TRAFICOM
Liikenne- ja viestintävirasto

Huoltovarmuuskeskus
Försörjningsberedskapscentralen
National Emergency Supply Agency

Tekoälypohjaiset kyberturvallisuusratkaisut

Traficomin tutkimuksia ja selvityksiä
07/2024

TRAFICOM
Liikenne- ja viestintävirasto

Huoltovarmuuskeskus
Försörjningsberedskapscentralen
National Emergency Supply Agency

Julkaisupäivämäärä
13.3.2024

Julkaisun nimi
Tekoälypohjaiset kyberturvallisuusratkaisut

Tekijät
Samuel Marchal, Bartosz Nawrotek, WithSecure

Toimeksiantaja
Liikenne- ja viestintävirasto Traficom

Julkaisusarjan nimi ja numero
Traficomin tutkimuksia ja selvityksiä
07/2024

ISSN (verkkojulkaisu) 2669-8781
ISBN (verkkojulkaisu) 978-952-311-917-8

Asiasanat
Tekoäly, koneoppiminen, kyberturvallisuus, tietoturva, suuret kielimallit, uhkien ennaltaehkäisy, uhkien tunnistaminen, riskienhallinta.

Tiivistelmä

Kyberturvallisuusalalla on jo yli kahden vuosikymmenen ajan hyödynnetty tekoälyteknologioita, mikä on merkittävästi tehostanut kyberuhkien torjuntaa. Tekoälyä on käytetty laajasti muun muassa roskapostin suodattamisessa, haittaohjelmien tunnistamisessa ja tunkeutumisen estämisessä, mikä on lisännyt kyberturvallisuusratkaisujen automaatiota, nopeutta, skaalautuvuutta ja sopeutumiskykyä. Vaikka suurin osa tekoälyn hyödyistä on nähty reaktiivisissa toimenpiteissä, on kiinnostusta herättänyt myös tekoälyn potentiaali myös ennaltaehkäisevissä toimissa, kuten uhkatiedustelussa ja turvallisuusriskien hallinnassa.

Tekoälyn soveltaminen kyberturvallisuusratkaisuissa on kuitenkin osoittautunut haastavaksi, ja monet onnistuneet sovellukset ovat syntyneet pitkällisten kokeilujen ja epäonnistumisten jälkeen. Onnistunut soveltaminen vaatii sekä tekoälyn että kyberturvallisuuden syvällistä ymmärrystä ja osaamista. Nykyinen tekoälybuumi on saanut monet organisaatiot pohtimaan, miten ne voisivat hyödyntää tekoälyä parantaakseen kyberturvallisuuttaan.

Selvitys tarjoaa kattavan katsauksen tekoälyn soveltamisesta kyberturvallisuuden parantamisessa ja paneutuu tekoälyn hyötyihin, haasteisiin ja tulevaisuuden kehitysmahdollisuuksiin. Tekoälyä hyödynnetään kyberturvallisuudessa laajaan sovellusten kirjoon, alkaen perinteisistä suodatus- ja tunnistusmekanismeista kehittyneempiin ennaltaehkäiseviin toimenpiteisiin. Vaikka tekoäly on edistynyt uhkien havaitsemisessa ja päätelaitteiden suojauksessa, sen soveltaminen esimerkiksi uhkatiedusteluun ja haavoittuvuuksien hallintaan on vielä kehitysvaiheessa. Selvityksessä muodostetaan kuvaa tulevaisuuden kehityskuluista ja potentiaalisista käyttötapauksista.

Selvityksen tulokset osoittavat, että tekoälyn onnistunut hyödyntäminen kyberturvallisuudessa vaatii monialaista osaamista, ja että organisaatioiden tulisi arvioida huolellisesti tekoälyyn pohjautuvien ratkaisujen soveltumista kriittisiin käyttötapauksiin. Lisäksi selvitys painottaa datan ymmärtämisen, saatavuuden ja laadun merkitystä, sekä prototyyppien testaamista todellisissa ympäristöissä ennen laajamittaisen käyttöönoton harkitsemista.

Uusien tekoälyteknologioiden, kuten suurten kielimallien, yleistyminen tarjoaa merkittäviä mahdollisuuksia kyberturvallisuussovellusten kehittämiselle. Näitä teknologioita voidaan hyödyntää muun muassa turvallisuuskoulutuksessa, analytiikassa ja uhkatiedustelussa, mahdollistaen laajojen datamäärien tehokkaan käsittelyn ja asiayhteyksien tunnistamisen. Tekoäly tuo mukanaan myös haasteita, kuten eettisiä kysymyksiä, teknisiä riskejä ja sääntelyyn liittyviä seikkoja, jotka kaikki vaativat huolellista harkintaa ja strategista lähestymistä.

Selvitys on toteutettu yhteistyössä Huoltovarmuuskeskuksen kanssa.

Yhteyshenkilö
Aleksi Blomqvist, Markus Mettälä

Raportin kieli
Suomi

Luottamuksellisuus
Julkinen

Kokonaissivumäärä
40

Jakaja
Liikenne- ja viestintävirasto Traficom, Kyberturvallisuuskeskus

Kustantaja
Liikenne- ja viestintävirasto Traficom, Kyberturvallisuuskeskus

TRAFICOM
Finnish Transport and Communications Agency

Huoltovarmuuskeskus
Försörjningsberedskapscentralen
National Emergency Supply Agency

Date of publication
13.3.2024

Title of publication
Applying artificial intelligence in cybersecurity

Author(s)
Samuel Marchal, Bartosz Nawrotek, WithSecure

Commissioned by
Finnish Transport and Communications Agency Traficom

Publication series and number
Traficom Research Reports 07/2024

ISSN (e-publication) 2669-8781
ISBN (e-publication) 978-952-311-917-8

Keywords
Artificial Intelligence, Machine Learning, Cybersecurity, Large Language Models, Threat prevention, Threat detection, Risk management.

Abstract

The cybersecurity industry has utilized Artificial Intelligence (AI) for over two decades, applying it across various domains such as spam filtering, malware detection, and intrusion detection to enhance performance through automation, speed, scalability, and adaptability. While AI's impact has been predominantly in reactive cybersecurity measures, new AI technologies are promising for proactive security efforts, including advanced threat intelligence, security risk management, and heightened security awareness.

Nevertheless, the path to successfully using AI for cybersecurity is paved with many pitfalls, and successful applications have been developed at the cost of many failures. This success requires advanced knowledge, skills and experience in both AI and cybersecurity, which a few specialized organizations have been able to gather and reap the benefits from. In the context of the current AI hype, a widespread interest has risen into exploring the possible applications of AI, including those to cybersecurity. Many organizations seek to know if and how they could use AI to improve their security posture. This turns out to be challenging without field experience, considering the shortage of skilled AI and security experts.

This report aims to provide organizations with insights into AI's capabilities and potential benefits for cybersecurity, detailing existing applications, their maturity levels, and associated challenges. AI's effectiveness in improving threat detection and endpoint security is noted, though its applications in threat intelligence and vulnerability management remain nascent. The report emphasizes the importance of a meticulous development process and the integration of AI into security measures, underscoring the necessity of aligning AI solutions with business objectives, thorough understanding of data, and testing early prototypes in real-world settings. Developing cross-competency among experts in AI and cybersecurity is crucial for success.

Looking ahead, emerging AI technologies like Large Language Models (LLMs) are set to revolutionize cybersecurity applications by supporting security education, analytics, threat intelligence, and vulnerability management. These models promise to enhance the processing and correlation of information from vast amounts of unstructured data, potentially leading to more autonomous and complex task management. Nonetheless, as AI applications evolve, they will encounter new ethical, technical, and regulatory challenges that could impede progress.

The report was conducted in collaboration with the National Emergency Supply Agency.

Contact person	**Language**	**Confidence status**	**Pages, total**
Aleksi Blomqvist, Markus Mettälä	Finnish	Public	40

Distributed by	**Published by**
Transport and Communications Agency, National Cyber Security Centre Finland	Finnish Transport and Communications Agency Traficom, National Cyber Security Centre Finland

TRAFICOM
Transport- och kommunikationsverket

Huoltovarmuuskeskus
Försörjningsberedskapscentralen
National Emergency Supply Agency

Utgivningsdatum
13.3.2024

Publikation
Cybersäkerhetslösningar baserade på artificiell intelligens

Författare
Samuel Marchal, Bartosz Nawrotek, WithSecure

Tillsatt av och datum
Transport- och kommunikationsverket Traficom

Publikationsseriens namn och nummer
Traficoms forskningsrapporter och utredningar 07/2024

ISSN (elektronisk publikation) 2669-8781
ISBN (elektronisk publikation) 978-952-311-917-8

Ämnesord
Artificiell intelligens, maskininlärning, cybersäkerhet, informationssäkerhet, stora språkmodeller, förebyggande av hot, identifiering av hot, riskhantering.

Sammandrag

Inom cybersäkerhetsområdet har man redan i över två årtionden använt sig av tekniker med artificiell intelligens, vilket i hög grad har effektiviserat bekämpningen av cyberhot. Artificiell intelligens har använts i stor utsträckning vid bland annat filtrering av skräppost, identifiering av skadliga program och förhindrande av intrång, vilket har ökat automatiseringen av samt snabbheten, skalbarheten och anpassningsförmågan hos cybersäkerhetslösningar. Även om största delen av fördelarna med artificiell intelligens har setts vid reaktiva åtgärder, har den artificiella intelligensens potential väckt intresse även vid förebyggande åtgärder, såsom underrättelser om hot och hantering av säkerhetsrisker.

Tillämpning av artificiell intelligens i cybersäkerhetslösningar har dock visat sig vara utmanande, och många framgångsrika applikationer har uppkommit efter långvariga försök och misslyckanden. En lyckad tillämpning kräver en djup förståelse av och kompetens i såväl artificiell intelligens som cybersäkerhet. Det nuvarande uppsvinget för artificiell intelligens har fått många organisationer att fundera över hur de skulle kunna utnyttja artificiell intelligens för att förbättra sin cybersäkerhet.

Utredningen ger en heltäckande översikt över tillämpningen av artificiell intelligens vid förbättring av cybersäkerheten och fördjupar sig i fördelarna med samt utmaningarna och de framtida utvecklingsmöjligheterna för artificiell intelligens. Artificiell intelligens utnyttjas inom cybersäkerhet för ett brett spektrum av applikationer, allt från traditionella filtrerings- och identifieringsmekanismer till mer avancerade förebyggande åtgärder. Även om den artificiella intelligensen har gjort framsteg i att upptäcka hot och skydda terminalutrustning, är tillämpningen av den för exempelvis underrättelse om hot och hantering av sårbarheter ännu under utveckling. I utredningen skapas en bild av utvecklingsgången och potentiella användningsfall i framtiden.

Resultaten av utredningen visar att ett framgångsrikt utnyttjande av artificiell intelligens inom cybersäkerhet kräver tvärvetenskapligt kunnande och att organisationerna noggrant borde överväga tillämpning av lösningar baserade på artificiell intelligens för kritiska användningsfall. Dessutom understryker utredningen vikten av att förstå data, tillgången till och kvaliteten på data samt testning av prototyper i verkliga miljöer innan ett storskaligt ibruktagande övervägs.

Det faktum att nya tekniker med artificiell intelligens, såsom stora språkmodeller, blir allt vanligare innebär stora möjligheter att utveckla cybersäkerhetsapplikationer. Dessa tekniker kan utnyttjas inom bland annat säkerhetsutbildning, analys och underrättelse om hot, vilket möjliggör effektiv behandling av stora datamängder och identifiering av kontexter. Artificiell intelligens innebär även utmaningar, såsom etiska frågor, tekniska risker och frågor om reglering, vilka alla kräver noggrant övervägande och ett strategiskt angreppssätt.

Utredningen har gjorts i samarbete med Försörjningsberedskapscentralen.

Kontaktperson	**Språk**	**Sekretessgrad**	**Sidoantal**
Aleksi Blomqvist, Markus Mettälä	Finska	Offentlig	40

Distribution	**Förlag**
Transport- och kommunikationsverket Traficom, Cybersäkerhetscentret	Transport- och kommunikationsverket Traficom, Cybersäkerhetscentret

Sisällysluettelo

Kuvaajat

Taulukot

Lyhenteet

C&C	Command-and-Control, komentopalvelin
CDR	Cloud Detection & Response
CVE	Common Vulnerabilities and Exposure, yleiset haavoittuvuudet ja paljastuneet tietoturvapuutteet
DNN	Deep Neural Networks, syvät neuroverkot
EDR	Endpoint Detection & Response
IoT	Internet of Things, esineiden internet
LIME	Local Interpretable Model-agnostic Explanations
PE	Portable Executable, tiedostomuoto suoritettaville ohjelmille
SIEM	Security Information and Event Management, Tietoturvatapahtumien ja -tietojen hallintajärjestelmä
SOC	Security Operation Centre, tietoturvanhallintakeskus
UEBA	User and Entity Behaviour Analytics, käyttäjän ja entiteetin käyttäytymisanalyysi

1. Tekoäly kyberturvallisuuden edistäjänä

Tekoäly, koneoppiminen, syväoppiminen ja generatiivinen tekoäly ovat trendikkäitä, monitulkintaisia ja osittain päällekkäisiä aloja, joiden ymmärtäminen edellyttää niiden eroavaisuuksien selkeyttämistä. Edellä mainituissa tutkimuskohteissa on otettu viime aikoina merkittäviä edistysaskeleita, joiden myötä on syntynyt uusia sovelluksia, joita organisaatiot voivat hyödyntää kyberturvallisuuden edistämisessä. Tekoälypohjaiset ratkaisut kykenevät automatisoimaan ja tehostamaan prosesseja sekä tukemaan tietoturva-asiantuntijoiden työtä.

1.1 Mitä tekoäly on?

Tekoäly (artificial intelligence, AI) viittaa kykyyn, jolla kone, kuten tietokonejärjestelmä tai -ohjelma, voi suorittaa tehtäviä, joita on aiemmin pidetty mahdollisena vain ihmisille tai eläimille. Tekoälyjärjestelmien ”älykkyys” viittaa kykyihin päättää, ratkaista ongelmia, löytää merkityksiä, yleistää, suunnitella ja oppia kokemuksesta. Tekoäly on kuitenkin laaja käsite, joka kattaa monia eri osa-alueita, kuten asiantuntijajärjestelmät, robotiikka ja sumea logiikka (*fuzzy logic*).

Useimmat nykyiset tekoälyn osa-alueet eivät edusta lähelläkään ihmisen tasoista älykkyyttä, eivätkä ne pysty organisoimaan tai käynnistämään kyberhyökkäyksiä automaattisesti. Koneoppimista hyödyntävät sovellukset pystyvät kuitenkin suoriutumaan useista tehtävistä, jotka on yhdistetty ihmisen älykkyyteen. Tällaisia tehtäviä ovat esimerkiksi kuvien luokittelu, tekstin kääntäminen, ja shakin, gon ja muiden vastaavien pelien pelaaminen. Koneoppiminen on saanut viime aikoina paljon huomiota siinä otettujen valtavien edistysaskeleiden vuoksi. Suurin osa nykyisestä tekoälyä koskevasta julkisesta keskustelusta liittyykin koneoppimisen sovelluksiin, ja termiä tekoäly käytetään usein lyhenteenä kuvaamaan koneoppimista.

Koneoppiminen (machine learning, ML) on termi, jota käytetään kuvaamaan sellaista asiantuntijajärjestelmää, joka käyttää dataa oppimiseen, päätöksentekoon, ja itsensä kehittämiseen ilman, että järjestelmän tarvitsee seurata tarkkoja ohjeita. Järjestelmä käyttää algoritmeja ja tilastollisia malleja datan analysointiin ja johtopäätösten tekemiseen. Koneoppiminen eroaa useimmista tekoälyn osa-alueet, jotka vaativat selkeitä käskyjä tai sääntöjä tulosten tuottamiseksi. Toisin kuin muut tekoälyn osa-alueet, koneoppiminen käyttää sopeutuvia algoritmeja, jotka sopeuttavat käyttäytymisensä itsenäisesti datan perusteella. Koneoppiminen on jaettu kolmeen päätyyppiin:

1) valvottuun oppimiseen, joka on suunniteltu suorittamaan tai toistamaan tunnettuja tehtäviä (*tehtävävetoisuus*),

2) valvomattomaan oppimiseen, joka on suunniteltu poimimaan datasta tuntematonta tietoa (*datavetoisuus*),

3) vahvistusoppimiseen, joka on suunniteltu uusien tehtävien oppimiseen kokeilu- ja virheprosessin kautta samalla, kun määriteltyä lopputulosta yritetään maksimoida (*kokeilu- ja virhevetoisuus*).

Syväoppiminen (deep learning, DL) tai **syvät neuroverkot (deep neural networks, DNNs)** ovat koneoppimisalgoritmeja, joilla on korkea suorituskyky luonnollisen datan automaattisessa käsittelyssä. Luonnollista dataa on esimerkiksi teksti, kuvat, ääni ja video. Syväoppimisen viimeaikaiset edistysaskeleet ovat pääasiallinen syy nykyiseen tekoälyyn ja koneoppimiseen liittyvään innostukseen. Syväoppimistekniikat ovat saavuttaneet vertaansa vailla olevan suorituskyvyn monimutkaisissa tehtävissä, kuten kuvien luokittelussa, tekstin kääntämisessä tai monimutkaisten pelien pelaamisessa. Ne osaavat päätellä, ratkaista ongelmia, löytää merkityksiä, ja oppia itsenäisesti kokemuksen avulla käyttäen pelkkää dataa. Keskeisiä mahdollistajia syväoppimisen viimeaikaiselle kehitykselle ovat olleet algoritmiparannukset, suurten tietojoukkojen saatavuus ja halpa laskentateho.

