kaikki ovat vaarassa, hyökkäysten yhteyksien ja älykkyyden vuoksi (Kaloudi & Li, 2020). Tekoälyä voidaan käyttää fyysisiin rikoksiin ohjaamalla autonomisia laitteita, kuten älykästä autoa, lennokkia, esineiden internetin laitteita ja muita vastaavia (Jeong, 2020). Esimerkiksi älykkäiden autojen hallinnan häiriöt voivat olla kohtalokkaita. Koneoppimista voidaan käyttää puolestaan aseena sosiaaliseen manipulointiin. Tekoälyn avulla voidaan luoda massatuotettuja viestejä, joissa on kalasteluviittauksia, ja julkaista niitä esimerkiksi Twitterissä ilman keskeytyksiä (Jeong, 2020). Tekoälyä hyödynnetään perinteisten kyberrikosten, kuten talouspetosten, kyberterrorismin ja kyberkiristyksen, tekniikoiden terävöittämiseen. Esimerkiksi, kun hakkerit tekevät puhekalastelua (voice phishing), he voivat huijata uhreja käyttämällä realistisesti jäljiteltyjä uhrin perheenjäsenten tai ystävien ääniä (Jeong, 2020).

4.2 Tekoälyn hyökkäysmenetelmät

Tässä alaluvussa tarkastellaan tekoälyn mahdollistamia hyökkäysmenetelmiä ja niiden vaikutuksia kyberturvallisuuteen. Tekoälyn käyttö kyberhyökkäyksissä on tehostanut perinteisiä menetelmiä monin tavoin, kuten hyökkäysten laajuuden kasvattamisella, heikkouksien automaattisella tunnistamisella ja hyökkäysten huomaamattomuuden lisäämisellä. Lisäksi tekoäly mahdollistaa hyökkäyksien sulauttamisen normaaliin verkkoliikenteeseen, mikä vaikeuttaa niiden havaitsemista ja torjumista. Tekoälyn käyttö kyberhyökkäysten tehokkuuden parantamiseksi ja ATS-omaisuuden suojaamiseksi on erittäin tärkeä tutkimus- ja kehityssuunta turvallisuuden ja suojauksen näkökulmasta (Illiashenko ym.,

2023). Hyökkäysten vastatoimien kehittämiseksi on keskeistä tutkia ja analysoida eri hyökkäysvektoreita, tekniikoita ja niihin liittyviä haavoittuvuuksia. Tämä on välttämätöntä ATS-järjestelmien (Automatic Transfer Switch) kohtaamien riskien ymmärtämiseksi ja erityisesti tekoälypohjaisten hyökkäysten huomioimiseksi (Illiashenko ym., 2023). Eri hyökkäysvektoreiden, tekniikoiden ja niihin liittyvien haavoittuvuuksien tutkiminen ja analysointi on välttämätöntä ATS-järjestelmien kohtaamien riskien ymmärtämiseksi ja tehokkaiden vastatoimien kehittämiseksi, erityisesti tekoälypohjaisia hyökkäyksiä ja suojaustoimenpiteitä silmällä pitäen (Illiashenko ym., 2023). Toisin kuin perinteiset kyberhyökkäykset, jotka ovat tyypillisesti manuaalisia tai ennalta kirjoitettuja skriptejä, tekoälyohjatut kyberhyökkäykset voivat itsenäisesti oppia ja kehittää taktiikoitaan, tekniikoitaan ja menettelytapojaan reaaliaikaisen palautteen ja ympäristön muutosten perusteella (Osman & El-Gendy, 2024).

Osman ja El-Gendy (2024) käyvät hyvin läpi tekoälyn hyökkäysmenetelmiä tutkimuksessaan, joka liittyy tekoälyn käyttöön kyberhyökkäyksissä kohdistuen maailmankauppaan. Tutkimuksessa esitetyt hyökkäysmenetelmät antavat erinomaisen kuvan tekoälypohjaisten hyökkäysten kyvykkyydestä, kehittyneisyydestä ja vaarallisuudesta kaikilla aloilla. Alla oleva lista perustuu Osman ja El-Gendyn (2024) tutkimukseen.

Adversariaaliset hyökkäykset

Tekoälyä käytetään tarkoituksellisesti harhauttamaan koneoppimismalleja luomalla vilpillisiä syötteitä.

- Kaupankäyntialgoritmien häiriöt, markkinoiden epävakaus ja taloudelliset tappiot.
- Väärät tulliluokitukset, jotka vaikuttavat kauppasopimuksiin aiheuttaen viivästyksiä tai sakkoja kansainvälisessä kaupassa.
- Luottamuksen väheneminen tekoälyteknologioihin, mikä hidastaa taloudellista toimintaa, kuten toimitusketjun hallintaa ja rahoituspalveluja.
- Virheelliset markkinointistrategiat heikentävät markkinoiden kilpailukykyä.
- Yli- tai aliresurssien käyttö toimitusketjuissa heikentää taloudellista tuottavuutta.

Automatisoidut sosiaalisen manipuloinnin hyökkäykset

Tekoälyä hyödynnetään räätälöityjen phishing-sähköpostien ja petosviestien luomisessa.

