Koneoppiminen digitaalisen turvallisuuden teknisessä valvonnassa

Julkisen hallinnon ICT

VALTIOVARAINMINISTERIÖN JULKAISUJA – 2023:56

VALTIOVARAINMINISTERIÖ
FINANSMINISTERIET

Valtiovarainministeriön julkaisuja 2023:56

Koneoppiminen digitaalisen turvallisuuden teknisessä valvonnassa

Niko Mäkilä (toim.)

Valtiovarainministeriö Helsinki 2023

Julkaisujen jakelu
Distribution av publikationer

Valtioneuvoston julkaisuarkisto Valto

Publikations-
arkivet Valto

julkaisut.valtioneuvosto.fi

Julkaisumyynti
Beställningar av publikationer

Valtioneuvoston verkkokirjakauppa

Statsrådets
nätbokhandel

vnjulkaisumyynti.fi

Valtiovarainministeriö
CC BY-SA 4.0

ISBN pdf: 978-952-367-641-1
ISSN pdf: 1797-9714

Taitto: Valtioneuvoston hallintoyksikkö, Julkaisutuotanto

Helsinki 2023

Kuvailulehti
1.8.2023

Koneoppiminen digitaalisen turvallisuuden teknisessä valvonnassa

Valtiovarainministeriön julkaisuja 2023:56		**Teema**	Julkisen hallinnon ICT
Julkaisija	Valtiovarainministeriö		
Toimittaja/t	Niko Mäkilä		
Kieli	suomi	**Sivumäärä**	43

Tiivistelmä

Koneoppiminen tuo uusia mahdollisuuksia digitaalisen turvallisuuden tekniseen valvontaan. Tämä raportti kertoo yleistajuisesti mistä koneoppimisessa on kyse ja kuinka tällaista teknologiaa voidaan käyttää digitaalisen turvallisuuden tekniseen valvontaan. Raportissa kerrotaan myös koneoppimisen hyödyntämiseen vaikuttavasta lainsäädännöstä, tekniikan mahdollisuuksista ja riskeistä, mihin suuntaan koneoppiva valvonta on menossa sekä mitä asioita erityisesti koneoppivan järjestelmän hankinnassa tulisi ottaa huomioon.

Asiasanat julkisen hallinnon ict, tekoäly, koneoppiminen, tietoturva, tietosuoja, riskienhallinta

ISBN PDF	978-952-367-641-1	**ISSN PDF**	1797-9714
Asianumero	VN/14455/2020	**Hankenumero**	VM174:00/2020

Julkaisun osoite https://urn.fi/URN:ISBN:978-952-367-641-1

Presentationsblad
1.8.2023

Användning av maskininlärning vid teknisk övervakning av digital säkerhet

Finansministeriets publikationer 2023:56		**Tema**	Offentliga förvaltningens ICT
Utgivare	Finansministeriet		
Redigerare	Niko Mäkilä		
Språk	finska	**Sidantal**	43

Referat

Maskininlärning ger nya möjligheter till teknisk övervakning av den digitala säkerheten. Denna rapport redogör allmänt för vad maskininlärning handlar om och hur sådan teknik kan användas för teknisk övervakning av den digitala säkerheten. Rapporten redogör också för den lagstiftning som påverkar utnyttjandet av maskininlärning, teknikens möjligheter och risker, i vilken riktning den maskininlärande övervakaren är på väg samt vilka frågor som särskilt bör beaktas vid anskaffning av maskininlärningssystemet.

Nyckelord	offentliga förvaltningens ict, artificiell intelligens, maskinlärning, informationssäkerhet, datasekretess, riskhantering		
ISBN PDF	978-952-367-641-1	**ISSN PDF**	1797-9714
Ärendenummer	VN/14455/2020	**Projektnummer**	VM174:00/2020
URN-adress	https://urn.fi/URN:ISBN:978-952-367-641-1		

Description sheet
1 August 2023

Use of machine learning in technical monitoring of digital security

Publications of the Ministry of Finance 2023:56		**Subject**	Public Sector ICT
Publisher	Ministry of Finance		
Editor(s)	Niko Mäkilä		
Language	Finnish	**Pages**	43

Abstract

Machine learning provides new opportunities for technical monitoring of digital security. This report provides a general overview of what machine learning is about and how such technology can be used for technical monitoring of digital security. The report also describes the legislation affecting the utilisation of machine learning, the opportunities and risks of technology, the direction in which machine learning is going and what matters should be taken into account especially when purchasing machine learning systems.