Generatiivinen tekoäly (generative AI, GenAI) viittaa koneoppimisalgoritmiin, joka luo tai tuottaa uutta sisältöä, olipa kyseessä sitten kuvat, teksti, musiikki tai videot. Se ei pelkästään nouda tietoa tai tee päätöksiä ennalta määriteltyjen valintojen perusteella, vaan tuottaa annetun datan pohjalta kokonaan uutta dataa oppimansa kaavan perusteella.

Suuret kielimallit (large language model LLM) keskittyvät tekstien tuottamiseen. Ne kuuluvat yhteisen generatiivisen tekoälyn, syväoppimisen ja luonnollisen kielenkäsittelyn (*natural language processing, NLP*) käsitteiden alle. Suuret kielimallit on koulutettu valtavilla määrillä tekstidataa ymmärtämään kielen kaavoja ja tuottamaan ihmismäistä tekstiä. Suuret kielimallit voivat täydentää lauseita, luoda tarinoita, vastata kysymyksiin ja koodata oppimiensa kaavojen perusteella. Nämä mallit ovat tähän mennessä lähin esimerkki tekoälystä, joka vastaa hypoteettista yleistä tekoälyä (*artificial general intelligence, AGI*). Yleisellä tekoälyllä on ihmisen tapaan kyky ymmärtää, oppia, sopeutua ja soveltaa älykkyyttään laajalla tehtäväalueella tai -alueilla.

Tässä selvityksessä käytetään pääasiassa termiä tekoäly, ja sillä viitataan johdonmukaisesti koneoppimiseen. Tämä vastaa sen yleistä käyttöä valtavirran viestinnässä.

1.2 Keskeiset tekoälykyvykkyydet

Tekoälyn ja koneoppimisen kyvykkyyksiä voidaan soveltaa useisiin eri käyttötapauksiin. Koneoppimistekniikoita voidaan räätälöidä ja soveltaa tarkkojen ennusteiden, luokittelujen tai näkemysten tuottamiseksi, ongelmanasettelun tarkkuuden ja datan saatavuuden asettamissa rajoissa.

Koneoppimistekniikoita voidaan soveltaa kyberturvallisuuden prosessien automatisointiin. Tekoälysovellukset keskittyvät haitallisen sisällön, tapahtumien tai käyttäytymisen tunnistamiseen ja erottamiseen normaalista toiminnasta:

- **Luokittelu (classification)** on tehtävä, jossa objekti (tai syöteaineisto) sijoitetaan ennalta määriteltyihin luokkiin tai kategorioihin. Luokka opitaan aiemmin havaitusta datasta, ja sitä voidaan soveltaa uuteen tuntemattomaan dataan. Luokittelu on laajalti käytössä kyberturvallisuudessa haitallisen sisällön havaitsemiseksi haittaohjelmista sekä tietojenkalastelu- ja roskapostisähköposteista. Luokittelu edellyttää esimerkkejä sekä tavanomaisesta että haitallisesta sisällöstä, jotka halutaan erottaa toisistaan.

- **Poikkeamien havaitseminen (anomaly detection)** on tehtävä, jossa tunnistetaan kaavat tai tapaukset, jotka poikkeavat merkittävästi siitä, mikä katsotaan normaaliksi tai odotetuksi järjestelmässä tai tietoaineistossa. Se toimii määrittelemällä normaalin käyttäytymisen rajat, ja merkitsemällä kaikki datapisteet tai tapahtumat, jotka poikkeavat normaalista käyttäytymisestä. Poikkeamat ovat jatkuvasti valvottuja, ja kaikki poikkeamat merkitään mahdollisiksi kyberturvallisuusuhkiksi.

- **Käyttäytymisanalyysi (behavioural analysis)** on lähellä poikkeaman havaitsemista. Käyttäytymisanalyysissa kartoitetaan tyypilliset käyttäytymismallit järjestelmässä, verkossa tai käyttäjän toiminnassa. Toisin kuin poikkeaman havaitseminen, käyttäytymisanalyysi ei keskity

pelkästään anomalioiden tunnistamiseen. Sen tavoitteena on sen sijaan ymmärtää järjestelmän sisällä olevien yksiköiden tavallista tai odotettua käyttäytymistä. Tietoturvallisuudessa käyttäytymisanalyysiin kuuluu usein normaalin käyttäytymisen profiilien luominen käyttäjille, laitteille tai järjestelmille historiallisen datan, käyttäjän toimien ja järjestelmän toimintojen perusteella. Poikkeamien tai muutosten tunnistaminen vakiintuneista profiileista voi viitata mahdollisiin kyberturvallisuusriskeihin, kuten sisäpiirin uhkiin tai luvattomiin pääsyihin.

- **Hahmontunnistuksella (pattern recognition)** tarkoitetaan merkityksellisiä toimintamalleja kuvaavien kaavojen automaattista löytämistä tietoaineistosta. Se antaa järjestelmille kyvyn tunnistaa ja tulkita tietoa, ominaisuuksia tai piirteitä, jotka toistuvat tai ovat samankaltaisia. Kaavojen tunnistaminen on kyberturvallisuusratkaisuissa pääasiassa käytössä ominaisuuksien erottelussa, missä poimitaan keskeiset ominaisuudet datasta edustamaan kaavoja. Kyberturvallisuudessa kaavat edustavat yleensä haitallista sisältöä ja käyttäytymistä, joista voidaan erottaa toiminnalle tyypilliset tunnusmerkit. Tunnusmerkkejä käytetään uusien kaavojen tunnistamiseen ja tunnettujen tietojenkalasteluyritysten, haittaohjelmatartuntojen, haitallisen käyttäjäkäyttäytymisen ym. tunnistamiseen.

  Tekoälyä voidaan soveltaa kyberturvallisuusasiantuntijoiden avustamiseen. Sovellusten tavoitteena on virtaviivaistaa päätöksentekoprosesseja ja mahdollistaa asiantuntijoiden keskittyminen osaamista vastaaviin ja merkittäviin työtehtäviin.

- **Ryhmittely (clustering)** on koneoppimistekniikka, mitä käytetään samankaltaisten datapisteiden luokitteluun niiden ominaisuuksien tai piirteiden perusteella. Ryhmittelyn tavoitteena on löytää luonnollisia ryhmittymiä tietojoukosta ilman ennalta määriteltyjä tunnisteita tai luokkia. Ryhmittelyä voidaan käyttää uhkatiedustelun suuren tietomäärän analysointiin ja luokitteluun tai erilaisten haittaohjelmaversioiden tunnistamiseen. Sitä voidaan myös yhdistää käyttäytymisanalyysiin erilaisten käyttäjä- tai järjestelmäprofiilien ryhmittelyyn ja tunnistamiseen.

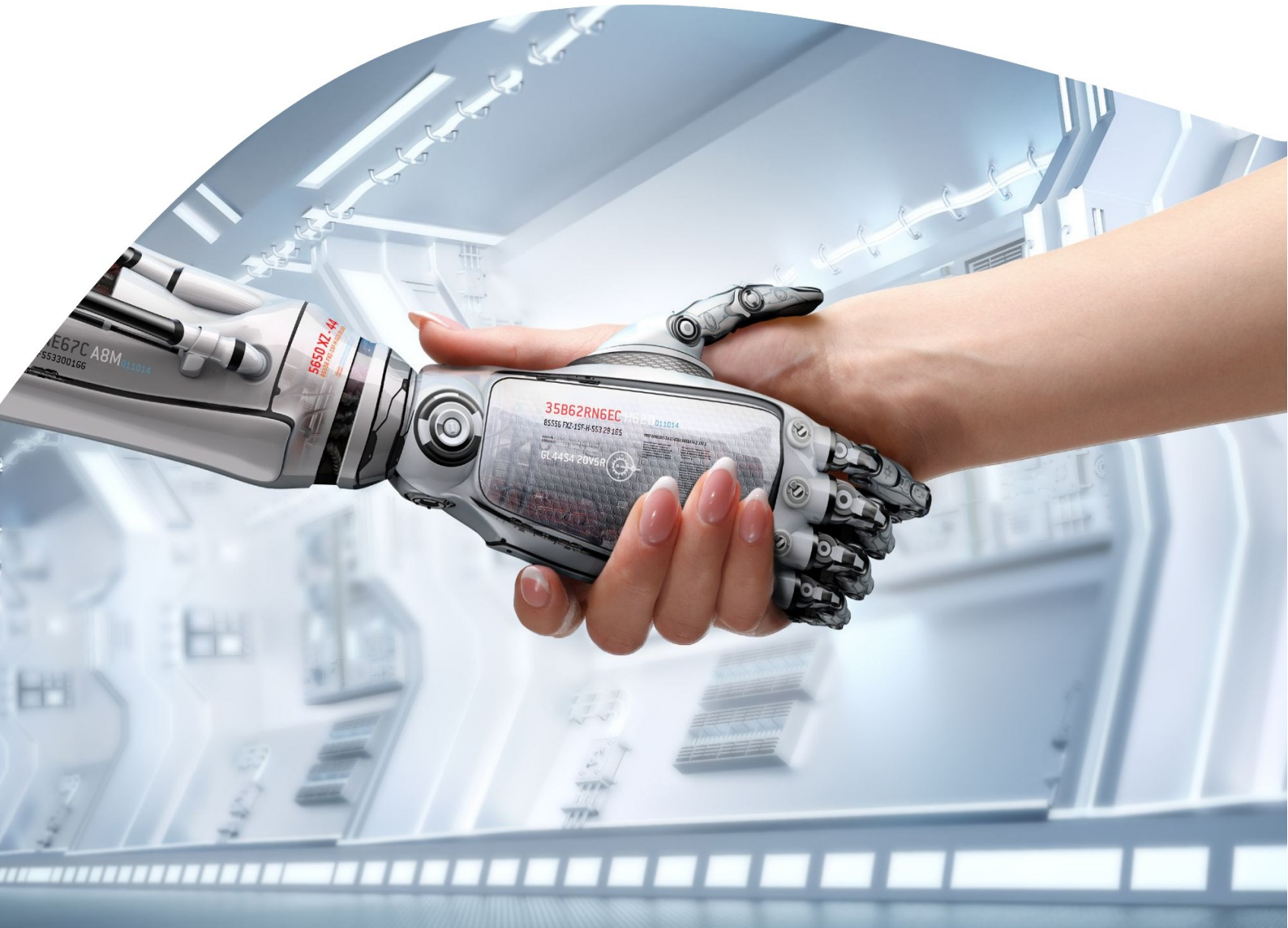
- **Tiedonhaulla (information retrieval)** louhitaan ja poimitaan informaatiota sisällöstä, joka vastaa tai on semanttisesti samankaltainen annetun kyselyn kanssa. Tiedonhakua voidaan käyttää uhkatiedustelussa. Tekoälyalgoritmit kykenevät käsittelemään, analysoimaan, tiivistämään ja luokittelemaan suuria määriä uhkatiedusteludataa eri lähteistä auttaen turvallisuustiimejä tunnistamaan ja vastaamaan uusiin uhkiin tehokkaammin.

- **Järjestäminen (ranking)** tarkoittaa joukon kohteiden lajittelua tiettyyn järjestykseen niiden merkityksen, tärkeyden tai todennäköisyyden perusteella. Koneoppimisen järjestelmä oppii luomaan järjestelymallin käsittelemällä luokiteltua tai epäsuorasti lajiteltua dataa. Järjestämistä käytetään yleensä kyberturvallisuusasiantuntijoiden tehtävien priorisointiin. Järjestämistä voidaan käyttää esimerkiksi haavoittuvuuksien korjaamisen priorisointiin haavoittuvuushallinnassa, sekä tapahtumien luokitteluun ja käsittelyyn tietoturvatapahtumien ja -tietojen hallintajärjestelmissä (SIEM-järjestelmissä).

- **Generointi (generation)** on toiminto, jossa luodaan uutta sisältöä, joka vastaa aiemmin määriteltyä kohdejakaumaa. Generointia voidaan käyttää kyberturvallisuusratkaisuissa haavoittuvuusarvioinnissa ja penetraatiotestauksessa. Tekoälypohjaisia generointityökaluja voidaan käyttää hyökkäysten simuloimiseen ja haavoittuvuuksien tunnistamiseen järjestelmissä tai tietoverkoissa. Tämä auttaa organisaatioita löytämään haavoittuvuuksia ennen niiden hyödyntämistä ja vahvistamaan puolustustaan ennakoivasti.

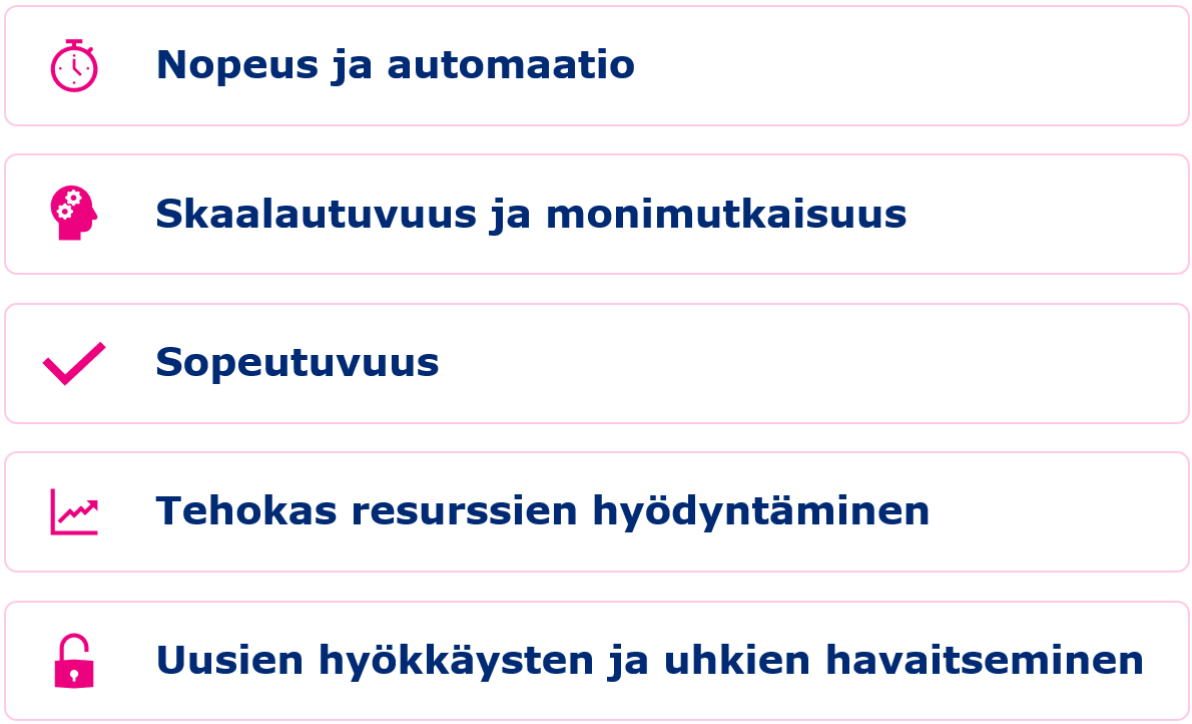
2. Tekoälyn käytössä huomioitavat asiat

Kyberturvallisuusosaajilla on vielä rajallinen tuntemus tekoälystä ja sen hyödyistä. Tavoiteltujen hyötyjen saavuttamiseksi organisaatioiden tulee tiedostaa tekoälyn soveltamisen erityispiirteet. Esimerkkejä huomioitavista seikoista ovat reaaliaikainen vaste, ympäristön dynaamisuus, vihamielisille toimijoille tarjoutuvat mahdollisuudet, kompromissit käytettävyydessä ja turvallisuudessa, ymmärrettävyys ja yksityisyydensuoja.



2.1 Tekoälyn tuomat hyödyt

Tekoälypohjaisten ratkaisuiden käyttöönotto yksittäisen kyberturvallisuusongelman ratkaisemiseksi ei ole mutkatonta. Kyberturvallisuusasiantuntijoilla voi olla rajallinen tuntemus tekoälyn tekniikoista ja niiden hyödyistä. Tämän takia perinteinen ongelmanratkaisu perustuu ihmisten asiantuntemukseen ja manuaalisiin tehtäviin, kunnes ne osoittautuvat riittämättömiksi. Organisaatiot valitsevat tekoälyn usein tavoitellessaan kuvaajassa 1 esitettyjä etuja.



Kuvaaja 1: Tekoälyn hyödyt

Nopeus ja automaatio. Kyberhyökkäykset voivat tapahtua muutamissa sekunneissa ja niiden vaikutusten minimoimiseksi tarvitaan nopeaa reagointia. Digitaaliset ympäristöt tuottavat valtavan määrän dataa, kuten tiedostoja, turvallisuustapahtumia, turvallisuushälytyksiä ym., jotka ovat oleellisia turvallisuuden ylläpidossa. Tekoäly mahdollistaa suuren ja monipuolisen datamäärän prosessoinnin ja analyysin reaaliajassa. Se voi tiivistää, tarkkailla ja havaita malleja valtavista, monimutkaisista tietoaineistoista hyvin nopeasti. Tekoälyyn pohjautuva automaatio voi mahdollistaa aiempaa nopeammat ja tehokkaammat toimet tietomurtojen rajaamiseksi ja lieventämiseksi sekä niistä palautumiseksi.