- Taloudelliset tappiot petollisten transaktioiden ja luvattomien tilikäyttöjen vuoksi.

- Toimitusketjujen toiminnan häiriintyminen ja kauppatapahtumien viivästykset.
- Luottamuksen väheneminen digitaalisiin viestintäkanaviin, mikä heikentää sähköisen kaupankäynnin aloitteita.

Tekoälyä hyödyntävä haittaohjelmisto

Malware, joka käyttää tekoälyä esimerkiksi tunnistamisen välttelyyn ja muuntautumiskykyyn.

- Tuotannon, jakelun ja tavarantoimitusten viivästykset vaikuttavat globaaleihin kauppavirtoihin.
- Kilpailukyvyn heikkeneminen omistusoikeudellisen tiedon varkauden vuoksi.
- Asiakkaiden luottamuksen menetys heikentää liiketoimintaa ja mainetta.
- Kyberturvallisuuskustannukset ohjaavat resursseja pois ydintaloudellisista toiminnoista.

Generatiiviset vastakkaisverkostohyökkäykset (GANs)

GAN-verkostoja käytetään synteettisen sisällön, kuten disinformaation, rahoituspetosten ja väärennösten luomiseen.

- Vaikutukset julkiseen mielipiteeseen ja yhteiskunnan vakauteen disinformaatiokampanjoiden kautta.
- Taloudelliset tappiot väärennetyistä tuotteista ja markkinoiden epävarmuudesta.
- Henkilöiden identiteetin väärinkäyttö vahingoittaa mainetta ja rikkoo yksityisyyttä.
- Sääntelyhaasteet ja kyberturvallisuusriskit heikentävät digitaalisen median luotettavuutta.

Tekoälyä hyödyntävät datan saastuttamishyökkäykset

Malware syöttää haitallista dataa koulutusaineistoihin manipuloidakseen tekoälymalleja.

- Heikentää AI-mallien tarkkuutta ja luotettavuutta kauppaan liittyvissä päätöksissä.
- Taloudelliset tappiot suboptimaalisten investointipäätösten ja resurssien väärän allokoinnin vuoksi.

- Markkinavääristymät, kuten epäreilu kilpailu ja syrjivät käytännöt.

Mallin käänteishyökkäykset

Hyödynnetään koneoppimismallien läpinäkyvyyttä arkaluonteisten tietojen paljastamiseksi.

- Yksityisyysloukkaukset ja luottamuksellisen tiedon luvaton käyttö.
- Teollisuusvakoilu heikentää kilpailuetuja ja oikeudenmukaisia kauppakäytäntöjä.
- AI-järjestelmien uskottavuuden heikkeneminen kauppaan liittyvässä päätöksenteossa.

Tekoälyohjatut kiristysohjelmat

Tekoälyä käytetään kohdentamaan arvokasta dataa ja järjestelmiä kiristysohjelmilla.

- Suorat taloudelliset tappiot ja pitkäkestoiset taloudelliset vaikutukset.
- Tuotannon ja liiketoiminnan häiriöt aiheuttavat tuottavuuden laskua ja taloudellisia vastoinkäymisiä.
- Kuluttajien luottamuksen väheneminen vaikuttaa kauppasuhteisiin.

Tekoälyä hyödyntävä tiedustelu

Automaattinen tietojen keruu ja analyysi järjestelmien haavoittuvuuksien tunnistamiseksi.

- Markkinoiden epävakaus ja toimitusketjujen haavoittuvuudet.
- Teollisuuden tietovarkaudet ja luottamuksen mureneminen kauppakumppaneiden välillä.
- Sääntelytaakka ja operatiivisten kustannusten kasvu.

Tekoälyohjattu kiertäminen

Tekoälyä käytetään hyökkäysstrategioiden dynaamiseen mukauttamiseen reaaliaikaisen palautteen perusteella.

- Kaupan toimintojen häiriöt ja kustannusten nousu.
- Luottamuksen heikentyminen kaupankäyntijärjestelmiin.

Tekoälyohjatut DDoS-hyökkäykset

Tekoälyä hyödynnetään tehokkaampien hajautettujen palvelunestohyökkäysten toteuttamiseen.

- Kauppasuhteiden häiriintyminen ja verkkopalveluiden luottamuksen heikkeneminen.
- Sääntelytoimenpiteet lisäävät vaatimustenmukaisuuden kustannuksia ja vaikuttavat kilpailukykyyn.