Keywords public sector ict, artificial intelligence, machine learning, information security, privacy, risk management

ISBN PDF	978-952-367-641-1	**ISSN PDF**	1797-9714
Reference number	VN/14455/2020	**Project number**	VM174:00/2020

URN address https://urn.fi/URN:ISBN:978-952-367-641-1

Sisältö

KONEOPPIMINEN DIGITAALISEN TURVALLISUUDEN TEKNISESSÄ VALVONNASSA

Valtioneuvoston periaatepäätöstä digitaalisesta turvallisuudesta vuodelta 2020 (VM 2020:23) on toteuttanut Haukka-toimeenpanosuunnitelma (VM 2020:33). Yhtenä alkuperäisen suunnitelman tehtävänä oli huolehtia julkisen hallinnon autonomisten ja oppivien järjestelmien valvonnasta ja määrittää siihen tarvittavat turvallisuusperiaatteet ja kontrolliympäristöt. Sekä lainsäädännön että teknologioiden kehitys on ollut sen suuntaista, että tällaista kokonaisuutta ei olisi ollut mahdollista toteuttaa.

Sen sijaan päätimme laatia julkisen hallinnon käyttöön selvityksen koneoppivien järjestelmien käyttämisestä digitaalisen turvallisuuden teknisessä valvonnassa. Selvitys pyrkii myös kertomaan yleistajuisesti mistä koneoppimisessa ja sitä hyödyntävissä järjestelmissä on kyse.

Tämän selvityksen kirjoittamisessa ei ole hyödynnetty tekoälyä tai koneoppimista.

Niko Mäkilä, neuvotteleva virkamies

Elokuu 2023

1 Johdanto

Tekoäly ja sen kyvykkyydet ovat olleet viime aikoina voimakkaasti pinnalla yhteiskunnallisessa keskustelussa. Teknologia-alan tutkimusyritys Gartnerin toukokuussa 2023 julkaiseman kyselytutkimuksen mukaan tekoälyn koetaan olevan murroksellisin eri toimialoihin tällä hetkellä vaikuttava teknologia (1). Tämä näkyy myös tekoälyyn tehtävien investointien määrässä, sillä tekoälyä hyödyntävien ohjelmistomarkkinoiden koon on ennustettu saavuttavan yli 120 miljardia dollaria vuoden 2025 aikana (2). Yritysten näkökulmasta tekoälyn uskotaan auttavan sekä nykyisten liitetoimintamallien kehittämisessä että kokonaan uusien palveluiden synnyttämisessä. Yksityishenkilöille ajankohtaisin tekoälysovellutus tämän selvityksen kirjoitushetkellä keväällä 2023 on yhdysvaltalaisen OpenAI-tekoälytutkimuskeskuksen kehittämä ChatGPT-keskustelubotti, jonka tuottamat kirjalliset tuotokset ovat olleet laajasti esillä niin uutisoinnissa, sosiaalisessa mediassa kuin työpaikkojen kahvihuonekeskusteluissa.

Tekoälyn tutkimuksessa ja kehityksessä saavutetut edistysaskeleet ovat herättäneet tutkijoiden ja yhteiskunnallisten vaikuttajien keskuudessa myös huolta. Ihmiskunnan olemassaoloa uhkaavien riskien ennakointiin ja hallintaan keskittynyt Future of Life Institute julkaisi maaliskuussa 2023 avoimen kirjeen, jonka pääviestinä oli vaatimus tietyn tekoälyä koskevan kehityksen hidastamisesta (3). Kirjeen taustalla on riski siitä, että pian hallussamme saattaa olla kyvykkyyksiä, jotka eivät enää ole hallinnassamme ja joilla voi olla ennalta arvaamattomia negatiivisia seurauksia koko maapallolle.