Skaalautuvuus ja monimutkaisuus. Nykyaikaiset digitaaliset ympäristöt ovat monimutkaisia ja alttiita lukuisille uhille. Ne tuottavat valtavia määriä turvallisuuteen liittyvää dataa, joka pitää analysoida mahdollisimman nopeasti. Tekoäly mahdollistaa skaalautuvuuden tietoturvatoimissa käsittelemällä dataa tehokkaasti. Uhkien määrän kasvaessa sääntöjä ei enää ole mielekästä määritellä manuaalisesti. Tekoäly kykenee automatisoimaan sääntöjen luomista, ja yksittäiseen tekoälymalliin voi sisällyttää potentiaalisesti tuhansia epäsuorasti ja automaattisesti opittuja sääntöjä. Tekoälyä voidaan hyödyntää myös ratkaisemaan

ongelmia, jotka ovat samanlaisia luonteeltaan, mutta vaihtelevia ilmenemismuodoltaan. Tekoäly mahdollistaa esimerkiksi räätälöidyn käyttäytymisanalyysin käyttäjistä, laitteista tai tietoverkoista tunnistamalla poikkeamia tai epätavallisia malleja, jotka voivat olla merkkejä mahdollisista uhkista. Mukauttamisessa voidaan huomioida ryhmät muun muassa roolin, organisaation tai sijainnin perusteella.

Sopeutumiskyky. Kehittyvä uhkaympäristö on kyberturvallisuuden ominaispiirre. Uhat, haavoittuvuudet ja hyökkäykset muuttuvat nopeasti ajan myötä, ja vaativat puolustukselta yhtä nopeaa sopeutumista. Havaitsemisen sääntöjä pitää määritellä uudelleen, sekä tunnistaa uusien haittaohjelmien tunnusmerkit ja hyökkäysten toimintamallit. Tekoälymallit voivat oppia nopeasti uudesta datasta automaattisen uudelleenkoulutuksen avulla, mikä ei vaadi muutoksia tekoälysovelluksen malliin tai ominaisuuksiin. Ne voivat sopeutua kehittyviin hyökkäyskuvioihin, mikä parantaa niiden omaa kykyä havaita ja ennaltaehkäistä uhkia. Säännöllisten harjoittelusyklien avulla mallit voivat oppia jatkuvasti kehittyvästä näytekirjastosta, joka sisältää analyytikkojen varmistamia havaintoja ja hälytyksiä. Tämä vähentää virheiden uusiutumista ja mahdollistaa mallien oppimisen ja asiantuntijatiedon hyödyntämisen.

Tehokas resurssien hyödyntäminen. Turvallisuusasiantuntemus on niukkaa ja tekoäly mahdollistaa arkipäiväisten toistuvien tehtävien automatisoinnin. Tämä vapauttaa asiantuntijoille lisää aikaa strategisiin ja monimutkaisiin turvallisuustoimiin. Tekoäly pystyy esimerkiksi analysoimaan ja yhdistelemään suuria määriä historiallista ja dynaamista uhkatietoa mahdollistaen datan käyttämisen useista lähteistä lähes reaaliajassa. Se mahdollistaa turvallisuusanalyytikoiden ajan ja osaamisen tehokkaan priorisoinnin kriittisten haavoittuvuuksien käsittelemiseksi sekä aikaherkkien turvallisuusvaroitusten ja havaintojen tutkimiseksi. Tekoäly voi myös vähentää hälytysväsymystä havaitsemis- ja vastetoimintatilanteissa korjaamalla virheitä ja järjestämällä turvallisuustapahtumat tärkeysjärjestykseen.

Uusien hyökkäysten ja uhkien löytäminen. Poikkeuksien havaitsemisen ja käyttäytymisanalyysin kaltaisten kykyjen avulla tekoälytekniikoilla on potentiaalia löytää uusia ja nousevia uhkia. Tekoälymallien ei tarvitse määritellä uhkaa tai hyökkäystä sen havaitsemiseksi, vaan ne voivat tunnistaa mahdolliset uudet hyökkäykset mallintamalla normaaleja käyttäytymismalleja ja havaitsemalla poikkeamia. Ne voivat ennaltaehkäistä hyökkäyksiä ja auttaa turvallisuustiimejä uusien uhkien tunnistamisessa ja lieventämisessä. Uusien uhkien löytämistä rajoittaa kuitenkin häiriöiden yleisyys, sillä useimmat poikkeamat eivät liity haitalliseen käyttäytymiseen.

2.2 Erityispiirteet ja haasteet tekoälyn soveltamiselle kyberturvallisuusratkaisuissa

"Tekoälymallin tekemän päätöksen ymmärtäminen on usein tarpeen kyberturvallisuudessa. Ymmärrettävyys luo luottamusta ja lisää päätösten läpinäkyvyyttä tietoturva-asiantuntijoille, jotka perustavat toimintaansa tekoälyn käsittelemään tietoon."

Tekoälypohjaisissa kyberturvallisuusratkaisuissa on omat erityispiirteensä ja haasteensa. Erityispiirteiden tunnistaminen auttaa suunnittelemaan ja toteuttamaan tekoälyratkaisuja tehokkaasti samalla, kun hallitaan mahdollisia riskejä ja varmistetaan kyberturvallisuussovelluksien luotettavuus.

Vasteen reaaliaikaisuus. Kyberhyökkäykset voivat aiheuttaa merkittävää vahinkoa nopeasti. Reaaliaikainen toiminta uhkien havaitsemisen jälkeen voi estää ne tai ainakin minimoida niiden vaikutuksia. Nopeaa automatisointia tarvitaan uhkien lieventämiseen, havaitsemiseen ja poikkeamien hallintaan. Kyberturvallisuuden edistämisessä käytettävien tekoälymallien on oltava tehokkaita, skaalautuvia ja kykeneviä toimimaan reaaliajassa. Tämä edellyttää mallien optimointia nopeaa päättelyä varten, laskennallisen ylikuormituksen vähentämistä ja sitä, että ennusteet ovat tarkkoja, ajankohtaisia, eivätkä vaaranna turvallisuutta.

Muuttuva uhkaympäristö. Kyberturvallisuusympäristö on dynaaminen. Uhkaympäristö kehittyy jatkuvasti uhkatoimijoiden kehittäessä uusia hyökkäystekniikoita ja haittaohjelmaversioita hyödyntääkseen haavoittuvuuksia. Myös tavanomainen toiminta on luonteeltaan muuttuvaa, ja vaihtelee järjestelmien kokoonpanojen ja käyttäjien toiminnan mukaan. Kyberturvallisuuden tekoälymallien on oltava ketteriä, kestäviä ja kykeneviä oppimaan ja sopeutumaan jatkuvasti muuttuvaan uhkaympäristöön. Tämä edellyttää jatkuvaa seurantaa, päivitystä ja uudelleenkoulutusta, jotta tekoälymallien tehokkuus uhkien havaitsemisessa ja uusiin uhkiin vastaamisessa säilyy.

Vastustajien kekseliäs luonne. Kyberturvallisuusratkaisuissa tulee olettaa, että vastustaja on aina läsnä. Hyökkääjät kehittävät jatkuvasti tekniikoitaan puolustuksen ja turvallisuustoimien ohittamiseksi. Hyökätessään tekoälypohjaista puolustusta vastaan, hyökkääjät voivat käyttää erityisesti tekoälymallien huijaamiseen tarkoitettuja hyökkäyksiä, kuten mallin kiertämistä ja saastuttamista. Vastustajan luonteen huomioiminen edellyttää kattavaa ymmärrystä mahdollisista uhista ja ennaltaehkäisevien puolustusstrategioiden käyttöä tekoälymallien vahvistamiseksi. Kyberturvallisuusratkaisuissa käytettävien tekoälymallien on kyettävä vastustamaan myös sellaisia hyökkäysmuotoja, joissa hyökkääjät manipuloivat tietoja tarkoituksellisesti tekoälyjärjestelmien harhauttamiseksi.

Virheistä koituvat kustannukset. Turvallisuusratkaisujen tehokkuus on usein ristiriidassa niiden suojaamien järjestelmien käytettävyyden kanssa. Virheet, kuten väärät positiiviset tulokset (normaalin käyttäytymisen merkitseminen pahantahtoiseksi) ja väärät negatiiviset (todellisten uhkien havaitsematta jääminen), vaikuttavat merkittävästi käytettävyyteen ja turvallisuuteen. Näiden virheiden vaikutukset ja kustannukset riippuvat asiayhteydestä ja niiden suojaamien järjestelmien kriittisyydestä. Molempien virhetyyppien minimoiminen ja järkevän kompromissin saavuttaminen vaatii tasapainottelua turvallisuuden ja käytettävyyden välillä. Kyberturvallisuuden tekoälymallien tarkkuustaso (esimerkiksi täsmällisyys, palautuminen, väärien positiivisten määrä ynnä muut tekijät) on pystyttävä säätämään tukemaan käytettävyyden ja turvallisuuden tasapainoa käyttötapauksen vaatimusten mukaisesti.

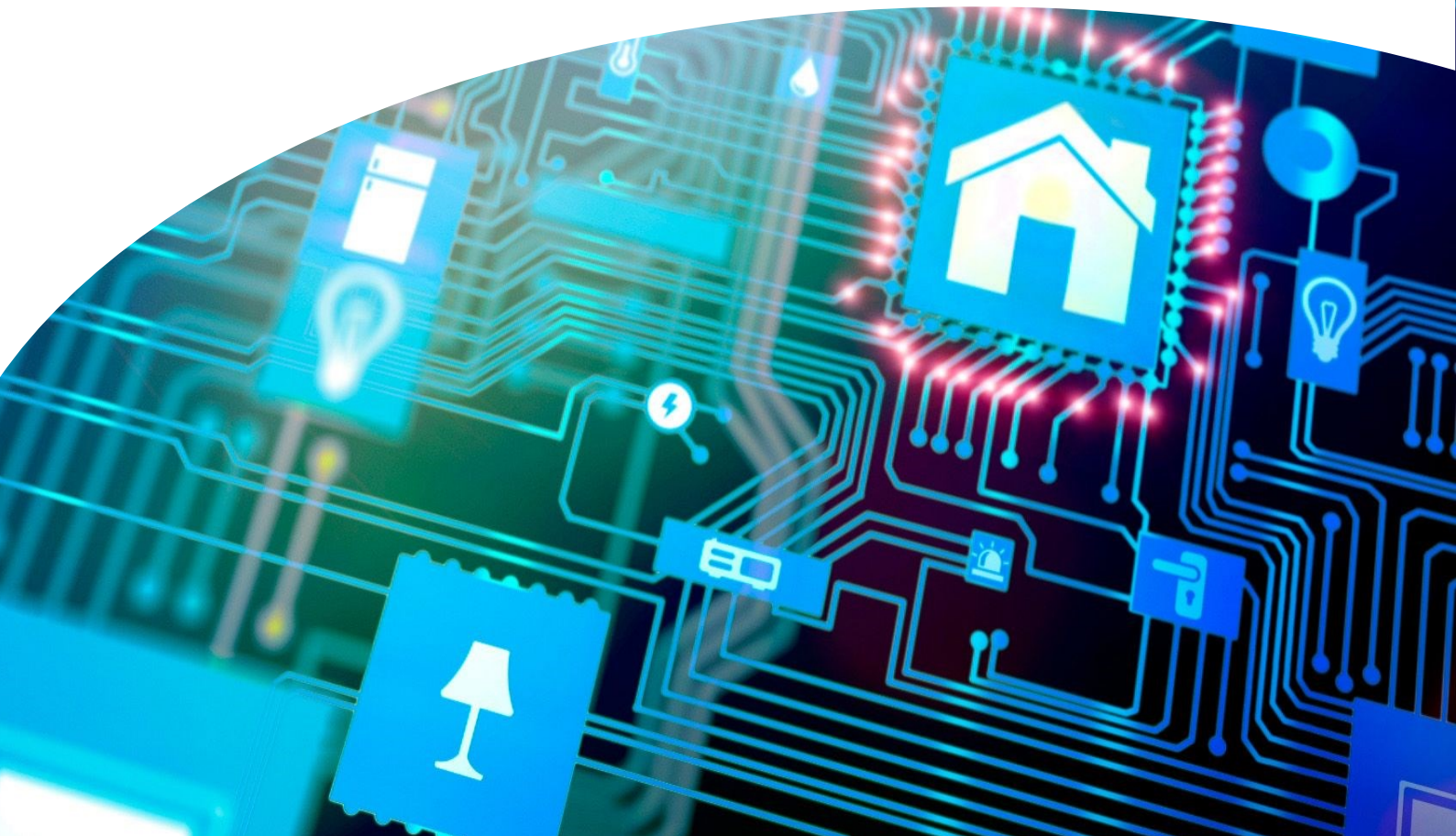
Dataan liittyvät haasteet. Kyberturvallisuudessa hyödynnettävät tietoaineistot ovat usein epätasapainossa siten, että tavanomaista käyttäytymistä edustavaa dataa on pahantahtoista toimintaa edustavaa dataa enemmän. Poikkeamien havaitsemista ja uhkien tunnistamista vaikeuttaa luokitellun datan puuttuminen tai

tunnettujen hyökkäysten vähäinen määrä. Kyberturvallisuusratkaisujen tekoälymallien on tehtävä yleistyksiä hyvin rajallisen aineiston pohjalta ja otettava huomioon aineiston epätasapaino. Tämä voidaan osittain ratkaista tekoälyn mallin valinnassa (esim. *ensemble-menetelmät*) ja mallin koulutusmenetelmässä (esimerkiksi *osittain valvottu* tai *aktiivinen oppiminen*) datamäärän kasvattamisen yhteydessä.

Tulkittavuus ja ymmärrettävyys. Tekoälymallin tekemän päätöksen ymmärtäminen on usein tarpeen kyberturvallisuudessa. Ymmärrettävyys luo luottamusta ja lisää päätösten läpinäkyvyyttä tietoturva-asiantuntijoille, jotka perustavat toimintaansa tekoälyn käsittelemään tietoon. Päätösten ymmärrettävyys auttaa tunnistamaan ongelman juurisyyn, ymmärtämään hyökkäysvektoreita ja kehittämään tehokkaita vastatoimenpiteitä. Läpinäkyvyyttä tulee edistää käyttämällä tulkittava tekoälymalleja tai selitettäviä menetelmiä (esimerkiksi *SHA*, *LIME*) yhdessä dokumentoinnin ja raportoinnin kanssa.

Yksityisyydensuojaan liittyvät huolenaiheet. Kyberturvallisuusratkaisut sisältävät usein salassa pidettävien ja arkaluontoisten tietojen analysointia, mikä aiheuttaa huolta tietovuodoista ja väärinkäytöksistä. Yksityisyydensuojaan liittyviin huoliin voidaan vastata teknisin toimenpitein, noudattamalla sääntelyä, viestimällä läpinäkyvästi ja ottamalla huomioon eettiset näkökulmat. Vastuullisen ja luotettavan tekoälyn käytön edellytyksenä on tasapaino tehokkaiden kyberturvallisuustoimien ja yksityisyyden suojelun välillä. Kyberturvallisuuden tekoälymallien on säilytettävä luottamuksellisuus niiden analysoidessa mahdollisesti arkaluonteisia tietoja.

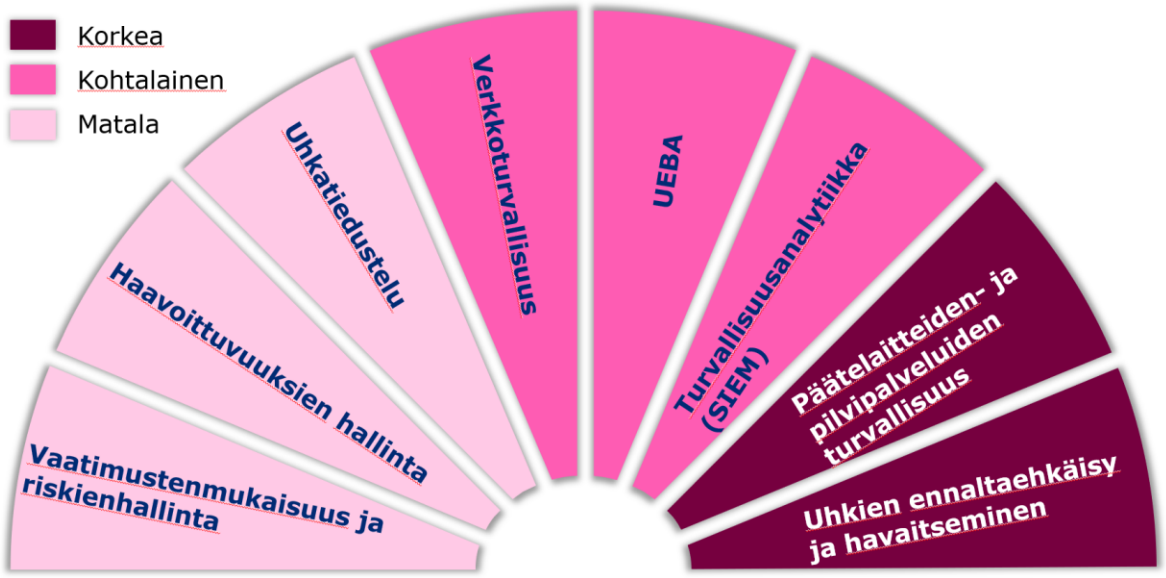
Yllä kuvailtujen erityispiirteiden käsittelyyn liittyy usein erikoistuneiden algoritmien kehittämistä, poikkeaman havaitsemistekniikoiden käyttöä, ensemble-mallien käyttöä tarkkuuden lisäämiseksi, käyttötapauskohtaisen tiedon hyödyntämistä ja mallin ymmärrettävyyden korostamista.