Skaalautuvat uhkat

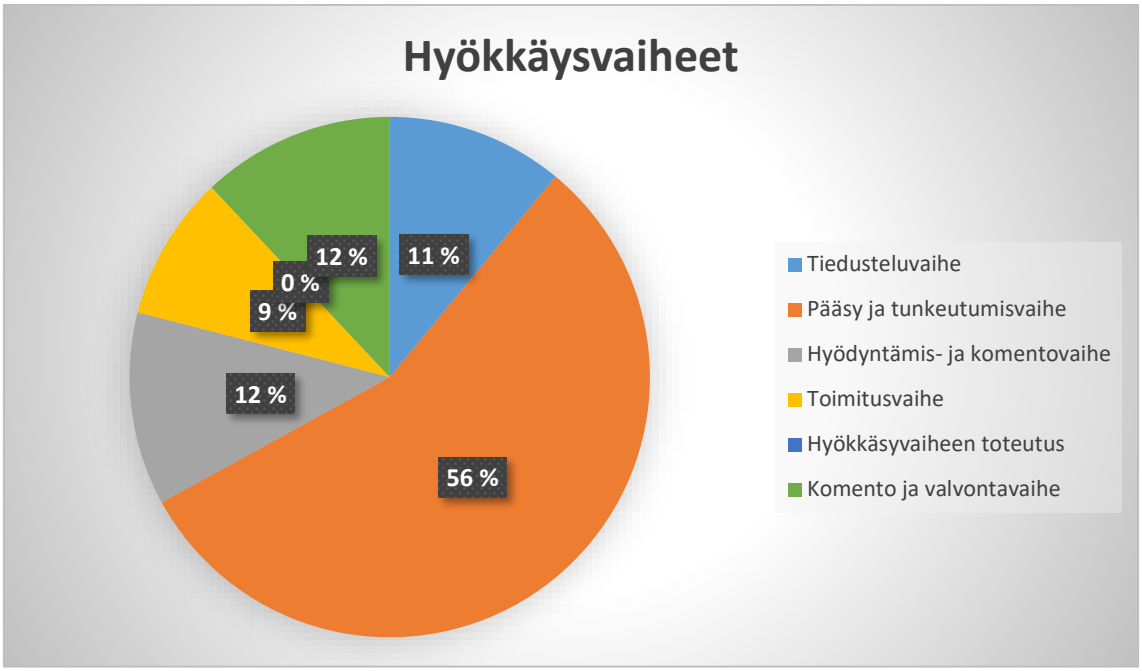
Tekoälyä käytetään laajamittaisiin ja koordinoituihin hyökkäyksiin.

- Kriittisen infrastruktuurin ja toimitusketjujen häiriöt aiheuttavat tuotannon viivästyksiä ja kustannusten nousua.
- Suurten hyökkäysten laajuus ylittää kyberturvallisuuspuolustuksen kapasiteetin.

(Osman & El-Gendy, 2024, s. 3-4)

4.3 Tekoäly hyökkäyksen eri vaiheissa

Tässä luvussa tarkastellaan tekoälyn roolia kyberhyökkäysten eri vaiheissa, jotka seuraavat perinteistä tietoturvahyökkäyksen toimintaketjua (cybersecurity kill chain). Tekoälyä voidaan hyödyntää jokaisessa hyökkäyksen vaiheessa erilaisiin tarkoituksiin, kuten tiedonkeruuseen, tunkeutumiseen, hyödyntämiseen, toimitukseen, hallintaan sekä hyökkäysten lopullisiin tavoitteisiin. Luvussa analysoidaan, miten tekoälypohjaiset menetelmät tehostavat hyökkäysten tehokkuutta, suurentavat skaalaa ja huomaamattomuutta eri vaiheissa, sekä pohditaan niiden vaikutuksia tietoturvajärjestelmien haavoittuvuuksiin. Pääsy- ja tunkeutumisvaiheessa (tekoälyavusteinen hyökkäys) tunnistettiin kuusi erilaista tekoälyllä toteutettua hyökkäystyyppiä, kun taas tiedusteluvaiheessa (tekoälykohdennettu hyökkäys) tunnistettiin neljä erilaista hyökkäystyyppiä. Hyödyntämisvaiheessa (tekoälyautomaattinen hyökkäys) tunnistettiin puolestaan kolme hyökkäystyyppiä, ja toimitusvaiheessa (tekoälykätkemishyökkäys) sekä C2-vaiheessa (tekoälymonivaihehyökkäys) kummassakin tunnistettiin kaksi erilaista hyökkäystyyppiä. Sen sijaan tavoitteiden toteuttamisvaiheessa (tekoälymalware-hyökkäys) tunnistettiin vain yksi tekoälyllä toteutettu hyökkäystyyppi (Guempe ym., 2022).



Guempe ym., (2024, s. 10)

K-tavan klusterointia käytettiin myös demonstroimaan, kuinka tekoälypohjainen itseoppiva haittaohjelma voi onnistuneesti hyödyntää haavoittuvuuksia turvallisuusvalvontajärjestelmissä. Haittaohjelmat voivat toimia ikään kuin ne olisivat tahattomia vikoja tietokonesovelluksissa, hyödyntäen ja vaarantaen herkkiä ympäristön ohjausinfrastruktuureja kyberturvallisuuden tappoketjun hyödyntämisvaiheessa (Guempe ym., 2024). Tekoälypohjaisista kyberhyökkäystekniikoista 56 % havaittiin pääsy- ja tunkeutumisvaiheessa, 12 % hyödyntämis- ja komentohallintavaiheessa, 11 % tiedusteluvaiheessa ja 9 % toimitusvaiheessa (Guempe ym., 2024). Hyökkäystavoitteiden toteutusvaiheessa ei havaittu tekoälytekniikoita hyökkäysten suorittamiseen (Guempe ym., 2024).