Tekoälyn mukanaan tuoma murros koskee myös digitaalista turvallisuutta. Maailman talousfoorumi listasi vuonna 2020 ilmestyneessä kyberturvallisuutta ja uusia teknologioita koskevassa raportissaan tekoälyn ja edistyneen koneoppimisen yhdeksi neljästä kybertoimintaympäristöön tulevaisuudessa vaikuttavasta teknologiasta (4). KPMG:n vuonna 2022 yritysten johtohenkilöille teettämässä kyselytutkimuksessa puolestaan lähes 80 prosenttia vastaajista koki tekoälyn tuovan mukanaan uusia ja ainutlaatuisia haasteita kyberturvallisuudelle (5). Digitaalisessa turvallisuudessa tekoälyä hyödynnetään ja tutkitaan sekä edistyneempien puolustusmekanismien että hyökkäysmenetelmien kehittämisessä. Liikenne- ja viestintävirasto Traficom on esimerkiksi julkaissut viimeisten kahden vuoden aikana kaksi tekoälyä ja kyberturvallisuutta koskevaa raporttia: Tekoälyn soveltamisen kyberturvallisuus ja riskienhallinta (6) sekä Tekoälyn mahdollistamat kyberhyökkäykset (7). Näissä raporteissa keskitytään tekoälyn sovellutuksiin yleisesti liittyviin riskeihin sekä tekoälyn hyödyntämiseen hyökkäyksellisissä käyttötarkoituksissa.

Tässä selvityksessä täydennetään näiden raporttien tuottamaa tietoa puolustuksellisesta näkökulmasta. Selvityksessä tarkastellaan tekoälyn yhtä merkittävää osa-aluetta, koneoppimista, digitaalisen turvallisuuden teknisissä valvontakäyttötapauksissa. Selvityksen tarkoituksena on tuottaa yleiskuvaa sekä lisätä ymmärrystä ja tietoisuutta koneoppimisen käytöstä digitaalisen turvallisuuden valvonnassa, hyödyistä joita sillä on mahdollista saavuttaa sekä riskeistä joita sen käytössä pitää huomioida. Lisäksi tämän selvityksen lähdeluetteloon on pyritty kokoamaan sellaisia artikkeleita, raportteja, säädöksiä ja muita julkaisuja, joihin tutustumalla omaa ymmärrystä aihepiiristä on hyvä lähteä syventämään ja kasvattamaan. Selvitys on suunnattu kaikille julkisessa hallinnossa digitaalisen turvallisuuden parissa työskenteleville henkilöille.

Selvitystä varten valtionvarainministeriö haastatteli vuodenvaihteessa 2022–2023 digitaalisen turvallisuuden ja koneoppimisen asiantuntijoita tutkimussektorilta, julkisesta hallinnosta ja elinkeinoelämästä. Elinkeinoelämästä oli mukana haastateltavia sekä koneoppivien valvontaratkaisuiden ympärille palveluita toimittavista organisaatiosta että näitä ratkaisuja kehittävistä yrityksistä. Haastatteluihin osallistui asiantuntijoita seuraavista organisaatioista: CGI, Cisco, Suomen Erillisverkot, Helsingin yliopisto, Insta Group, Istekki, Jyväskylän yliopisto, KSTieto, Maanpuolustuskorkeakoulu, Microsoft, Netox, Nixu, Oulun yliopisto, Puolustusvoimat, Traficom, valtioneuvoston kanslia, Valtori, Verohallinto ja WithSecure. Raportissa esitetyt havainnot perustuvat pääasiassa näissä haastatteluissa esitettyihin näkemyksiin ja huomioihin. Haastatteluista kerättyä tietoa on lisäksi täydennetty tutkimusartikkeleista sekä muiden toimijoiden, kuten tietoturvayhtiöiden, julkaisemista raporteista tai muista julkaisuista kerätyillä tiedoilla ja näkemyksillä.