3. Tekoälyn soveltaminen kyberturvallisuusratkaisuissa

Tekoäly on ollut olennainen osa kyberturvallisuusratkaisuja jo 25 vuoden ajan, aina roskapostin ja tietojenkalasteluviestien suodattimista lähtien. Ajan myötä tekoälyä on sovellettu useisiin kyberturvallisuuden osa-alueisiin. Tekoälyn soveltaminen on edistyksellistä reaktiivisissa turvallisuustoimenpiteissä, kuten uhkien havaitsemisessa tai päätelaiteturvallisuudessa, mutta ennakoivissa turvallisuussovelluksissa hyödyntäminen on tuottanut haasteita.

Tämä osio esittelee tekoälyn yleisimmät käyttötapaukset ja mahdolliset tulevaisuuden sovellukset kahdeksalla keskeisellä kyberturvallisuusalueella. Osio tarjoaa esimerkkejä tekoälyn käytöstä ja määrittelee tekoälytehtävät tai kyvykkyydet, jotka liittyvät kyseiseen sovellukseen (ks. kappale 1.2). Lisäksi osiossa korostetaan kuhunkin sovellukseen liittyviä haasteita ja arvioidaan sovelluksen kypsyystasoa. Kuvaaja 2 ja taulukko 1 tiivistävät havainnot. Osio on tarkoitettu luettavaksi valikoivasti lukijan kiinnostuksen kohteiden perusteella. Eri sovellukset käyttävät samanlaisia tekoälytekniikoita ja kohtaavat samanlaisia haasteita; näin ollen sisällöissä voi olla päällekkäisyyksiä. Kiinnostuneet voivat tutustua kyselytutkimuksiin, jotka sisältävät lisätietoa sovelluksista sekä tarkemmat tekniset yksityiskohdat[1][2][3].



Kuvaaja 2: Tekoälyn kyberturvallisuussovellusten kehittyneisyys

[1] Apruzzese, G. et al. "The role of machine learning in cybersecurity." *Digital Threats: Research and Practice*. 2023

[2] Dasgupta, D. et al. "Machine learning in cybersecurity: a comprehensive survey." *The Journal of Defense Modeling and Simulation*. 2022

[3] Sarker, I.H. et al. "Ai-driven cybersecurity: an overview, security intelligence modeling and research directions." *SN Computer Science.* 2021

3.1 Uhkien ennaltaehkäisy ja havaitseminen

Uhkien ennaltaehkäisy ja havaitseminen edellyttää strategioiden ja työkalujen käyttöönottoa, jotta organisaation järjestelmät voidaan suojata mahdollisilta luottamuksellisuuden, eheyden tai saatavuuden vaarantavilta kyberturvallisuusuhilta. Maailmassa, jossa ihmiset ja laitteet ovat yhä enemmän yhteydessä toisiinsa, tekoälyn soveltaminen uhkien ennaltaehkäisyyn ja havaitsemiseen on uskottavan kyberturvallisuusstrategian kulmakivi. Tekoälyä on käytetty tähän tarkoitukseen yli 20 vuoden ajan. Varhainen ja onnistunut sovellus tekoälylle haitallisen sisällön havaitsemiseksi on ollut hahmontunnistus. Hahmontunnistuksen avulla kyetään erottamaan haitallisen sisällön tunnusmerkit tuntemattomien ohjelmistotiedostojen arviointia ja vertailua varten.

Havainnoinnissa tekoälyä käytetään ensisijaisesti haittaohjelmien tunnistamisessa. Pääasiallinen lähestymistapa perustuu binääristen luokkien erottamiseen vaarattomien ja haitallisten tiedostojen välillä. Havainnoinnin tehtävänä on tunnistaa haitallinen koodi, joka on piilotettu näennäisesti asianmukaisiin ajotiedostoihin, asiakirjoihin tai URL-osoitteisiin. Haitallisten tiedostojen tunnistamiseen käytetään tiedostoanalysoijia, jotka ovat joko staattisia tai dynaamisia. Staattiset analysoijat tarkastelevat lähde-/binäärikoodia nopeita vastauksia varten, kun taas dynaamiset suorittavat koodin hiekkalaatikkoympäristössä tarkkojen, mutta hitaampien tulosten saavuttamiseksi. Haittaohjelmien havaitsemiseen suunnitellut tekoälymallit kykenevät sopeutumaan jatkuvasti kehittyvään haittaohjelmaympäristöön. Ne ovat erinomaisia tiedon ekstrapoloinnissa, mikä auttaa tunnistamaan uusia haitallisia näytteitä, joissa hyödynnetään tunnettujen haittaohjelmien toimintatapoja muistuttavia hyökkäystekniikoita.

Tekoäly on osoittanut tehokkuutensa tietojenkalasteluviestien ja roskapostin havaitsemisessa, mutta syväoppimisen ja luonnollisen kielen käsittelyn viimeaikaiset harppaukset ovat olleet merkittäviä edistysaskeleita. Perinteiset binääriset luokittelumenetelmät ovat olleet ja ovat edelleen ensisijainen lähestymistapa roskapostin, tietojenkalasteluyritysten, sekä verkkosivustojen sisällön analysointiin. Syväoppimisen menetelmien kehittyminen on kuitenkin vähentänyt tarvetta ominaisuuksien suunnitteluprosessille. Syväoppivien mallien kehitys mahdollistaa sähköposti- tai SMS-viestien suoran käsittelyn raakatekstinä, mikä virtaviivaistaa analyysiprosessia ja parantaa tekoälypohjaisten havaitsemistekniikoiden sopeutumiskykyä. Luokittelumenetelmien käytössä haitallisten verkkotunnusten tunnistamiseen on saavutettu vastaavan kaltaista menestystä hyödyntämällä verkkosivuilta ja URL-osoitteista havaittuja piirteitä[4]. Parhaat tulokset saavutetaan monikerroksisella lähestymistavalla, joka hyödyntää erilaisia ominaisuuksia varmistaakseen kattavan ja vahvan suojan haitallisia URL-osoitteita vastaan[5].

Tekoäly on tehokas lähestymistapa uhkien ennaltaehkäisyyn ja havaitsemiseen, mutta sen käytössä on useita ongelmakohtia. Yksi keskeisimmistä haasteista liittyy jatkuvasti kehittyvään uhkaympäristöön, jossa uusia haittaohjelmatyyppejä ja hyökkäysmenetelmiä ilmenee säännöllisesti. Tämä edellyttää jatkuvaa sopeutumista ja tekoälymallien koulutusta. Tekoälymallin tekemät väärät

[4] Marchal, S. et al. "Know your phish: Novel techniques for detecting phishing sites and their targets." *36th IEEE International Conference on Distributed Computing Systems (ICDCS).* 2016

[5] Cohen, D. et al. "Website categorization via design attribute learning." *Computers & Security.* 2021

päätökset, kuten vaarattomien toimintojen tunnistaminen virheellisesti uhiksi tai todellisten uhkien huomiotta jättäminen, johtaa tarpeettomiin hälytyksiin ja riskitilanteisiin. Tekoälymallien vaikea tulkittavuus ja ymmärrettävyys on myös ongelma, sillä on tärkeää ymmärtää, mihin haittaohjelmien havaitseminen perustuu. Sivuston sisällön piilottaminen CAPTCHA:n (*kokonaan automatisoitu julkinen Turingin testi tietokoneiden ja ihmisten erottamiseksi toisistaan*) taakse vaikeuttaa haitallisten verkkotunnusten havaitsemista, sillä nykyiset menetelmät eivät kykene tulkitsemaan CAPTCHA-kyselyitä.

Tekoälyn käyttö uhkien ennaltaehkäisyssä ja havaitsemisessa on kehittynyttä. Tekoälyllä on merkittäviä käyttömahdollisuuksia niin ohjelmistojen hyväksikäytön ja haittaohjelmien tunnistamisessa, tietojenkalastelun ja roskapostin havaitsemisessa kuin haitallisten verkkotunnusten estämisessäkin. Kone- ja syväoppimista on käytetty näillä osa-alueilla tehokkaasti, mikä on mahdollistanut uhkien ennakoivan tunnistamisen ja lieventämisen. Tämä on merkittävästi parantanut kyberturvallisuustoimien tehokkuutta ja kykyä käsitellä modernien IT-järjestelmien kasvavaa monimutkaisuutta.

3.2 Päätelaitteiden ja pilvipalveluiden kyberturvallisuus

Päätepisteiden ja pilven tunnistamisen ja reagoinnin -palvelut (*endpoint and cloud detection & response*, *EDR & CDR*) ovat olennaisia tietoturvaratkaisuja paikallisille laitteille ja pilvipalveluille. EDR- ja CDR-palvelut seuraavat ja analysoivat laitteilla ja palvelimilla ilmeneviä tapahtumia, kuten tiedostotoimintoja, verkkoyhteyksiä ja käyttäjien toimia, ja luovat tarkan kuvan tapahtumista mahdollisten uhkien havaitsemiseksi ja lieventämiseksi.

Tekoälyllä on ratkaiseva rooli EDR- ja CDR-palveluissa. Tekoälyn avulla palvelut pystyvät käsittelemään tehokkaasti laitteista ja pilvipalveluista peräisin olevaa runsasta ja monimutkaista dataa. Yksi ensimmäisistä tekoälyn onnistuneista käyttötavoista uhkien tunnistamisessa ja reagoinnissa on harmittomien ja epäolennaisten turvallisuustapahtumien suodattaminen esiintymistiheyden perusteella. EDR- ja CDR-järjestelmät tuottavat merkintöjä turvallisuustapahtumista, jotka voivat olla hyödyllisiä kyberhyökkäysten havaitsemisessa. Valtaosa tapahtumista on kuitenkin hyvin tavallisia, eivätkä ne liity hyökkäyksiin. Tilastolliset mallinnustekniikat voivat helposti tunnistaa ja suodattaa vaarattomat tapahtumat pois jatkotarkastelusta niiden esiintymistiheyden, keston ja samankaltaisuuden perusteella. Tämä taas vähentää tapahtumien korrelaatiojärjestelmien käsittelykuormitusta ja vääriä hälytyksiä. Tekoäly voi myös auttaa havaitsemaan poikkeavia ketjuja ja tapahtumien sekvenssejä, jotka eroavat tyypillisestä käyttäytymisestä. Ketjut voivat liittyä haitallisiin toimintoihin, kuten haittaohjelmien suorittamiseen, tietojen poistoon tai käyttöoikeuksien laajentamiseen. Hyödyntämällä sekvenssilouhintaan ja graafianalyysiin perustuvia poikkeamien havaitsemistekniikoita tekoäly voi tunnistaa kaavoja ja tuottaa järjestelmähälytyksiä tietoturva-asiantuntijoille tarkempaa tutkimusta ja toimenpiteitä varten.

Tekoäly helpottaa epäilyttävien tapahtumien priorisointia ja analysointia. Ryhmittelytekniikoita voidaan käyttää samankaltaisten tapahtumien järjestelyyn ja erottamaan koko ryhmää edustava prototyyppi. Tämä säästää tietoturva-asiantuntijoiden aikaa, kun heidän tarvitsee analysoida vain yhtä tapahtumaa

useiden samankaltaisten sijaan. Tekoäly voi auttaa arvioimaan tapahtumien vakavuutta, vaikutusta ja kiireellisyyttä ja antaa ohjeita parhaasta toimintatavasta.

Tekoälymallit tulee räätälöidä tiettyyn tarkoitukseen ja laite- ja palvelukohtaisesti, kuten käyttöjärjestelmiin, sovelluksiin ja toimintoihin. Tekoälymallien suorituskyvyn arviointi on haastavaa, koska vain pieni osa havaituista tapahtumista liittyy tietoturvauhkiin. Tunnistamiseen ja reagointiin käytettävien tekoälymallien on oltava tarkkoja ja luotettavia, koska ratkaisun käytettävyys ja luotettavuus heikentyvät väärien havaintojen vuoksi. Mallien on myös oltava nopeita ja responsiivisia, koska viive tunnistamisessa ja reagoinnissa voi lisätä onnistuneiden hyökkäysten vaikuttavuutta.

Tekoälyn kypsyystaso EDR- ja CDR-sovelluksissa on korkea. Tekoälystä on tullut näille palveluille olennainen mahdollisten uhkien havaitsemista ja niihin vastaamista tehostava osa.

3.3 Verkon tietoturva

Tekoäly tarjoaa edistyksellisiä mahdollisuuksia suurten verkkoliikennemäärien seurantaan ja analysointiin monimutkaisissa verkkoinfrastruktuureissa. Tekoälyn tärkeimmät sovellusalueet verkon tietoturvassa ovat tunkeutumisen havaitsemis- ja ehkäisyjärjestelmät (*intrusion detection, IDS* ja *intrusion prevention, IPS*), joissa verkkoliikenteestä etsitään poikkeamia ja epäilyttäviä toimintoja, jotka saattavat viitata luvattomaan käyttöön, vaarantuneisiin tietovarantoihin tai mahdollisiin hyökkäyksiin. Toinen suosittu verkkoturvallisuuden sovellusalue on tuntemattoman verkkoliikenteen analysointi ja liikenteen lähteen tunnistaminen, olipa kyseessä sitten käyttäjä, laite tai sovellus, potentiaalisesti haitallinen viestintä, kuten komento- ja hallintaliikenne (*command-and-control, C&C*) tai matojen leviäminen verkossa.

Havaitsemis- ja ehkäisyjärjestelmät perustavat poikkeamien havaitsemiskyvyt yleensä tavanomaisen verkkoliikennekäyttäytymisen mallintamiseen, ja tulkitsee poikkeamat mahdollisina hyökkäyksen merkkeinä. Tämä lähestymistapa on tehokas, kun hyökkäykset aiheuttavat ilmeisiä ja merkittäviä muutoksia verkkoliikenteessä, kuten portti- ja palveluskannauksia tai palvelunestohyökkäyksiä. Poikkeamien havaitsemisjärjestelmät eivät kuitenkaan kykene tunnistamaan tehokkaasti hienovaraisia hyökkäyksiä tai haittaohjelmaviestintää, jotka muistuttavat tavanomaista viestinvaihtoa.

Edistyneiden ja hienovaraisten hyökkäysten tunnistaminen edellyttää poikkeamien havaitsemisen hienosäätöä herkemmäksi, mikä johtaa väärien hälytysten määrän kasvuun. Hienovaraisten uhkien käsittelemiseksi havaitsemis- ja ehkäisyjärjestelmät hyödyntävät tarkempia luokittelumenetelmiä. Binääriluokittelua käytetään erottamaan hyödyllinen viestintä haitallisesta viestinnästä, kun taas moniluokkaluokittelua käytetään tunnistamaan erilaisia hyökkäyksiä ja haitallisia käyttäytymismalleja. Hahmontunnistusta käytetään käyttäjän, laitteen ja sovelluksen tunnistamiseen. Tunnistettavan kohteen verkkoviestintä mallinnetaan yksityiskohtaisesti ja mallia käytetään viestinnän tarkempaan luokitteluun (*pattern matching*). Esimerkiksi

signaalinkäsittelytekniikoita voidaan yhdistää yksinkertaisiin luokittelumenetelmiin ja näin tunnistaa IoT-laitteita kotiverkoissa[6].

Koneoppimisalgoritmien käyttämät verkkoliikenteen tunnusmerkit poimitaan verkkoliikenteestä, jotta verkkoliikenteessä tapahtuvan viestinnän tyyppi voidaan osoittaa. Poiminta suoritetaan vaihtelevilla yksityiskohtaisuuden tasoilla. Tasot valikoituvat havaittavan toiminnan, analyysiviiveen, käytettävissä olevan laskentatehon sekä protokollatiedon saatavuuden mukaan. Verkkoliikenteen salaus rajoittaa poimittavissa olevaa tietoa. Tietoa voidaan poimia yksittäisistä verkkopaketeista, verkkopakettien ryhmistä, viestintävirroista tai kaikesta verkkoliikenteestä.

Verkon turvallisuussovelluksien merkittävin haaste on analysoitavan verkkoliikenteen monimuotoisuus. Viestintäprotokollien runsaus vaikeuttaa monimutkaisten ympäristöjen ja käyttäytymisen mallintamista. Verkkoliikenne on yhä enemmän salattua, mikä estää pääsyn oleellisiin paketti- ja kuormatietoihin. Salattu verkkoliikenne ja verkkoliikenteen monimuotoisuus altistavat hienovaraisille hyökkäyksille, jotka kykenevät naamioitumaan ja sopeutumaan hyväksyttävän verkkoviestinnän sekaan.