Tiedusteluvaihe

Tiedusteluvaiheessa tunnistettiin kolme erilaista tekoälytekniikkaa. Tutkimukset osoittivat, kuinka haitalliset toimijat voivat hyödyntää Markov-ketjuja/LTSM:ää, neuroverkkoja (NN) ja syviä neuroverkkoja (DNN) haavoittuvuuden ennakoimiseen, End-to-End (E2E) suorahyökkäyksiin ja älykkääseen kohdeprofiilointiin/tiedonkeruuseen (Guempe ym., 2024).

Tunkeutumisvaihe

Pääsy- ja tunkeutumistilanteessa tutkimuksessaan Guempe ym., (2024) tunnistivat kuusi tekoälypohjaista hyökkäystapaa, kuten salasanan murtamisen laskentatehoa käyttämällä (brute force -hyökkäys), älykkään captcha-menetelmän manipuloinnin, poikkeavan käyttäytymisen luomisen, tekoälymallin manipuloinnin ja älykkäiden valearvostelujen tuottamisen. Tutkimus osoitti myös

yhdeksäntoista tekoälytekniikkaa, joita haitalliset toimijat voivat hyödyntää pääsy- ja tunkeutumishyökkäyksissä.

Toimitusvaihe

Toimitusvaiheessa tunnistettiin kaksi tekoälypohjaista kyberhyökkäystyyppiä: älykäs peittely ja väistävä haittaohjelma. Tutkimukset osoittivat kolme tekoälytekniikkaa, joita haitalliset toimijat voivat käyttää peittely- ja väistämishyökkäysten toteuttamiseen (Guempe ym., 2024).

Hyväksikäyttövaihe

Hyökkäyksen hyödyntämisvaiheessa saadaan valtuutettu pääsy tietokoneohjelmiin ja resursseihin. Pääsyn jälkeen haitalliset toimijat voivat käyttää tekoälytekniikoita monimutkaisten ja havaitsemista vaikeuttavien hyökkäysten toteuttamiseen (Guempe ym., 2024).

Komento ja ohjausvaihe

Komento- ja ohjausvaiheessa havaittiin kaksi tekoälypohjaista kyberhyökkäystyyppiä: älykäs itseoppiva haittaohjelma ja automatisoitu verkkotunnusten generointi (Guempe ym., 2024).

5 TEKOÄLYPOHJAISTEN UHKIEN VAIKUTUS TIETOJÄRJESTELMIIN JA NIIDEN TORJUNTA

Tässä luvussa käsitellään tekoälypohjaisten uhkien vaikutusta tietojärjestelmiin ja niiden kyberturvallisuutta kehittäviä menetelmiä. Tietojärjestelmät ovat kehittyneet vastaamaan lähes kaikkiin elämäämme ympäröiviin toimintoihin, kuten pankkipalveluihin, viestintään, kaupankäyntiin, turvallisuuteen ja terveydenhuoltoon. Siksi onkin erittäin tärkeää ymmärtää tekoälypohjaisten uhkien vaikutus näitä toimia ylläpitäviin tietojärjestelmiin. Tekoälypohjaiset uhat ja kyberhyökkäykset tulevat lisääntymään tulevaisuudessa ja tietojärjestelmien kyberturvallisuuden kehittäminen on kriittisen tärkeää. Monet tietojärjestelmät ovat kyberrikollisille houkuttelevia kohteita, sillä ne sisältävät lukuisten ihmisten tärkeää dataa, kuten henkilötietoja ja salasanoja. Näillä tietomurroilla voidaan esimerkiksi kiristää organisaatioita, yrityksiä ja ihmisiä.

Tietojärjestelmiä on moneen eri tarkoitukseen, joten jokaisella erilaisella tietojärjestelmällä on omat haavoittuvuutensa ja haasteensa kyberturvallisuudessa. Siksi kyberturvallisuuden jatkuva kehittäminen on tärkeää palvelujen, datan ja ihmisten luottamuksen ylläpitämiseksi.

5.1 Tekoälypohjaisten uhkien vaikutus tietojärjestelmiin

Larriva-Novo ym., (2020) toteavat tutkimuksessaan että, internet-elementtien sisällyttäminen ihmisten päivittäiseen elämään, kuten esineiden internet, puettavat teknologiat tai Ubicomp, sovellettuna herkkään käyttäjäkontekstiin, kuten terveydenhuoltoon, tarjoaa uusia tunkeutumisvektoreita, jotka vaikuttavat suoraan ihmisten elämään. Sama riski koskee myös tietojärjestelmiä useissa erilaisissa konteksteissa: mitä enemmän erilaiset elementit käyttävät tietojärjestelmää, sitä useampi hyökkäysvektori siihen löytyy. Kaur, Gabrijelčič ja Klobučar, (2023) toteavat, että kyberturvallisuus on muuttumassa monimutkaisemmaksi, koska laitteiden, järjestelmien ja verkkojen välinen yhteyksien kasvu on

eksponentiaalista. Tutkimuksessaan Azambuja ym., (2023) kertovat, että kyberhyökkäykset vaihtelevat menetelmiään ja hyökkäysstrategioitaan lisätäkseen hyökkäyskykyään keskittyen tekoälyteknologioiden soveltamiseen.