Tietoverkkojen turvallisuuden tekoälysovellukset soveltuvat parhaiten yksinkertaisten, vakaiden ja ennustettavien verkkoympäristöjen valvontaan, jotka koostuvat laitteista, joilla on yksinkertaisia käyttäytymismalleja, kuten esimerkiksi esineiden Internet -verkot (*Internet-of-Things, IoT*) tai teollisuuden ohjausjärjestelmät (*Industrial Control Systems, ICS*). Esimerkiksi menestyksekäs verkon poikkeamien havaitsemisjärjestelmä IoT-verkossa perustuu matalan tarkkuustason verkkotietoihin ja laitekohtaisiin tekoälymalleihin, jotka käyttävät sekvenssien mallintamisen tarkoitettua DNN-tyyppiä, toistuvaa neuroverkkoa (*Recurrent Neural Network, RNN*)[7].

Tekoälyä on sovellettu verkon turvallisuuteen pitkään, mutta sen kypsyystasossa ei ole saavutettu merkittävää kehitystä. Tämä johtuu pääasiassa tietoverkkojen jatkuvasta kehityksestä ja monimutkaisuudesta, joka luo verkkoturvallisuudelle säännöllisesti uusia haasteita tekoälytekniikoiden soveltamisessa.

3.4 Käyttäjän ja entiteetin käyttäytymisanalyysi (User and Entity Behaviour Analytics, UEBA)

Käyttäjän ja entiteetin käyttäytymisanalyysi (*UEBA*) analysoi käyttäjien, laitteiden ja sovellusten käyttäytymistä ja pyrkii havaitsemaan mahdollisia turvallisuusuhkia ja poikkeavia toimintoja organisaatioissa[8]. UEBA:n avulla organisaatiot voivat vahvistaa ennakoivaa ja mukautuvaa turvallisuuslähestymistapaa. UEBA hyödyntää tekoälyä hienovaraisten ja kehittyvien uhkien havaitsemiseksi, vasteaikojen lyhentämiseksi sekä turvallisuusuhkien tunnistamiseksi ja käsittelemiseksi.

Käyttäytymisanalyysia sovelletaan käyttäjän ja yksikön tapahtumadataan, joka on tyypillisesti runsasta ja monimuotoista. Tyypillisiä mallinnettavia ja seurattavia

[6] Marchal, S. et al. "AUDI: Toward autonomous iot device-type identification using periodic communication." *IEEE Journal on Selected Areas in Communications*. 2019

[7] Nguyen, T.D. et al. "DÏoT: A federated self-learning anomaly detection system for IoT." *39th IEEE International conference on distributed computing systems (ICDCS).* 2019

[8] Shashanka, M. et al. "User and entity behavior analytics for enterprise security." *IEEE International Conference on Big Data.* 2016

tapahtumia ovat kirjautumisyritykset, asiakirjojen lataukset/lähetykset sekä tiettyjen sovellusten ja prosessien käyttö. Mallinnusta varten huomioidaan tapahtuman tyyppi, sen tapahtuma-aika, sijainti, kesto, tekijä ja vastaavat ominaisuudet. Edellä mainittuja tekijöitä voidaan mallintaa käyttämällä ohjaamatonta klusterointitekniikkaa ja aikasarja-analyysiä, joiden avulla havaitaan tapahtumien sijainti, ajallisuus, kausiluonteisuus ja samankaltaisuus. Järjestelmä luo *profiilin*, joka kuvastaa tietyn käyttäjän tai yksikön tyypillistä toimintaa ja vuorovaikutusta. Yksiköiden ja käyttäjien toimintaa verrataan opittuihin profiileihin ja tunnistetaan poikkeamia. Poikkeamiksi liputetut tapahtumat tutkitaan mahdollisten sisäpiirin uhkien, luvattoman pääsyn tai muiden huolestuttavien merkkien varalta. Tekoälypohjaisia UEBA-palveluita tarjotaan organisaatioille identiteetinhallintaan kohdistuvien riskien lieventämiseksi jo nyt.[9]

Tarkan käyttäytymismallin rakentaminen edellyttää suurta määrää historiallista tapahtumadataa profiloiduista yksiköistä ja käyttäjistä. UEBA tarvitsee pitkän oppimisjakson, jonka aikana se ei voi suojata uusia käyttäjiä. Oppimisjakson aikana yksiköiden ei tulisi kuvata tavanomaisesta poikkeavaa käyttäytymistä, jotta profiilit eivät sisältäisi poikkeamia, joita järjestelmä pyrkii havaitsemaan. Tämä lisää UEBA:n soveltamisen haasteellisuutta.

UEBA:n kohteena olevien yksiköiden käyttäytyminen on tyypillisesti monimutkaista, mikä aiheuttaa ajan myötä monimutkaistuvaa luonnollista vaihtelua. Käyttäjien ja yksiköiden käyttäytymisen luonnollinen vaihtelu tekee UEBA:sta alttiin tekemään vääriä hälytyksiä poikkeavista tapahtumista, jotka eivät edusta todellista uhkaa. Käyttäytymisprofiileja on säännöllisesti koulutettava ja päivitettävä, jotta käyttäytymisen kehitys voidaan huomioida. Onnistumisen avain UEBA:ssa on arvioida havaittujen poikkeamien merkityksellisyys ja tarkkuus. Havaintoihin vastaamisessa on pyrittävä saavuttamaan tasapaino uhkien lieventämisen ja ennaltaehkäisyn, sekä käyttäjälle koituvien häiriöiden välillä.

Uhkiin kohdistetut toimenpiteet on sovitettava turvallisuusuhkan tasoon ja mahdollisiin seurauksiin. Toimenpiteet pitää olla myös helposti mukautettavissa. Toimenpiteet määritellään yleensä manuaalisesti ja ne voivat olla erilaisia eri käyttäjäryhmille saman tyyppisessä tapahtumassa. Esimerkkejä toimenpiteistä tapahtumalle, jossa ladataan salainen tiedosto keskellä yötä epätavallisesta sijainnista voivat olla

a) toiminnan estäminen,

b) lisäautentikoinnin pyytäminen (esim. monivaiheinen tunnistautuminen),

c) hiljainen hälytys, jolloin hälytyksellä ei ole suoraa vaikutusta käyttäjään.

UEBA-järjestelmän kouluttaminen ja oikeasuhtaisten toimenpiteiden hienosäätö sopivan turvallisuustason saavuttamiseksi voi osoittautua pitkäksi prosessiksi. UEBA-suorituskyvyn parantamiseksi käyttäjien tulee voida antaa palautetta järjestelmän tekemistä päätöksistä.

Käyttäjien käyttäytymisen monimuotoisuudesta johtuvat mallintamisen haasteet johtavat verrattain alhaiseen tarkkuuteen UEBA-sovelluksissa, ja havaitsemiskyvyn

[9] *What is Entra ID protection?*. https://learn.microsoft.com/en-us/entra/id-protection/overview-identity-protection

kehittäminen vaatii runsaasti hienosäätöä. Näiden seikkojen vuoksi UEBA-sovelluksia ei voida käyttää kriittisessä automatisoidussa päätöksenteossa.

3.5 Turvallisuusanalytiikka

Tietoturvatiedon ja -tapahtumien hallintajärjestelmät (*Security Information and Event Management, SIEM*) integroivat ja analysoivat tietoja eri lähteistä, kuten tapahtumien tietojen analyysistä (*EDR* ja *CDR*), käyttäjien ja yksiköiden käyttäytymisanalytiikasta (*UEBA*), verkon turvallisuudesta ja muista turvallisuusjärjestelmistä. Integraation avulla organisaatio saa kokonaisvaltaisen näkymän koko yritysverkon turvallisuuden tasosta. Perinteisissä SIEM-ratkaisuissa on useita rajoitteitta: Ne tuottavat usein vääriä hälytyksiä, perustavat analyysinsä rajoitetulle määrälle tietoa ja ovat usein kyvyttömiä vastamaan tapahtumiin nopeasti ja tehokkaasti.

Tekoäly voi automatisoida esimerkiksi tiedonhaun ja hahmontunnistamisen avulla tapahtuvan tietojen keräämisen, normalisoinnin, rikastamisen ja korreloinnin. Tekoäly voi yhdistää menneitä turvallisuustapahtumia ja uhkatiedustella havaitakseen ja torjuakseen edistyksellisiä, jatkuvia uhkia (*advanced persistent threat, APT*), jotka osaavat välttää perinteisiä turvallisuustoimenpiteitä. Tekoälyn avulla voi automatisoida hälytysten luomisen ja toteuttaa ennalta määriteltyjä toimenpiteitä, tai järjestää osittain monimutkaisten poikkeamien hallintaa generointimenetelmillä. Täydellistä automatisointia on kuitenkin vaikea saavuttaa, sillä nykyisillä generatiivisilla tekoälymenetelmillä on hallusinoinnin kaltaisia rajoituksia.

Täydellinen automatisointi vaatisi huolellista tekoälymallien koulutusta. Malleille tulisi antaa tietoa tiimin organisaatiosta, yksilöiden kyvyistä, olemassa olevista työnkuluista ja muista vaikuttavista tekijöistä. Lupaavampia lähestymistapoja tekoälyn käyttöön SIEM:issä on tietoturva-analyytikkojen tekemän turvallisuustapahtumien tutkimuksen priorisointi ja tukeminen. Klusterointitekniikoilla turvallisuustapahtumia voidaan ryhmitellä ja vähentää manuaalisen analyysin aiheuttamaa työmäärää. Järjestelmä voi käyttää järjestelyä myös turvallisuustapahtumien manuaalisen analyysin priorisoimiseen.

Tekoälyn integroinnissa SIEM-järjestelmiin on haasteita. Datan keruu voi olla vaikeaa, koska data on usein saatavilla huonolaatuisena, tai kerääminen voi olla rajoitettua yksityisyydensuojan ja arkaluonteisuuden vuoksi. Lisäksi eri järjestelmistä peräisin olevien erilaisten tietojen normalisointi ja korrelointi on haastavaa. Tekoälymallit voivat virheellisesti luokitella vaarattomia toimintoja uhkiksi tai jättää todelliset uhat havaitsematta. Molemmat tilanteet voivat johtaa vakaviin seurauksiin.

Tekoälyä on sovellettu SIEM-sovelluksiin, mutta teknologian kypsyys on keskitasoa. SIEM-sovelluksissa on vielä rutkasti mahdollisuuksia innovaatioille. Tekoäly voi käyttää luonnollista kielenkäsittelyä ja päättelyä auttaakseen tietoturva-asiantuntijoita tietokantojen kyselyissä ja alustavissa turvallisuusarvioissa. Tekoäly voi myös käyttää generatiivisia malleja ja luoda realistisia skenaarioita ja kyberhyökkäyssimulaatioita. Simulointi auttaisi turvallisuustiimejä testaamaan ja parantamaan SIEM-ratkaisujaan, kouluttautumaan ja arvioimaan valmiuttaan.

3.6 Uhkatiedustelu

"Tekoäly rikkoo kielirajoja, sillä se soveltuu hyvin uhkatiedon kääntämiseen. Luokittelumenetelmät ja ymmärrettävät tekoälymenetelmät voivat ohjata tietoturva-asiantuntijoita uhka-analyysiin tarkoitetuilla ohjauspaneeleilla kohdistamalla heidän huomionsa nouseviin hyökkäystrendeihin."

Uhkatiedustelu perustuu tietojen keräämiseen, käsittelyyn ja analysointiin uhkatoimijoiden tavoitteiden, kohteiden ja hyökkäysstrategioiden ymmärtämiseksi. Uhkatiedustelu on turvallisuustoimenpide, jonka tarkoituksena on ennakoida ja estää kyberhyökkäykset ennen niiden tapahtumista. Tekoälyn integrointi uhkatiedustelualustoihin vaikuttaa lupaavalta. Tekoäly voi automatisoida toistuvia tehtäviä, joka mahdollistaa tietoturva-asiantuntijoiden keskittymisen monimutkaisiin uhka-analyyseihin. Tekoälyn kyky käsitellä suuria tietomääriä ja havaita kaavoja voi tarjota ratkaisevia oivalluksia, joiden avulla ennaltaehkäistään ja havaitaan uhkia. Nykypäivänä tiedonkäsittely on erityisen arvokasta, sillä uhkien määrä kasvaa ja uhista tulee yhä monimutkaisempia.

Tekoäly voi rikastuttaa uhkatiedustelualustoja automatisoimalla kyberuhkatiedusteluraporttien ja avoimen lähdekoodin uhkatiedustelun (*open-source threat intelligence, OSTI*) syötteiden hakua ja integrointia, mikä varmistaa uhkatiedon jatkuvan päivityksen tuoreimmalla tiedolla.[10] Suuria kielimalleja voidaan myös käyttää kybertiedusteluraporttien yhteenvedon ja synteesin tekemiseen, mikä säästää aikaa uhkien analysoinnissa. Tunnistettujen uhkatietolähteiden lisäksi tietojen noutotekniikoita voidaan käyttää uhkatiedon löytämiseen moninaisista lähteistä, kuten sosiaalisesta mediasta, jossa haavoittuvuustietoja voidaan löytää syväoppimistekniikoita käyttämällä.[11] Lisäksi tekoäly voi seurata kyberuhkien keskeisten teemojen ajallista kehitystä. Teemojen ajallisen kehityksen seuranta edistää uhkien hallinnan ennakoivaa lähestymistapaa tunnistamalla kaavoja ja mahdollisia tulevaisuuden uhkia.[12] Tekoäly rikkoo kielirajoja, sillä se soveltuu hyvin uhkatiedon kääntämiseen. Luokittelumenetelmät ja ymmärrettävät tekoälymenetelmät voivat ohjata tietoturva-asiantuntijoita uhka-analyysiin tarkoitetuilla ohjauspaneeleilla kohdistamalla heidän huomionsa nouseviin hyökkäystrendeihin.

Tekoäly tarvitsee suuria määriä korkealaatuista dataa, ja se voi tuottaa virheellisiä tuloksia. Virheelliset tulokset voivat merkittävästi vahingoittaa uhkatiedustelualustassa saatavilla olevan tiedon laatua. Samalla kun käsitellään suurta määrää monimuotoisia syötteitä, on varmistettava, että olennainen data ei jää huomaamatta ja että erilaisten tietojen välille luodut korrelaatiot tuottavat järkeviä tuloksia.

[10] Pantelis, G. et al. "On Strengthening SMEs and MEs Threat Intelligence and Awareness by Identifying Data Breaches, Stolen Credentials and Illegal Activities on the Dark Web." *16th International Conference on Availability, Reliability and Security.* 2021

[11] Iorga, D. et al. "Yggdrasil—early detection of cybernetic vulnerabilities from Twitter." *23rd International Conference on Control Systems and Computer Science.* 2021

[12] Kim, G. et al. "Automatic extraction of named entities of cyber threats using a deep Bi-LSTM-CRF network." *International journal of machine learning and cybernetics*. 2020

3.7 Haavoittuvuuksien hallinta

Tietoturvallisuuden haavoittuvuuksien hallinta on prosessi, jossa tunnistetaan, arvioidaan ja lievennetään IT-järjestelmien tietoturva-aukkojen aiheuttamia riskejä. Kyberturvallisuudessa yhteistyö ja tiedonjako edistävät tietoturvan haavoittuvuuksien hallintaa. Yhteistyöstä ja tiedonjaosta hyvä esimerkki on Common Vulnerabilities and Exposures (*CVE*) -luettelo, joka on tarkoitettu tunnettujen haavoittuvuuksien julkisten tietokantojen tallentamiseen ja ylläpitämiseen. Ajan tasalla olevan tiedon ylläpitäminen vastikään löydetyistä haavoittuvuuksista ja niiden vaikutuksista organisaatioon, sekä uusien haavoittuvuuksien löytäminen alkuperäisestä lähdekoodista, ovat vaativia tehtäviä. Tekoäly voi auttaa näiden ongelmien ratkaisemisessa.

Tekoälyä voidaan käyttää uusien haavoittuvuuksien seuraamiseen tietokannoissa ja testaamaan, koskevatko ne organisaation ohjelmistoa tai koodia. Esimerkiksi luonnollisen kielen käsittelyä yhdistettynä järjestelytekniikoihin voidaan käyttää tiettyyn ohjelmistoon todennäköisimmin vaikuttavien, tunnettujen haavoittuvuuksien luetteloimiseen ja priorisointiin.[13] Syväoppimista voidaan käyttää hahmontunnistuksen kanssa tunnettujen haavoittuvuuksien tunnistamiseen lähdekoodista. [14] Automaattinen haavoittuvuuskategorisointi on toinen alue, jossa tekoälyn kaltaisia luokittelukykyjä voidaan käyttää organisaatiossa havaittujen haavoittuvuuksien jakamiseen ennalta määritettyihin ryhmiin Common Weakness Enumeration (*CWE*) -luokituksen mukaisesti.[15] Luokittelu auttaa järjestämään ja priorisoimaan haavoittuvuuksia niiden vakavuuden, vaikutuksen ja CVE-tiedon mukaan. Tekoälymenetelmiä on käytetty myös uusien tuntemattomien haavoittuvuuksien löytämiseen fuzz-testauksessa.