Tekoälyn haitallinen käyttö on muuttanut kyberympäristön potentiaalisten uhkien tilannetta. Uhkamaisemaan liittyy useita toimijoita, ja hyökkääjät etsivät erilaisia haavoittuvuuksia hyökätäkseen. Näihin hyökkäyksiin sisältyvät muun muassa kehittyneiden jatkuvien uhkien monimutkaisuus ja hienostuneisuus, haitalliset toimet kyberavaruudessa sekä kyberrikollisuuden kaupallistaminen (Azembuja ym., 2023). Heidän mukaansa, tekoälypohjaiset uhat ovat riski suurille tietojärjestelmille ominaisuuksiensa takia. Kaloudi ja Li (2020) kertovat että, rikolliset käyttävät tekoälyä automatisoidakseen hyökkäysten toimintoja ja oppimaan optimaaliset hyökkäysvektorit. Tulevaisuudessa tekoälypohjaiset kyberhyökkäykset pystyvät mukautumaan ja muuttumaan kesken hyökkäyksen, perustuen hyökättävään kohteeseen (Guempe ym., 2022). Tulevaisuudessa tietojärjestelmät tulevat olemaan entistä haavoittuvaisempia tekoälypohjaisten kyberhyökkäysten edessä, ellei kyberturvallisuutta paranneta. Lisäksi vain suuret tietojärjestelmät eivät ole uhkien kohteena. Mirskyn ym., (2022) mukaan rikolliset voivat kehittää tekoälyn, joka pystyy valitsemaan hyökkäyksiä eri organisaatioihin niiden helppouden ja vahvimpien hyökkäysvektoreiden perusteella. Tämän takia myös pienten ja keskisuurten yritysten tulisi panostaa tietoturvaan, sillä pahimmillaan tekoälypohjainen hyökkäys voi kaataa yritysten kaikki toiminnot.

5.2 Tekoälypohjaisten kyberhyökkäysten torjunta

Tekoälypohjaisten kyberhyökkäysten torjumiseksi esitetään monenlaisia ratkaisuja. Esitettyihin ratkaisuihin kuuluvat tekoälyn käyttö hyökkäysten tunnistamiseen ja estämiseen eri menetelmillä, kuten selitettävällä tekoälyllä, ja erilaiset kone- ja syväoppimismallit. Selitettävä tekoäly on tehokas työkalu lisäämään perinteisten kone- ja syväoppimismenetelmiä käyttävien tekoälymallien selitettävyyttä ja läpinäkyvyyttä (Zhang, Damiani, Al Hamadi, Yeun, & Taher, 2022).

Tutkimuksissa ehdotetaan sitkeiden tietojärjestelmien kehittämistä tekoälyyn integroituna, jolloin tekoäly pystyy tekemään reaaliaikaisia korjauksia ja muutoksia palvelun jatkumiseksi hyökkäyksen keskellä. Kaur ym., (2023) kertovat monien vaihtoehtojen joukosta kolme keskeistä uutta tekniikkaa, jotka voivat vaikuttaa merkittävästi käytännöllisen ja käyttökelpoisen tekoälyn kehittämiseen kyberturvallisuudessa. Nämä tekniikat ovat tietolähteiden analyysi, selitettävä tekoäly ja täydennetty älykkyys (ihmisen ja tekoälyn rajapinnat) (Kaur ym., 2023). Hussain, Du, Sun ja Han (2020) puolestaan ehdottavat yhtenäistä runkoa, joka hyödyntää syviä konvoluutioisia neuroverkkoja ja todellisia verkon tietoja, joka tarjoaa varhaisen havaitsemisen hajautetuille palvelunestohyökkäyksille. Mirsky ym., (2022) suosittelevat panostusta haavoittuvaisuuksia havaitsevan ja tutkivan tekoälyn kehitykseen. Shih, Yang, Jiang, ja Kristiani (2023) ehdottavat

tutkimuksessaan tekoälyn ja data laken yhdistämistä luodakseen alustan, joka tukee hyökkäyksen tunnistamista ilmoittamalla operaattorille siitä. Tällaisella tekoälyllä voidaan etsiä organisaatioiden tietojärjestelmissä olevia haavoittuvuuksia ja korjata ne ennen kuin niitä hyväksikäytetään.

Mirsky ym., (2022) toteavat tutkimuksessaan, että kyberturvallisuudessa käytettäviä ML-malleja tulisi kehittää siten, että ne sisältävät kyvykkyydet turvallisuustestaukseen, suojaukseen ja valvontaan. Tekoälypohjaiset hyökkäykset alkavat vääjäämättä käyttämään automatisoituja hyökkäyksiä, jolloin niiden volyymin ja tarkkuuden takia tarvitaan kyberturvallisuudessa tekoälyä vastatoimena sitä vastaan. Tekoälyn sisällyttäminen tunkeilijan havaitsemisjärjestelmään voi tuottaa erittäin positiivisia tuloksia (Larriva-Novo ym., 2020). Larriva-Novo ym., (2020) mukaan tekoälyn ja varsinkin koneoppimisen sisällyttäminen tunkeilijoiden havaitsemis- ja suojausjärjestelmiin on todella tehokasta. Heidän mukaansa haasteena onkin tekoälyn koulutukseen vaadittavan valtavan datamäärän saaminen, sillä dataa omaavat yritykset ja organisaatiot pitävät siitä tiukasti kiinni. Tämä hidastaa tekoälyn koulutusta ja kehitystä kyberturvallisuudessa. Hwang:n, Shinin, ja Kimin (2022) mukaan kyberturvallisuuden muutosten ja trendien seuranta on tehokas tapa vastata kyberturvallisuusriskeihin proaktiivisesti ja se on kriittistä seuraavan sukupolven kyberturvallisuuden kehittämisessä. Tulevaisuudessa tietojärjestelmien kyberturvallisuus on entistä haastavampaa. Kasvavat uhat ja monimutkaiset ja moniulotteiset tietojärjestelmät ovat entistä haavoittuvaisempia tekoälypohjaisille hyökkäyksille.

5.3 Sitkeät tietojärjestelmät

Tekoälyllä on merkittävä rooli sitkeiden tietojärjestelmien (Resilient Information Systems, RIS) kehittämisessä erityisesti saatavuuden ja kyberturvallisuuden parantamiseksi. Guptan ym., (2024) mukaan perinteiset tietojärjestelmät eivät välttämättä omaa riittävästi analyyttistä tehoa vastatakseen nykypäivän monimutkaisiin ja nopeasti kehittyviin kyberuhkiin. He korostavat, että järjestelmä määritellään sitkeäksi, jos se suojaa nopeasti ja tehokkaasti keskeisiä toimintakykyjään häiriöltä, joka johtuu vastakkaisista tapahtumista ja olosuhteista (Gupta ym., 2024). Sitkeiden tietojärjestelmien kehityksessä korostetaan päivitysvaraa ja kykyä laajentua uusilla moniulotteisilla kyvyillä, joista tekoälyn integroiminen on keskeinen esimerkki. Gupta ym. (2024) osoittavat tutkimuksessaan, että tekoälypohjaiset tietojärjestelmät voivat parantaa kestävyyttä erityisesti häiriöiden ja kyberhyökkäysten aikana. Kaurin, Gabrijelčičin ja Klobučarin (2023) tutkimus tukee tätä näkemystä tuomalla esiin, että kyberturvallisuuden hallinta muuttuu jatkuvasti haastavammaksi johtuen verkottuneiden järjestelmien ja teknologioiden nopeasta kasvusta. Tässä yhteydessä tekoälyn integroiminen kyberturvallisuuteen tarjoaa merkittävän lisän, mikäli se implementoidaan oikein. Tekoäly voi esimerkiksi tuottaa analytiikkaa ja reaaliaikaista uhkatietoa, minkä avulla

kyberturvallisuustoimenpiteet voidaan sopeuttaa dynaamisesti muuttuvaan uhkakenttään (Kaur ym., 2023).

6 YHTEENVETO

Tämä tutkielma tarkastelee tekoälypohjaisia kyberuhkia ja niiden vaikutuksia tietojärjestelmiin sekä ratkaisuja näiden uhkien hallintaan. Tekoälypohjaiset uhat ovat ja tulevat kehittymään vaarallisiksi työkaluiksi rikollisten käsissä. Tutkielman tutkimusongelmana oli, kuinka tekoälypohjaiset kyberuhat vaikuttavat tietojärjestelmiin sekä uhkien torjunta. Tutkimuskysymyksenä oli: Miten tekoälypohjaiset uhat vaikuttavat tietojärjestelmiin, ja kuinka niitä voidaan torjua?

Tekoäly -luvussa käytiin läpi tekoälyn ominaisuuksia, kuten suuren datamäärän käsittelyä ja käyttöä, käyttötarkoituksia yhteiskunnassa, haasteita ja kehitystarpeita etenkin mustan laatikon malliin liittyen. Luvussa ''Tietojärjestelmien kyberturvallisuus'' käsitellään tietojärjestelmien kyberturvallisuutta, minkä haasteena on järjestelmien monimutkaisuus ja riskienhallinnan tärkeys. Tekoäly kyberhyökkäyksissä -luvussa käsitellään puolestaan tekoälypohjaisten kyberhyökkäysten käyttökohteita, hyökkäysmenetelmiä ja tekoälyn käyttöä hyökkäysten eri vaiheissa. Tekoäly kyberturvallisuudessa -luvussa käydään läpi tekoälypohjaisten uhkien torjuntaa ja haasteita, mitä tekoälypohjaisten kyberhyökkäysten luoma uhka tekee. Luvussa lisäksi käsitellään käytännön menetelmiä, millä suojautua tekoälypohjaisilta kyberuhilta, kuten sitkeitä tietojärjestelmiä, tekoälyn ja ihmisten yhteistyö kyberturvallisuudessa ja tekoälyn integrointi vastatoimena kyberhyökkäyksiä vastaan.