Generatiivista tekoälyä voidaan käyttää luomaan fuzz-testitapauksia niin, että ne kattavat paremmin haavoittuvuuksia pelkkään globaaliin koodikattavuuteen perustuvan ohjauksen sijaan.[16] Tulevaisuudessa generatiivisia tekoälymenetelmiä voidaan mahdollisesti käyttää ohjelmistojen fuzz-testauksessa suoraan hyökkäävällä tavalla, mutta tällaisesta sovelluksesta ei ole vielä kovin montaa onnistunutta esimerkkiä. Generatiivisiin tekoälymenetelmiin perustuvat uudet työkalut ja suuret kielimallit vaikuttavat lupaavilta myös hyökkäystiimien ja tunkeutumistestausten harjoituksille.[17] Tekoälyratkaisut voivat tukea hyökkäyksessä käytettyjen työkalujen valinnassa, testien seuraavissa vaiheissa ja testitulosten tulkinnassa.

Riittävän ja luotettavan kouluttamis- ja testaamisdatan puute on yksi tekoälyn soveltamisen keskeisistä haasteista, erityisesti kun kyse on tuntemattomien- ja nollapäivähaavoittuvuuksien havaitsemisesta. Generatiiviset tekoälymenetelmät kamppailevat edelleen kääntämiseen ja suorittamiseen tiukkaa syntaksia vaativan koodin tai suoritettavien komentojen generoimisessa. Yksi väärä merkki ihmisen

[13] Huff, P. et al. "A recommender system for tracking vulnerabilities." *16th International Conference on Availability, Reliability and Security* (ARES). 2021

[14] Jeon, S., and Kim, H.K. "AutoVAS: An automated vulnerability analysis system with a deep learning approach." *Computers & Security.* 2021

[15] Saha, T. et al. "SHARKS: Smart hacking approaches for risk scanning in Internet-of-Things and cyber-physical systems based on machine learning." *IEEE Transactions on Emerging Topics in Computing*. 2021

[16] Saha, T. et al. "SHARKS: Smart hacking approaches for risk scanning in Internet-of-Things and cyber-physical systems based on machine learning." *IEEE Transactions on Emerging Topics in Computing*. 2021

[17] Deng, G. et al. *PentestGPT: An LLM-empowered Automatic Penetration Testing Tool.* arXiv. 2023

luettavaksi tarkoitetussa tekstissä ei merkittävästi haittaa tekstin ymmärrettävyyttä, kun taas suoritettavissa komennoissa, kuten lähdekoodissa tai verkkopaketeissa, se voi tehdä tulkinnan mahdottomaksi. Näiden lisäksi tekoälyn käytön eettiset ja oikeudelliset vaikutukset haavoittuvuuksien löytämiseen ja hyväksikäyttöön voivat muodostaa uhan IT-järjestelmien turvallisuudelle ja yksityisyydelle, sillä näitä tekoälyjärjestelmiä voivat hyödyntää myös hyökkääjät kyberhyökkäysten suorittamiseen.

Tekoälyn käyttö haavoittuvuuksien hallinnassa on kypsyystasoltaan matalaa, koska useimmat sovellutukset ovat edelleen kehitys- tai testivaiheessa. Tekoälyn potentiaaliset hyödyt haavoittuvuuksien hallinnassa ovat kuitenkin merkittäviä. Tekoäly tuo jo nyt etuja tukemalla analyytikoita tunkeutumistestauksessa ja vähentämällä manuaalisen haavoittuvuusanalyysin kustannuksia ja vaivaa. Generatiivisten tekoälymenetelmien kehittyessä ne voivat tulevaisuudessa ottaa merkittävämpää roolia haavoittuvuuksien hallinnassa.

3.8 Vaatimustenmukaisuus ja riskienhallinta

Tietoturvariskien hallinta on prosessi, jossa tunnistetaan, arvioidaan ja priorisoidaan organisaation toimintaan, omaisuuteen ja yksilöihin liittyviä kyberturvallisuusriskejä. Traficomin luoma Kybermittari[18] on esimerkki työkalusta, jota voidaan käyttää tietoturvariskien hallintaan ja organisaation kyberturvallisuuden kypsyystason arviointiin. Tekoäly voi tukea ja muokata tietoturvariskien hallintaa, joka on monimutkainen, kallis, aikaa vievä ja yleensä aktiivista ihmisen osallistumista ja asiantuntemusta vaativa prosessi. Riskianalyysin sekä tietoturvakyvykkyyksien ja vaikutusten arvioinnin automatisointi vahvistaa riskienhallintatiimien kykyjä hyödyntämällä sisäisiä ja ulkoisia riskitietoja, mikä taas helpottaa riskitekijöiden ja niihin liittyvien mittareiden reaaliaikaista arviointia. Tekoäly toimii katalyyttinä monimutkaisten tehtävien automatisoinnissa, kuten riskipisteiden laskemisessa, mahdollisten tietoturvatapahtumien todennäköisyyden päättelemisessä, kriittisten haavoittuvuusriskien tunnistamisessa ja kattavien riskiarviointien ja analyysien tekemisessä.

Tekoäly voi helpottaa digitaalisten informaatiohyödykkeiden (engl. *asset*) hallintaa tunnistamalla, luokittelemalla sekä seuraamalla tietoja, ihmisiä, laitteita ja järjestelmiä. Riskienhallinta perustuu informaatiohyödykkeiden hallinnalle. Tietojen hallinnan automatisointia voidaan tukea luokittelu- ja kaavioanalyysitekniikoilla, joita sovelletaan organisaation jokaiseen informaatiohyödykkeeseen. Yhdistettynä skannaustekniikoihin, jotka tunnistavat säännöllisesti uusien hyödykkeiden ilmestymisen tai vanhojen hyödykkeiden poistamisen, tekoäly voi auttaa pitämään hyödykeluettelon ajan tasaisena. Tunnistettujen hyödyke-erien ominaisuuksia voidaan käyttää syötteenä ranking-algoritmeille, jotta voidaan automaattisesti tunnistaa niiden suhteellinen tärkeys ja turvallisuusriski. Näitä tuloksia voi puolestaan hyödyntää turvatoimien priorisoinnissa, jotta voidaan suojata etenkin kriittisimpiä hyödykkeitä ja vähentää niihin kohdistuvia riskejä. Tekoälyä voidaan käyttää organisaation käyttämien turvatoimien luetteloimiseen ja arvioimaan, miten puolustukselliset toimet lieventävät tunnistettujen uhkien luomia turvallisuusriskejä. Tekoälysovellus voi auttaa arvioimaan organisaation

[18] Traficom. *Kybermittari* https://www.kyberturvallisuuskeskus.fi/fi/palvelumme/tilannekuva-ja-verkostojohtaminen/kybermittari

turvallisuustasoa automatisoimalla nämä tyypillisesti manuaaliset inventointi- ja arviointitehtävät, joita myös Kybermittarissa edellytetään.

Tekoälytekniikat voivat tukea toimitusketjun riskienhallintaa automatisoimalla uhka-analyysia ja ennakointia, optimoimalla tietoturvan taloudellisia kustannuksia ja luomalla arvioita toimitusketjujen tietoturvallisuudesta. Arviointia voidaan laajentaa koskemaan suojausta, havaitsemista, vastetta ja elpymismekanismeja, jotka liittyvät mihin tahansa organisaation informaatiohyödykkeisiin ja prosesseihin. Toimitusketjujen riskienhallintaan voidaan käyttää esimerkiksi tietojen nouto- ja korrelaatiotekniikoita. Tekniikat soveltuvat organisaation sisäisten ja ulkoisten järjestelmien, turvallisuusdokumentaation, inventaarion, historiallisten tapahtumatietojen, verkkotietojen, järjestelmälokien, haavoittuvuustietojen ja uhkatiedustelun tarkasteluun.

Tekoälyllä voi olla merkittävä rooli organisaation turvallisuuden vahvistamisessa ja hallinnassa. Tekoälyä on käytetty jo nyt esimerkiksi viestintäverkkojen toimintaperiaatteiden täytäntöönpanossa[19]. Reitittimissä voi olla toimintaperiaatteiden välityspalvelin (*policy proxy*), mikä tunnistaa verkkoliikennettä, jota toimintaperiaatteet koskevat. Reitittimet käyttävät luokittelua toimintaperiaatteiden toteutumisen seurantaan ja siten varmistavat, että sääntöjä noudatetaan. Tietojen nouto- ja generointikyvyt voivat poimia merkkejä riskeistä automaattisesti ja muodostaa niiden pohjalta toimintatapoja, jotta tietoturvaloukkauksien käsittelyllä ilmeneviä riskejä voidaan ennaltaehkäistä nopeasti.

Tekoälysovellusten valmius vaatimustenmukaiseen ja riskienhallintaan kykenevään toimintaan on edelleen alhainen. Riskille ei ole yhtä määritelmää, eikä riskinarvioinnille tai riskienhallinnalle yksiselitteisiä prosesseja. On siis vaikeaa kouluttaa tekoälyjärjestelmiä suorittamaan tehtävää, jolla ei ole selkeää määritelmää. Viimeaikaiset tiedonhaun ja generoinnin edistysaskeleet suurissa kielimalleissa tarkoittavat kuitenkin tälle sovellukselle potentiaalia kehittyä.

[19] Odegbile, O. et al. "Dependable Policy Enforcement in Traditional Non-SDN Networks." *39th IEEE International Conference on Distributed Computing Systems (ICDCS).* 2019

Sovellus	Kyvyt	Kypsyys
Uhkien ennaltaehkäisy ja havaitseminen	• Luokittelu • Hahmontunnistus • Ryhmittely	Korkea
Päätelaitteiden ja pilvipalvelujen tietoturva	• Luokittelu • Poikkeamien havaitseminen • Hahmontunnistus • Ryhmittely • Järjestäminen	Korkea
Verkon tietoturva	• Luokittelu • Poikkeamien havaitseminen • Hahmontunnistus	
UEBA	• Käyttäytymisanalyysi • Ryhmittely	Kohtalainen
Turvallisuusanalytiikka (SIEM)	• Hahmontunnistus • Poikkeamien havaitseminen • Tiedonhaku • Järjestäminen • Generointi	Kohtalainen
Uhkatiedustelu	• Luokittelu • Ryhmittely • Tiedonhaku • Generointi	Matala
Haavoittuvuuksien hallinta	• Luokittelu • Hahmontunnistus • Järjestäminen • Generointi	Matala
Vaatimustenmukaisuus ja riskienhallinta	• Luokittelu • Tiedonhaku • Järjestäminen • Generointi	Matala

Taulukko 1: Tekoälyn kyberturvallisuussovellukset, niiden kyvyt ja kypsyystaso

4. Suositukset ja parhaat toimintatavat tekoälyn käyttöön

Tekoälymallien onnistunut käyttöönotto ja kehittäminen edellyttää datatieteilijöiden ja kyberturvallisuusasiantuntijoiden osaamisen yhdistämistä. Tässä kappaleessa esiteltävät suositukset ja havainnollistukset antavat suuntaviivoja tekoälypohjaisten kyberturvallisuusratkaisujen käyttöönottoon ja kehittämiseen.

4.1 Esimerkki onnistuneesta koneoppimissovelluksen kehittämisestä

"Tekoälyprojekteissa epäonnistumisen todennäköisyys on suuri, koska käyttöönotto on monimutkaista ja tulokset epävarmoja. Haasteita on käsiteltävä mahdollisimman varhaisessa vaiheessa tavoitelähtöisesti ja toteutettavuustutkimusten avulla. Mahdolliset hyödyt ja käytettävyys tulee varmistaa ennen kuin järjestelmän kehitys käynnistetään."

Tekoälyn käyttöönotossa huomioitavia tekijöitä havainnollistetaan seuraavassa esimerkissä. Esimerkki kuvaa metodologiaa, jolla tekoälyn kehittämistä voidaan yleistäen soveltaa muihinkin kyberturvallisuuden käyttötapauksiin. Esimerkkisovellus on haitallisten suorittavien tiedostojen (P*ortable Executables, PE*) havainnointijärjestelmä.

Hyvä tekoälyn kehitysprosessi alkaa sovelluksen kehittämisen liiketoimintatavoitteiden määrittelystä. Tässä vaiheessa tulisi keskustella tekoälypohjaisen ratkaisun tarpeesta ja mahdollisista muista vaihtoehdoista. Tekoälypohjaisten tietoturvaratkaisujen hyötyjä ja haasteita tulisi punnita huolellisesti. PE-havainnointijärjestelmälle tarvittiin tekoälypohjainen ratkaisu tiedostojen analysoitavan määrän, haitallisten PE-tiedostojen monimuotoisuuden ja niiden tunnistamisen monimutkaisuuden vuoksi. Haitallisten PE-tiedostojen tunnistaminen oli ominaisuus, joka oli jo tarjolla ulkoisen toimittajan tuotteessa. Tämän uuden ratkaisun liiketoimintatavoitteet olivat seuraavat: 1) itsenäistyminen ulkoisesta toimittajasta ja 2) haitallisten PE-tiedostojen havaitsemisen suorituskyvyn ja tarkkuuden parantaminen. Tavoitteiden perusteella määriteltiin konkreettiset mittarit, joilla arvioitiin järjestelmän tavoitesuorituskykyä. Mittareiden perusteella asetettiin menestyskriteerit, joita vasten järjestelmää tulisi säännöllisesti arvioida. PE-havainnointijärjestelmän arviointimittareita olivat käyttökustannukset, tunnistamisen kokonaistarkkuus, väärien positiivisten määrä ja kattavuus. Menestyksen määrittävät kynnysarvot määriteltiin nykyisen, ulkoisen toimittajan tuotteen suorituskyvyn perusteella. Liiketoimintatavoitteiden määrittely ja niiden kääntäminen avainluvuiksi (*Key Performance Indicator, KPI*) selkeillä tavoitearvoilla on ratkaisevan tärkeää, koska sillä voidaan varmistaa tekoälyjärjestelmän käyttöönoton liiketoiminnalliset hyödyt.

Liiketoiminnan tavoitteiden määrittelyn jälkeen tulee määritellä vaatimukset tuotantoon siirtymiselle. Toteutettavuustutkimuksella varmistetaan datan saatavuus, ratkaisun integrointi olemassa oleviin prosesseihin, olemassa olevien järjestelmäkomponenttien muutostarpeet ja mahdollisuudet, hyväksyttävät toimintakustannukset ja muut merkittävät tekijät. Tekoälyputken eri vaiheet on määriteltävä. Näitä ovat datan keruu ja käsittely, puhdistus, koulutus, validointi, käyttö, seuranta ja kehittäminen. Jokainen vaihe on priorisoitava sen mukaan, kuinka tärkeä se on käyttöön otettavan koneoppimista hyödyntävän järjestelmän saavuttamiseksi. Esimerkiksi datan keruu ja koulutus voidaan määrittää pakollisiksi vaiheiksi, kun taas puhdistus, seuranta ja kehittäminen voidaan jättää pois. Jokaiselle vaiheelle tulee löytää teoriassa toteutettava ratkaisu, missä tunnistetaan myös mahdolliset ongelmat. PE-havainnointijärjestelmän vähimmäisvaatimukset toimivan tekoälyjärjestelmän käyttöönotolle olivat datan keruu, datan käsittely, ominaisuuspoiminta, koulutus ja validointi. Jokainen näistä vaiheista tarvitsi

realistisen ja käyttöön otettavan ratkaisun. Jos kriittisen vaiheen toteuttaminen ei näytä mahdolliselta suunnitteluvaiheessa, projekti tulisi lopettaa, jotta vältytään panostamasta käyttökelvottomaan ratkaisuun. Usein organisaatiossa toistuvat samat ongelmat eri tekoälyprojekteissa. Toistuvat ongelmat pitää tunnistaa ja niitä on seurattava, jotta voidaan kehittää prosesseja sekä työkaluja johdonmukaiseen ongelmanratkaisuun.

Tekoälyprojekteissa epäonnistumisen todennäköisyys on suuri, koska käyttöönotto on monimutkaista ja tulokset epävarmoja. Haasteita on käsiteltävä mahdollisimman varhaisessa vaiheessa tavoitelähtöisesti ja toteutettavuustutkimusten avulla. Mahdolliset hyödyt ja käytettävyys tulee varmistaa ennen kuin järjestelmän kehitys käynnistetään.

Seuraavassa vaiheessa kehitystyötä arvioidaan liiketoiminnan tavoitteiden saavutettavuutta tekoälyratkaisulla. Prototyyppi on rakennettava nopeasti, mutta datan keräämiseen ja prototyypin kouluttamiseen ja testaamiseen on kiinnitettävä paljon huomiota. On tärkeää kerätä kokeeksi pieni datajoukko, joka edustaa tuotantojärjestelmän analysoitavana olevan datan kokonaisuutta. Tätä kokeellista datajoukkoa käytetään koneoppimista hyödyntävän mallin kouluttamiseen ja validointiin aiemmin määriteltyjen tavoitelukujen mukaisesti. PE-havainnointijärjestelmää varten kerättiin dataa erilaisilta organisaatioilta eri maantieteellisiltä alueilta yhtäjaksoisesti. Koulutukseen ja validointiin käytetty data jaettiin keruuaikojen mukaan siten, että vanhempaa dataa käytettiin mallin kouluttamiseen ja uudempaa dataa sen tulosten validointiin. Kehitysvaiheen aikana datan keruun ja valmistelun tulee vastata todellista tuotantoympäristöä, jotta suorituskykytulokset ovat järkeviä ja lähellä varsinaisen järjestelmän odotuksia. Vaikka suorituskykytuloksissa tulee todennäköisesti olemaan eroja tämän vaiheen huolellisenkin toteutuksen jälkeen, on erojen minimoiminen tärkeää.

Tekoälyjärjestelmä tulee siirtää tuotantoon mahdollisimman pian sen jälkeen, kun datan kerääminen on hoidettu ja ensimmäinen prototyyppi järjestelmästä on kehitetty. Tuotantoon siirrytään kehittämällä tekoälyputken pakolliset vaiheet vähimmäisvaatimusten mukaisiksi. Tässä vaiheessa tekoälyjärjestelmän tulisi ihanteellisesti tuottaa sellaisia tuloksia, joita valmiskin järjestelmä tuottaa, mutta tuloksia ei tule vielä hyödyntää päätöksenteossa. Tuloksia tulisi käyttää kustannuksen ja tarkkuuden tavoitelukujen laskemiseen, jotta koneoppimisen mallia voidaan iteratiivisesti kouluttaa, validoida ja parantaa tavoitearvoja vastaaviksi.

Aluksi on suositeltavaa ottaa koneoppimisen malli käyttöön rajoitetulla datamäärällä, esimerkiksi 1-10 prosentilla normaalista datavirrasta, ja lisätä sen kuormitusta asteittain, kunnes järjestelmä saavuttaa täyden kuormituksen. Tämä mahdollistaa kustannussäästöt mallin hienosäädön aikana ja poistaa skaalaustekijän suorituskyvyn alustavasta arvioinnista. Tuotantoon siirtyminen aikaisessa vaiheessa mahdollistaa kehitys- ja testausympäristöissä ilmenemättömien ongelmien nopean tunnistamisen. Aikainen tuotantoon siirtyminen on hyödyllistä kahdesta syystä:

1) se mahdollistaa merkittävien ongelmien tunnistamisen, joiden vuoksi malli olisi käyttökelvoton, ja projekti voidaan lopettaa aikaisessa vaiheessa,

2) pakollisia toimintoja testaamalla voidaan tunnistaa, miten eri vaiheiden ongelmat vaikuttavat toisiinsa, esimerkiksi datan kerääminen puhdistukseen tai

kouluttamiseen. Testauksen myötä toiminnallisuuden edellytyksiä voidaan arvioida uudelleen vaatimustason määrittelyssä.

Edellä kuvattua prosessia sovellettiin PE-tunnistusjärjestelmän käyttöönotossa. Suorituskykyä seurattiin ja visualisoitiin testituotannon varhaisessa vaiheessa. Seuranta ja visualisointi mahdollistivat tavoitelukujen kattavuuden arvioinnin ja tarkkuuden laskemisen, sekä samaan tuotantodataan perustuvan vertailun tekoälypohjaisen PE-tunnistusjärjestelmän ja ulkoisen toimittajan järjestelmän välillä. Tulosten vertailun perusteella kunkin järjestelmän vahvuudet ja soveltamisalat voitiin tunnistaa. Lisäksi voitiin määritellä tarkkuus, käyttökustannukset ja latenssi huomioiden, voiko järjestelmiä yhdistää sarjaan vai pitäisikö niitä ajaa rinnakkain. Tietoturva-asiantuntijat käyttivät tulosten visualisointia manuaaliseen analysointiin silloin, kun kahden järjestelmän päätökset erosivat toisistaan. Tästä saatua palautetta käytettiin AI-järjestelmän kehittämiseksi. Tätä kehittämistä kutsutaan ihmisen avustamaksi vahvistusoppimiseksi. Suorituskyvyn vertailua tehtiin riippumattoman testausorganisaation suorittamilla testeillä, joita käytetään yleisesti kyberturvallisuusalalla eri tuotteiden suorituskyvyn vertailuun ja arviointiin.

Tekoälypohjaisen PE-tunnistusjärjestelmän tuloksia aloitettiin hyödyntämään päätöksenteossa haitallisten PE-tiedostojen tunnistamisessa sen jälkeen, kun se oli saavuttanut tavoitelukunsa ja pystyi toimimaan täydellä kuormalla hyväksyttävin kustannuksin. Tekoälyputken valinnaisia vaiheita, kuten datan muutosilmiön havaitsemista ja järjestelmän parantamista palautteen avulla, alettiin kehittämään järjestelmän suorituskyvyn ja ylläpidettävyyden parantamiseksi.

Tekoälyprojektin käyttöönoton jälkeen on tärkeää keskittyä helppoon ja edulliseen ylläpitoon. Järjestelmän suorituskyky voi heikentyä nopeasti, ja tällaisessa tilanteessa ongelmiin pitää pystyä puuttumaan nopeasti. Kun organisaatio haluaa kehittää, testata ja ottaa käyttöön enemmän koneoppimispohjaisia järjestelmiä, on tärkeää harkita prosessien ja kirjastojen kehittämistä, sillä ne tukevat järjestelmien kehitystä ja ylläpitoa. On suositeltavaa luoda yhteinen tietopankki mallien nopeaa käyttöönottoa ja jaettua suorituskyvyn seurantatoimintoa varten. Datan kerääminen yhtenäiseen tallennusjärjestelmään mahdollistaa datan paremman saatavuuden myös tekoälyn kehittämisen ja toiminnallisuuden kannalta. Edellä eritellyt toimintatavat parantavat tekoälyjärjestelmien kehityksen, testauksen, käyttöönoton ja ylläpidon tehokkuutta kustannuksia vähentäen.

4.2. Edellytykset tekoälyn soveltamiselle

Alla on listattu suosituksia tekoälyn kehittämiseen kyberturvallisuussovelluksissa. Suositukset perustuvat edellä annettuun esimerkkiin onnistuneesta tekoälyn soveltamisesta sekä järjestelmiä kehittävien data-analyytikoiden ja kyberturvallisuusasiantuntijoiden palautteeseen. Suositukset vedetään yhteen kuvaajassa 3.

Selkeä ongelmanasettelu: Organisaation on määriteltävä selkeästi, minkä ongelman se haluaa ratkaista, mihin tekoälypohjaista ratkaisua tarvitaan ja mitä hyötyä siitä voisi olla. Tekoälyä ei tule valita sen ympärillä yleisesti olevan innostuksen tai sen markkinoinnillisen arvon perusteella, tai yksin siksi, että organisaatiolla on paljon käyttämätöntä tai alikäytettyä dataa. On tärkeää ymmärtää, mitä haasteita ongelman ratkaisemisessa on ja mitkä ovat tekoälyn

tarjoamat mahdollisuudet. Tyypillisesti kyberturvallisuuden sovelluksilta vaaditaan erittäin korkeaa tarkkuutta tai alhaista väärien positiivisten määrää. On arvioitava, voiko tekoälypohjainen ratkaisu täyttää nämä vaatimukset.

Soveltamiskohteen kriittisyyden arviointi: Tekoälyä hyödyntävät järjestelmät tekevät aina virheitä. On tärkeää arvioida, kuinka kriittisiä AI-järjestelmän päätökset ovat. On parempi käyttää tekoälyratkaisuja vähemmän kriittisiin sovelluksiin ja varmistaa, että ihmisten läsnäoloa lisätään käyttötapauksen kriittisyyden kasvaessa. Näin voidaan ennaltaehkäistä vakavat epäonnistumiset. Aivan ensimmäisten projektien aikana tekoälyä kannattaa soveltaa vain ei-kriittisiin sovelluksiin.

Tekoälyn suorituskyvyn yhdistäminen liiketoiminnan tavoitteisiin: Liiketoiminnan tavoitteet on tunnistettava ja yhdistettävä mitattaviin tekoälyn suorituskykymittareihin projektin alusta alkaen. Mittareita voivat olla esimerkiksi tulosten tarkkuus, väärien positiivisten määrä, latenssi ja kustannukset. Suorituskykymittareita on arvioitava mahdollisimman varhaisessa vaiheessa ja niitä on laskettava säännöllisesti uudelleen projektin kehityksen aikana.

Käyttöönoton vaatimusten määrittely: Organisaation tulee tutkia käyttöönoton toteutettavuutta, tekoälyjärjestelmän riippuvuuksia, sen integrointia olemassa oleviin järjestelmiin ja olemassa olevien osa-alueiden muuttamisen tarvetta ja mahdollisuuksia jo ennen projektiin ryhtymistä. Organisaation on varmistuttava siitä, että tekoälyjärjestelmän käyttöönotto on toteutettavissa ja että kaikki riippuvuudet voidaan ratkaista realistisesti budjetin puitteissa. Monet turvallisuuteen liittyvät tekoälyprojektit eivät koskaan päädy tuotantoon integrointi- ja käyttöönotto-ongelmien vuoksi.

Datan merkityksen, saatavuuden ja laadun varmistaminen: On yleinen virhe aloittaa tekoälyprojekti yksinkertaisesti siksi, että dataa on saatavilla. Projekti tulee aloittaa ongelmasta, jonka perusteella määritellään data, jota sen ratkaisemiseksi tarvitaan. Sen jälkeen tulee kehittää kyky kerätä tätä määriteltyä dataa. Datan laadun, johdonmukaisuuden ja edustavuuden varmistaminen on ensiarvoisen tärkeää kehittämiselle, varmistamiselle ja tuotannolle. Tekoälyprojektin menestymisen edellytys on korkealaatuisen datan käyttö.

Datan kehityksen ymmärtäminen: Tuotantoympäristöissä datalle tapahtuvien muutosten vuoksi koneoppimisjärjestelmän suorituskyky voi muuttua. Turvallisuusprosesseihin syötettävä data on erittäin altis muutoksille, jota kutsutaan yleisesti datan muutosilmiöksi (*data drift*). Organisaation tulee seurata datan jakautumista ja ymmärtää, kuinka nopeasti se muuttuu. Muutosten hallitsemiseksi ja suorituskyvyn heikkenemisen estämiseksi tulee kehittää ratkaisuja, joista paras on usein koneoppimismallin säännöllinen uudelleenkoulutus tuoreella datalla. Uudelleenkoulutuksen tiheyden on perustuttava datan muutosilmiön nopeuteen ja uuden datan hankkimisen ja merkitsemisen kustannuksiin.

Vältä turhaa monimutkaisuutta: Organisaatioiden ei kannata turhaan lähteä mukaan syväoppimisen ja suurten kielimallien ympärillä olevaan intoiluun. Organisaation tulee aina valita ongelman ratkaisuun kykenevistä tekoälyratkaisuista yksinkertaisin. Monimutkaiset tekoälyratkaisut tuovat aina

mukanaan haittoja, kuten korkeita käyttökustannuksia, suuria datatarpeita ja vaikeuksia ymmärtää tekoälyn tekemiä päätöksiä.

Aikainen käyttöönotto: Testivaiheesta tuotantoympäristöön siirtymisen jälkeen tulee usein ilmi vielä ratkaisemattomia ongelmia. Tuotantoympäristöissä on uusia rajoituksia ja haasteita. Esimerkiksi, tuotantoympäristössä saatavilla oleva data on lähes aina erilaista kuin kehityksen ja validoinnin aikana saatavilla oleva data. Skaalautuvuus- ja latenssiongelmien lisäksi dataan liittyvät haasteet voivat johtaa tarkkuuseroihin. Kehittämällä aikaisin tuotantoon saatava ja minimivaatimuksia vastaava tuote on mahdollista ratkaista ongelmia jo tuotannossa eikä vain testiympäristössä.

Joustavuus käyttöönoton ja tulosten suhteen: Tekoälyjärjestelmän antamia tuloksia tulee arvioida avoimesti. Tekoälyprojektin tulokset pysyvät epävarmoina loppuun asti, ja eri vaihtoehtoja tulisi harkita siitä saatavien hyödyn varmistamiseksi. Tekoälyjärjestelmät ovat harvoin niin tarkkoja, että niitä voitaisiin käyttää itsenäisinä järjestelminä osana kriittistä päätöksentekoa. Käyttöönottoa joko rinnakkain tai putkessa muiden järjestelmien kanssa tulisi harkita, kunkin järjestelmän vahvuuden ja laajuuden perusteella. Esimerkiksi, ensimmäisenä putkessa voitaisiin käyttää nopeaa ja edullista tekoälyjärjestelmää, jonka tarkkuus on alhainen. Ensimmäistä järjestelmää hyödynnettäisiin vain niissä tapauksissa, kun sillä voidaan saavuttaa korkealla varmuudella oikeita tuloksia. Järjestelmä lähettäisi muut tapaukset tarkemmalle järjestelmälle, joka voi olla kalliimpi ja hitaampi. Lisäksi tekoälyjärjestelmien tekemiin päätöksiin tulisi suhtautua mukautuvasti ja tekoälyn tekemät toimenpiteet tulisi määritellä tarkkuuden ja kriittisyyden mukaan. Esimerkiksi alhaisen tarkkuuden tekoälyjärjestelmiä ei tulisi käyttää kriittisten päätösten tekemiseen tai tekemään toimenpiteitä, jotka heikentävät käyttäjäkokemusta ja käytettävyyttä.

Käsittely- ja laskentakustannusten huomiointi: Kyberturvallisuussovelluksissa käytetty tekoäly vaatii usein kouluttamiseen ja ennustamiseen käytettävien suurten datamäärien rinnakkaista käsittelyä. Monimutkaisten koneoppimismallien, kuten DNN:ien ja suurten kielimallien kouluttaminen voi olla laskennallisesti erittäin intensiivistä, minkä lisäksi on otettava huomioon, että koulutusta tulee toistaa säännöllisesti datan muuttumisen vuoksi. Datan käsittelyyn ja laskentaan liittyvät kustannukset voivat nousta pilvipalveluissa erittäin korkeiksi, kun huomioidaan käytetyn laitteiston hinta (esim. GPU). Kustannukset tulee arvioida ja optimoida aikaisessa vaiheessa.

Kehitä moniosaamista: Tekoälyratkaisuihin liittyvät tehokas ongelmanratkaisu edellyttää teknistä osaamista sekä itse ratkaisuissa että niiden soveltamisalueilla. Tehokkain tapa odotustenmukaisten ja tehokkaiden tekoälyratkaisujen kehittämiseksi on ottaa mukaan datatieteilijöitä, joilla on syvä ymmärrys kyberturvallisuudesta tai kyberturvallisuusasiantuntijoita, joilla on taustaa datatieteessä. Tällaista moniosaamista ei ole yleensä helposti saatavilla työmarkkinoilla, joten osaamista on luotava talon sisäisesti koulutusohjelmien avulla. Lähtökohtana koulutusohjelmissa on tarjota yleinen kyberturvallisuuskoulutus datatieteilijöille ja datatieteen koulutus tietoturva-asiantuntijoille, jotta asiantuntijoilla olisi yhteinen kieli ja he voisivat tehdä tehokkaampaa yhteistyötä.

Työkalut ja prosessit toistuviin tehtäviin: Ajallisesti kehitys tekoälyprojektin konseptin ja käyttöönoton välillä on melko pitkä. Jos organisaatio aikoo kehittää useita tekoälyjärjestelmiä, sen on systematisoitava prosesseja ja kehitettävä kirjastoja toistuvien, yhteisten tehtävien helpottamiseksi ja automatisoimiseksi. Esimerkiksi kirjastot ja järjestelmät datan lataamiselle, kirjastot tuotantoon viennille, yhteisen sovellusohjelmointirajapinnan tekoälyjärjestelmille, kirjastot havainnoimiselle ja alustat toimintakyvyn visualisoimiseksi sopivat tähän tarkoitukseen.

Kuvaaja 3: Onnistuneen tekoälysovelluksen edellytykset projektin vaiheiden mukaan

5. Tekoälyn tulevaisuus kyberturvallisuudessa

Tekoälyn käyttöönottoa kyberturvallisuuden sovelluksissa on vauhdittanut hyökkäysten kasvava monimutkaisuus, nopeus ja mittakaava. Tekoälyyn kohdistuvan mielenkiinnon ja tekoälyteknologioiden saatavuuden kasvun perusteella trendi tulee jatkumaan myös tulevaisuudessa. Soveltamalla tekoälyä oleellisiin ongelmiin voidaan saavuttaa kestäviä hyötyjä. Nykyisiin ja tuleviin tekoälysovelluksiin liittyy aiemmin vähemmälle huomiolle jääneitä eettisiä, teknisiä ja sääntelyyn liittyviä haasteita, jotka voivat hidastaa sovelluskehitystä.

5.1 Suurten kielimallien soveltaminen

Suurten kielimallien kyky ymmärtää ja tuottaa luonnollista kieltä helpottaa merkittävästi tekoälyn tarjoamien mahdollisuuksien tutkintaa. Uusi liikkuma-ala luo runsaasti erilaisia uusia mahdollisuuksia luovaan ja innovatiiviseen työskentelyyn kyberturvallisuuden sovelluksissa. Kokeellisten ideoiden saattaminen tuotantoon kelpaaviksi ratkaisuiksi vaatii kuitenkin asiantuntijoiden osallistumista. Suuret kielimallit tarjoavat mahdollisuuksia etenkin sellaisille organisaatioille, jotka eivät vielä ole kypsiä soveltamaan tekoälyä kybertuvallisuutensa parantamisessa.