Keskeisenä löydöksenä havaittiin tekoälyn kaksikäyttöisyys. Tekoälyn soveltaminen lisää sekä hyökkäysten että puolustuksen tehokkuutta, mikä korostaa tekoälyyn perustuvien tietojärjestelmien resilienssin (RIS) merkitystä kyberturvallisuudessa. Tärkeää on tekoälyn soveltaminen myös selitettävän tekoälyn (XAI) ja koneoppimisen (ML) avulla, mikä voi parantaa sekä tietoturvaa että käyttäjäluottamusta. Kyberturvallisuudessa tekoälyllä pystytään havaitsemaan tietojärjestelmien haavoittuvaisuudet, epäilyttävä toiminta sekä tekemään vastaiskuja kyberhyökkäykselle ja turvaamaan järjestelmän toimintoja. Kuitenkin parhaat tulokset saavutetaan ihmisen ja tekoälyn yhteistyöllä. Tutkielman tulokset perustuvat kirjallisuuskatsaukseen, mikä tarjoaa vahvan teoreettisen pohjan, mutta rajoitteena on empiirisen tutkimuksen puute. Lisäksi tekoälyn nopea kehittyminen asettaa haasteita tutkimuksen ajankohtaisuudelle. Sitkeiden

tietojärjestelmien käytännön soveltamisessa ja tekoälyratkaisujen todellisessa ympäristössä testaamisessa on vielä kehitystyötä tehtävänä.

Jatkotutkimuksissa voidaan keskittyä esimerkiksi tekoälyn käytännön soveltamiseen kyberturvallisuudessa erityisesti kriittisten infrastruktuurien suojaamisessa. Lisäksi jatkotutkimuksissa tulisi tutkia tekoälypohjaisten uhkien roolia kyberturvallisuuden näkökulmasta. Myös tekoälyn selitettävyyteen ja eettisiin haasteisiin liittyvien ratkaisujen kehittäminen sekä lainsäädännön roolin tarkastelu tarjoavat merkittäviä jatkotutkimusaiheita. Empiiriset tutkimukset, joissa arvioidaan tekoälyyn perustuvien ratkaisujen vaikutuksia kyberuhkien torjunnassa, ovat erityisen tarpeellisia. Mahdollisia etuja, joita tekoälyn ja ihmisen yhteistyöllä voidaan saavuttaa, on tärkeää tutkia etenkin kyberturvallisuuden näkökulmasta.

LÄHTEET

Abdullayeva, F. (2023). Cyber resilience and cybersecurity issues of intelligent cloud computing systems. Results in Control and Optimization, 12, 100268. https://doi.org/10.1016/j.rico.2023.100268

Abdullahi, M., Alhussian, H., Aziz, N., Abdulkadir, S. J., Muazu, A. A., Alwadain, A. & Bala, A. (2024). Comparison and investigation of AI-based approaches for cyberattack detection in cyber-physical systems. IEEE Access, 12, 14523-14537. https://doi.org/10.1109/ACCESS.2024.3370436

Alawida, M., Abu Shawar, B., Abiodun, O. I., Mehmood, A., Omolara, A. E. & Al Hwaitat, A. K. (2024). Unveiling the dark side of ChatGPT: Exploring cyberattacks and enhancing user awareness. Information, 15(1), 27. https://doi.org/10.3390/info15010027

Chen, Y., Cui, M., Wang, D., Cao, Y., Yang, P., Jiang, B., Lu, Z. & Liu, B. (2024). A survey of large language models for cyber threat detection. Institute of Information Engineering, Chinese Academy of Sciences; School of Cyber Security, University of Chinese Academy of Sciences.

de Azambuja, A. J. G., Plesker, C., Schützer, K., Anderl, R., Schleich, B. & Almeida, V. R. (2023). Artificial intelligence-based cybersecurity in the context of Industry 4.0—A survey. Electronics, 12(8), 1920. https://doi.org/10.3390/electronics12081920

Falco, G., Viswanathan, A., Caldera, C. & Shrobe, H. (2018). A master attack methodology for an AI-based automated attack planner for smart cities. IEEE Access, 6, 42917-42929. https://doi.org/10.1109/ACCESS.2018.2867556

Gignac, G. E. & Szodorai, E. T. (2024). Defining intelligence: Bridging the gap between human and artificial perspectives. Intelligence, 104, 101832. https://doi.org/10.1016/j.intell.2024.101832

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. Applied Artificial Intelligence, 36(1), 2037254. https://doi.org/10.1080/08839514.2022.2037254

Gupta, C., Johri, I., Srinivasan, K., Hu, Y.-C., Qaisar, S. M. & Huang, K.-Y. (2022). A systematic review on machine learning and deep learning models for electronic information security in mobile networks. Sensors, 22(5), 2017. https://doi.org/10.3390/s22052017

Gupta, S., Modgil, S., Meissonier, R. & Dwivedi, Y. K. (2024). Artificial intelligence and information system resilience to cope with supply chain disruption. IEEE Transactions on Engineering Management, 71, 1–12. https://doi.org/10.1109/TEM.2021.3116770

Hussain, B., Du, Q., Sun, B. & Han, Z. (2020). Deep learning-based DDoS attack detection for cyber-physical systems over 5G network. IEEE Transactions on Industrial Informatics, 17(2), 1329-1337. https://doi.org/10.1109/TII.2020.2974520

Hwang, S.-Y., Shin, D.-J. & Kim, J.-J. (2022). Systematic review on identification and prediction of deep learning-based cybersecurity technology and convergence fields. Symmetry, *14*(4), 683. https://doi.org/10.3390/sym14040683

Illiashenko, O., Kharchenko, V., Babeshko, I., Fesenko, H. & Di Giandomenico, F. (2023). Security-informed safety analysis of autonomous transport systems considering AI-powered cyberattacks and protection. Entropy, 25(8), 1123. https://doi.org/10.3390/e25081123