Suuret kielimallit eivät välttämättä paranna merkittävästi esimerkiksi uhkien havaitsemiseen tai päätelaiteturvallisuuteen tarkoitettuja kehittyneitä sovelluksia, varsinkaan mitä tulee havaitsemiskyvykkyyksiin. Suurten kielimallien kyky yhdistellä monimutkaisia syötteitä ja tuottaa helposti ymmärrettäviä tuloksia voi kuitenkin lisätä läpinäkyvyyttä ja ymmärrettävyyttä epäselvissä päätöksentekoprosesseissa. Järjestelmän käytettävyyttä voi parantaa perustelemalla käyttäjälle toimintoja, esimerkiksi sen, miksi tiedoston avaaminen tai verkkosivuilla vierailu on estetty. Lisäksi kielimalleja voi käyttää havaitsemisjärjestelmien kehittämiseen tarjoamalla näkemyksiä virheiden, kuten väärien positiivisten ja väärien negatiivisten, syistä. Suuret kielimallit voivat myös potentiaalisesti tukea perinteisempiä turvallisuusratkaisuja, joita käytetään uhkien havaitsemisessa ja päätelaiteturvallisuudessa, kuten havaitsemissääntöjen määrittelyä sääntömoottoreille. Turvallisuusanalyytikot voisivat muotoilla tutkimuksiinsa perustuvia havaitsemissääntöjä käyttäen luonnollista kieltä ja syöttää ne suuriin kielimalleihin, jotka kääntäisivät ja tuottaisivat muotoiltuja sääntöjä, joita voidaan soveltaa sääntömoottorilla. Suuret kielimallit voivat lisätä jo kehittyneempien tekoälyä hyödyntävien kyberturvallisuusratkaisujen automaatiota ja käytettävyyttä.

Kielimallien kyky käsitellä ja yhdistellä suuria määriä monimuotoista tietoa voi olla hyödyksi uhkatiedustelussa ja haavoittuvuudenhallinnassa, joissa kerättävän ja analysoitavan tiedon määrä on valtava. Suuret kielimallit voivat kerätä tietoja sekä sisäisistä että ulkoisista tietolähteistä, yhdistää ne ja erottaa olennaisen tiedon uusista organisaatioon vaikuttavista uhista ja haavoittuvuuksista. Turvallisuusanalytiikassa ja tietoturvanhallintakeskuksissa (*Security Operations Centre, SOC*) suuret kielimallit voivat tukea turvallisuushälytysten tutkimusta keräämällä tietoa eri järjestelmistä turvallisuustapahtumien taustoittamiseksi. Iteratiivinen vuorovaikutus mahdollistaa analyytikoiden syventymisen uhkatietoon sekä lisätiedon etsimisen.

Toimimalla sillanrakentajana eri tietolähteiden välillä ja tarjoamalla perusteltuja näkemyksiä suuret kielimallit voivat tarjota merkittävää apua turvallisuusoperaatioissa toimiville ihmisille. Tämä muutos näkyy uusissa turvallisuustiimejä avustavissa tuotteissa[20], mikä osoittaa, että suuret kielimallit ovat integroitumassa turvallisuusprosesseihin. Suuret kielimallit voivat myös vastata ja lieventää tutkittujen uhkien vaikutuksia. Suurten kielimallien luotettavuuden lisääntyessä tutkimusten ja toimenpiteiden automatisoituminenkin voi lisääntyä, mikä vähentää ihmisen tarvetta puuttua prosesseihin.

[20] *Microsoft Secutity Copilot.* https://www.microsoft.com/en-us/security/business/ai-machine-learning/microsoft-security-copilot

Turvallisuuskoulutus on erityisen lupaava käyttötapaus suurille kielimalleille. Samalla tavalla kuin aiemmin mainittu tuki tietoturvanhallintakeskuksille, suuret kielimallit voivat kouluttaa turvallisuusanalyytikoita hälytysten tutkimukseen. Turvallisuustapahtuman yhteydessä kielimallit voivat ehdottaa toimia, kuten tietojen keräämistä ja korrelaatiota ja tarkasteltavia indikaattoreita. Suuret kielimallit voivat myös opettaa ja parantaa turvallisuuskäytäntöjen soveltamista tavallisille kehittäjille ja järjestelmänvalvojille. Koodausavustajat voivat opettaa turvallista koodausta ja varmistaa turvallisen ohjelmistokehityksen periaatteiden soveltamisen ohjelmistokehityksen aikana. Esimerkiksi jotkut koodausavustajat sisältävät turvallisuusskannausominaisuuden, joka löytää ja ennaltaehkäisee mahdollisia haavoittuvuuksia koodissa. [21]

Suuret kielimallit voivat myös avustaa monimutkaisten järjestelmien, kuten pilvipalveluiden, turvallisessa määrityksessä. Pilvipalveluntarjoajat toteuttavat korkeatasoista tietoturvaa infrastruktuurissaan, mutta virheet käyttöönoton määrityksessä on merkittävä tietoturvariski, koska asianmukaisten asetusten määrittäminen on monimutkaista. Suuret kielimallit voivat ohjata järjestelmänvalvojia vuorovaikutteisesti määritysprosessin läpi, tarjota tietoja asetuksista ja ehdottaa tarpeiden perusteella määriteltyjä arvoja. Suuret kielimallit voivat kouluttaa myös IT-järjestelmien ei-asiantuntijakäyttäjiä tietoturvariskien havaitsemiseen. Suuria kielimalleja käytetään jo nyt esimerkiksi simuloimaan edistyneitä tietojenkalasteluhyökkäyksiä edistäen loppukäyttäjien kykyä havaita ja välttää niitä.

Kuvaaja 4 esittää alustavan ennusteen suurten kielimallien käyttöönotolle kyberturvallisuusratkaisuissa. Niiden ensimmäinen pääsovellus on todennäköisesti koulutuksellinen, joka mahdollistaa kokeilun ja perustason turvallisuuskäytäntöjen opettamisen tarvittavassa mittakaavassa. Varhaiset sovellukset kattavat myös tuen turvallisuusanalytiikalle, joka mahdollistaa turvallisuustapahtumien taustoittamisen ja toimenpiteiden suosittelemisen. Ajan myötä suuret kielimallit tukevat monimutkaisempia tehtäviä, jotka liittyvät uhkatiedusteluun ja haavoittuvuudenhallintaan, kuten tunnettujen haavoittuvuuksien löytämiseen järjestelmissä ja koodeissa tunkeutumistestauksen tukena. Pitkällä aikavälillä suuret kielimallit saattavat pystyä käsittelemään korkeampaa asiantuntemustasoa vaativia tehtäviä, ja kyetä löytämään täysin uusia haavoittuvuuksia, arvioimaan organisaatioiden turvallisuuden tasoa tai antamaan toimenpide-ehdotuksia monimutkaisten poikkeamien hallintaan.

[21] *Amazon CodeWhisperer.* https://aws.amazon.com/codewhisperer/

Ennuste suurten kielimallien käytöstä kyberturvallisuudessa

Kuvaaja 4

5.2 Riskit, uhat ja tulevaisuuden haasteet

Tekoälyn lisääntynyt sovelluskäyttö on paljastanut erilaisia uhkia ja vaaroja. Viime aikoina on kiinnitetty huomiota tekoälyjärjestelmien vinoumiin, huonoon selitettävyyteen, ja turvallisuus- ja yksityisyysongelmiin. Tekoälyjärjestelmät ovat alttiita uusille, vihamielisiksi koneoppimishyökkäyksiksi kutsutuille tietoturvariskeille, jotka vaikuttavat vain tekoälyyn perustuviin järjestelmiin. Huolestuttavia nousevia uhkia ovat esimerkiksi mallin saastutus (*model poisoning*) ja mallin kiertäminen (*model evasion)*, jotka vaarantavat tekoälyjärjestelmien tiedon eheyden ja niiden tekemien ennusteiden luotettavuuden. Mallin saastutus on hyökkäys, jossa hyökkääjä syöttää tai muokkaa tekoälymallin koulutusdataa tai koulutuslogiikkaa manipuloidakseen sen ennusteita. Mallin kiertämisessä taas vastapuoli rakentaa haitallisia syötteitä lähetettäväksi tekoälyjärjestelmään päätöksenteon aikana tuottaakseen vääriä ennusteita.

Kyberhyökkäyksissä tullaan todennäköisesti käyttämään turvallisuusuhkien havaitsemiseen ja ehkäisemiseen tarkoitettuja tekoälyjärjestelmiä huijauksiin ja puolustusten kiertämiseen. Tekoälyjärjestelmien, joita käytetään kyberturvallisuuteen, on varmistettava, että ne ovat yhtä turvallisia ja joustavia hyökkäyksille kuin ei-tekoälypohjaiset järjestelmät. Vaikka todellisia vihamielisiä koneoppimishyökkäyksiä on havaittu vain vähän, ne tulevat todennäköisesti lisääntymään tulevaisuudessa.

Yksityisyysongelmat ovat myös kasvava huolenaihe, sillä erityisesti generatiivisen tekoälyn mallit ja suuret kielimallit ovat alttiita vuotamaan koulutusdataa. Koulutusdataan liittyviä ongelmia voidaan torjua asettamalla rajoituksia tekoälyn kouluttamisessa käytettävälle tiedolle tai vaatimalla lisäsuojauksia sen varmistamiseksi, ettei tietovuotoja tapahdu.

Uusissa tekoälyteknologioissa on omat ongelmansa. Vaikka perinteiset tekoälymallit ovatkin erittäin luotettavia tietyissä tehtävissä, suurilla kielimalleilla on enemmän ongelmia tulosten laadun ja luotettavuuden kanssa. Hallusinaatio on suurten kielimallien erityinen ongelma. Hallusinoinnilla tarkoitetaan sitä, että malli tuottaa mahdolliselta tai koherentilta vaikuttavaa virheellistä tietoa. Luotettavuuden puute rajoittaa automatisoitujen päätösten käyttöönottoa uhkien havaitsemisessa ja ehkäisemisessä, sillä niissä tarvitaan korkeaa tarkkuutta. Hallusinaatio on huolenaihe myös sellaisten sovellusten käytössä, joissa käyttäjien on voitava luottaa tarkkaan tietoon. Tällaisia ovat esimerkiksi turvallisuustapahtumien tutkimus tai turvallisuuskoulutus. Niin kauan kuin hallusinoinnin luomia ongelmia ei voida ratkoa, suurten kielimallien soveltaminen rajoittuu sellaisiin käyttötapauksiin, joissa ihmisasiantuntijat pystyvät nopeasti tarkistamaan tuotettujen tulosten pätevyyden.

Tekoälyn laaja käyttöönotto kyberturvallisuusratkaisuissa edellyttää järjestelmien kehittämistä luotettaviksi, turvallisiksi, yksityisyydensuojaa kunnioittaviksi, oikeudenmukaisiksi ja läpinäkyviksi. Kehittyminen kaikilla osa-alueilla on kuitenkin pitkä prosessi. Sillä välin on toteutettava järkevää tekoälyä koskeva riskienhallintaprosessia parhaiden tutkittujen käytäntöjen mukaisesti, jotta turvallisuusriskit voidaan tunnistaa ja niihin voidaan puuttua. Tämä on haastavaa, mutta välttämätöntä, kun otetaan huomioon, että tekoäly on todennäköisesti ainoa ratkaisu vastata tekoälypohjaisiin kyberhyökkäyksiin.[22] Tekoäly tulee

[22] Traficom. *The security threat of AI-enabled cyberattacks.* 2022

todennäköisesti olemaan uusi ase jatkuvassa varustelukilvassa hyökkäyksen ja puolustuksen välillä.

5.3 Tekoälyn sääntely ja standardisointi

"Yhteisten standardien puuttuessa tekoälyä hyödyntävien organisaatioiden tulisi luoda vahvat kehykset, käytänteet ja ohjeistukset tekoälyn käyttämiseen. Näin varmistetaan tekoälyn eettinen, turvallinen ja tehokas integraatio järjestelmiin, ennaltaehkäistään haittoja sekä priorisoidaan tekoälystä saatavaa hyötyä."

Tekoälyn soveltaminen kyberturvallisuusratkaisuissa on ollut tähän mennessä avointa ja sitä on rajoitettu vain vähän. Tekoälyä voidaan käyttää käytännössä mihin tahansa kyberturvallisuuden sovellukseen, jos järjestelmille syötettävät tiedot kerätään ja käsitellään tietosuojasääntelyn, esimerkiksi EU:n yleisen tietosuoja-asetuksen eli GDPR:n mukaisesti.[23] Tietosuojasääntely rajoittaa henkilötietojen käyttöä tekoälysovelluksissa. Se myös rajoittaa tekoälyjärjestelmien kehittämistä ja kokeilua rajoittamalla tietojen saatavuutta. Asiakastietojen kerääminen on sallittua vain, jos olemassa olevan tuotteen tai palvelun toiminnallisuus edellyttää sitä. Kuitenkin tekoälyjärjestelmien kehittämiseen ja testaamiseen tarvitaan tietoja jo ennen kuin niiden tarjoamien lopputulosten hyödyllisyydestä voidaan olla varmoja. Siksi tietojen kerääminen voi vaatia asiakkaiden nimenomaisen suostumuksen silloin, kun ne eivät vielä palvele muuta tarkoitusta kuin järjestelmän kehittämistä.

Tekoälyn eettistä ja turvallista käyttöä koskevaa sääntelyä luodaan ja otetaan käyttöön paraikaa. Joillakin teollisuudenaloilla, kuten rahoitus- ja autoteollisuudella, on omaa sääntelyä tekoälyn käytöstä, mutta kyberturvallisuusala kuuluu yleisen tekoälylainsäädännön piiriin. Joitakin esimerkkejä tekoälysääntelystä ovat EU:n tuleva tekoälyasetus[24], tai Yhdysvalloissa ehdotettu tekoälylainsäädäntö (AI Bill of Rights). [25] Euroopan unioni on edistynein tekoälyn käytön sääntelyssä, myös kyberturvallisuussovelluksien osalta. Tekoälyasetus kieltää todennäköisesti joitakin tekoälysovelluksia ja asettaa vaatimuksia korkean riskin sovelluksille, jotka koskevat kriittisen infrastruktuurin hallintaa, terveydenhuollon tehtäviä, lainvalvontaa tai siirtolaisuuden hallintaa.

EU:n tekoälyasetuksen mukaisen luokittelun mukaan useimmat kyberturvallisuussovellukset kuuluisivat rajoitettuihin tai vähäriskisiin tekoälysovelluksiin, joilta vaaditaan vain läpinäkyvyyttä ja tietojen tarjoamista käyttäjille. On kuitenkin epäselvää, koskisiko korkean riskin sovellusten turvaamisessa käytettyjä tekoälypohjaisia turvallisuusratkaisuja samat vaatimukset, kuin niiden suojaamia sovelluksia. Näihin vaatimuksiin kuuluvat muun muassa riskienhallinta, ihmisen tekemä valvonta, vakaus ja turvallisuus. Kyberturvallisuusalalla oletetaan, että hyökkääjä, joka pyrkii kiertämään ja vaarantamaan puolustusjärjestelmiä on jatkuvasti läsnä, joten tietoturvallisuuden varmistaminen on erityisen tärkeää.

[23] Council of the European Union. *General Data Protection Regulation (GDPR).* 2016
[24] Council of the European Union. *Artificial Intelligence Act (AI act).* 2021
[25] US Office of Science and Technology Policy. *AI bill of rights.* 2023

Tulevaa tekoälysääntelyä ei ole vielä yhdistetty kattaviin teknisiin standardeihin, mikä antaisi tekoälyä hyödyntävälle taholle selkeät ohjeet laillisten vaatimusten täyttämiseksi. Vaikka joitakin aloitteita teknisten standardien määrittelemiseksi tekoälylle onkin olemassa, esimerkiksi ISO/IEC JTC 1/SC 42[26], tämä työ on vasta alkuvaiheessa. Teknisten standardien määrittelyä hidastaa esimerkiksi se, että vielä ei ole olemassa varmaa puolustusta tiettyjä koneoppimista hyödyntäviä hyökkäyksiä vastaan. Ohjeita ja suosituksia on kuitenkin jo saatavilla näiden turvallisuusriskien hallitsemiseksi ja vähentämiseksi, esimerkiksi NIST AI Risk Management Framework[27], the MITRE ATLAS framework[28] UK NCSC:n guidelines for secure AI development[29].

Tekoälysertifioinnin edelleen kehittäminen teknisten standardien perusteella selkeyttäisi tekoälypohjaisten kyberturvallisuusjärjestelmien vaatimustenmukaisuuden varmistamista. Nämä ovat kuitenkin vasta tulevaisuudenkuvia; Yhteisten standardien puuttuessa tekoälyä hyödyntävien organisaatioiden tulisi luoda vahvat kehykset, käytänteet ja ohjeistukset tekoälyn käyttämiseen. Näin varmistetaan tekoälyn eettinen, turvallinen ja tehokas integraatio järjestelmiin, ennaltaehkäistään haittoja sekä priorisoidaan tekoälystä saatavaa hyötyä.

[26] ISO/IEC. *ISO/IEC JTC 1/SC 42.* https://www.iso.org/committee/6794475.html
[27] NIST. *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* 2023
[28] MITRE *ATLAS.* https://atlas.mitre.org/
[29] UK National Cyber Security Center. *Guidelines for secure AI system development. 2023*

Liikenne- ja viestintävirasto Traficom
PL 320, 00059 TRAFICOM
p. 029 534 5000

traficom.fi

ISSN (verkkojulkaisu) 2669-8781
ISBN (verkkojulkaisu) 978-952-311-917-8

TRAFICOM
Liikenne- ja viestintävirasto