Jakhar, D. & Kaur, I. (2020). Artificial intelligence, machine learning and deep learning: Definitions and differences. Clinical and Experimental Dermatology, 45(1), 131–132. https://doi.org/10.1111/ced.14030

Jeong, D. (2020). Artificial intelligence security threat, crime, and forensics: Taxonomy and open issues. IEEE Access, 8, 195348–195373. https://doi.org/10.1109/ACCESS.2020.3029280

Kaloudi, N. & Li, J. (2020). The AI-based cyber threat landscape: A survey. ACM Computing Surveys, 53(1), Article 20, 34 pages. https://doi.org/10.1145/3372823

Kaur, R., Gabrijelčič, D. & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. Information Fusion, 97, 101804. https://doi.org/10.1016/j.inffus.2023.101804

Kliegr, T., Bahník, Š. & Fürnkranz, J. (2021). A review of possible effects of cognitive biases on interpretation of rule-based machine learning models. Artificial Intelligence, 298, 103458. https://doi.org/10.1016/j.artint.2021.103458

Larriva-Novo, X., Vega-Barbas, M., Villagrá, V. A., Álvarez-Campana, M., Berrocal, J. & Rivera, D. (2020). Efficient distributed preprocessing model for machine learning-based anomaly detection over large-scale cybersecurity datasets. Applied Sciences, 10(10), 3430. https://doi.org/10.3390/app10103430

Liu, Y. & Chen, M. (2024). The knowledge structure and development trend in artificial intelligence based on latent feature topic model. IEEE Transactions on Engineering Management, 71, 12593.

Mirsky, Y., Demontis, A., Kotak, J., Shankar, R., Gelei, D., Yang, L., Zhang, X., Pintor, M., Lee, W., Elovici, Y. & Biggio, B. (2022). The threat of offensive AI to organizations. Computers & Security, 122, 103006. https://doi.org/10.1016/j.cose.2022.103006

Mukhopadhyay, A. & Jain, S. (2024). A framework for cyber-risk insurance against ransomware: A mixed-method approach. International Journal of Information Management, 74, 102724. https://doi.org/10.1016/j.ijinfomgt.2024.102724

Murugesan, S. (2023). The AI-cybersecurity nexus: The good and the evil. COLUMN: From the Editors. BRITE Professional Services, Sydney, NSW, Australia.

Osman, R. & El-Gendy, S. (2024). Interconnected and resilient: A CGE analysis of AI-driven cyberattacks in global trade. Risk Analysis. https://doi.org/10.1111/risa.14321

Palko, D., Babenko, T., Bigdan, A., Kiktev, N., Hutsol, T., Kuboň, M., Hnatiienko, H., Tabor, S., Gorbovy, O. & Borusiewicz, A. (2023). Cyber security risk modeling in distributed information systems. Applied Sciences, 13, 2393. https://doi.org/10.3390/app13042393

Russell, S. (2021). The history and future of AI. Oxford Review of Economic Policy, 37(3), 509–520. https://doi.org/10.1093/oxrep/grab013

Shih, W.-C., Yang, C.-T., Jiang, C.-T. & Kristiani, E. (2023). Implementation and visualization of a netflow log data lake system for cyberattack detection using distributed deep learning. The Journal of Supercomputing, 79(9), 4983–5012. https://doi.org/10.1007/s11227-022-04802-y

Schramm, S., Wehner, C. & Schmid, U. (2023). Comprehensible artificial intelligence on knowledge graphs: A survey. Web Semantics: Science, Services and Agents on the World Wide Web, 79, 100806. https://doi.org/10.1016/j.websem.2023.100806

Surianarayanan, C., Lawrence, J. J., Chelliah, P. R., Prakash, E. & Hewage, C. (2023). A survey on optimization techniques for edge artificial intelligence (AI). Sensors, 23(3), 1279. https://doi.org/10.3390/s23031279

Tanya, S. M., Nguyen, A. X., Buchanan, S. & Jackman, C. S. (2023). Development of a cloud-based clinical decision support system for ophthalmology triage using decision tree artificial intelligence. Ophthalmology Science, 3(1), 100231. https://doi.org/10.1016/j.xops.2022.100231

Voskoglou, M. G. (2020). Bayesian reasoning and artificial intelligence. Journal of Applied Mathematics and Computation, 17(12). https://doi.org/10.37394/232010.2020.17.12

Wang, Y., Xue, W. & Zhang, A. (2023). Application of big data technology in enterprise information security management and risk assessment. Journal of Global Information Management, 31(3). https://doi.org/10.4018/JGIM

Yamin, M. M., Ullah, M., Ullah, H. & Katta, B. (2021). Weaponized AI for cyberattacks. Journal of Information Security and Applications, 57, 102722. https://doi.org/10.1016/j.jisa.2020.102722

Zhang, Z., Damiani, E., Al Hamadi, H., Yeun, C. Y. & Taher, F. (2022). Explainable artificial intelligence applications in cybersecurity: State-of-the-art in research. IEEE Access, 10, 113394–113412. https://doi.org/10.1109/ACCESS.2022.3204051