Raportti koostuu johdanto-osio mukaan lukien kahdeksasta pääosiosta. Johdannon jälkeen osiossa 2 määritellään raportin kannalta keskeisimmät käsitteet: digitaalinen turvallisuus, koneoppiminen sekä tekninen automaattinen valvonta. Osiossa 3 käydään läpi tekoälyyn ja koneoppimiseen liittyviä juridisia näkökulmia aiheeseen liittyvän sääntelykehikon ja pelisääntöjen ymmärtämiseksi. Osiossa 4 keskitytään hyötyihin, joita koneoppimisesta saadaan digitaalisen turvallisuuden valvonnassa sekä ihmisen rooliin koneoppivan valvontajärjestelmän rinnalla. Hyötyjen jälkeen osiossa 5 tarkastellaan koneoppimisen käyttöön liittyviä riskejä. Osio 6 käsittelee koneoppivien valvontaratkaisuiden hankinnassa huomioon otettavia tekijöitä. Osiossa 7 painopisteenä on koneoppivien valvontaratkaisuiden tulevaisuuden näkymät sekä aihepiiriin liittyvä tutkimus. Osiossa 8 tehdään yhteenveto selvityksen keskeisimmistä havainnoista ja annetaan yleisiä jatkosuosituksia koneoppivien valvontajärjestelmien kehityksen ja hyödyntämisen parantamiseksi selvitysten havaintojen pohjalta.

Tekoälyyn liittyvät eettiset kysymykset ja yhdenmukaisen kohtelun takaaminen ovat laajaa keskustelua herättäneitä aiheita. Tässä selvityksessä sivutaan tekoälyn etiikkaa, mutta syvällisemmän katsauksen saamiseksi tähän tekoälyä läpileikkaavaan osa-alueeseen saa

tutustumalla esimerkiksi Suomen kansallisen tekoälyohjelman AuroraAI:n eettiseen koodistoon (8), OECD:n julkaisuihin (9) sekä Euroopan unionin aihepiiriä koskeviin suosituksiin (10).

2 Käsitteitä

Tämän selvityksen kannalta keskeisimmät käsitteet ovat digitaalinen turvallisuus, koneoppiminen sekä tekninen automaattinen valvonta. Seuraavissa kappaleissa määritellään, mitä näillä käsitteillä tämän selvityksen yhteydessä tarkoitetaan.

2.1 Digitaalinen turvallisuus

Digitaalisen turvallisuuden käsitteelle ei ole olemassa vain yhtä vakiintunutta määritelmää, vaan sitä on määritelty eri foorumeilla eri tavoin. OECD:n joulukuussa 2022 ilmestyneessä OECD Policy Framework on Digital Security -julkaisussa digitaalinen turvallisuus määritellään kyberturvallisuuden taloudelliseksi ja sosiaaliseksi ulottuvuudeksi, joka koostuu taloudellista ja sosiaalista hyvinvointia uhkaavien digitaalisten riskien hallintakeinoista (11). Suomessa digitaalisen turvallisuuden käsite on puolestaan määritelty valtioneuvoston periaatepäätöksessä Julkisen hallinnon digitaalinen turvallisuus (12). Periaatepäätöksessä todetaan, että digitaalisella turvallisuudella tarkoitetaan usein samaa kuin kyberturvallisuudella. Näiden kahden käsitteen erottamiseksi periaatepäätöksessä digitaalinen turvallisuus määritellään viitekehykseksi, johon sisältyy niin riskienhallintaan, toiminnan jatkuvuuden hallintaan ja varautumiseen, kyberturvallisuuteen tietoturvallisuuteen kuin tietosuojaan liittyviä asioita. Periaatepäätöksen mukaan ”digitaalisen turvallisuuden tavoitteena on kokonaisturvallisuuden viitekehyksessä suojata kansalaisia, yhteisöjä ja yhteiskuntaa niiltä riskeiltä ja uhkilta, jotka voivat kohdistua henkilötietoihin ja kansalaisten palveluihin sekä yhteiskunnan ja viranomaisten toimintaan, prosesseihin, palveluihin ja tietoaineistoihin digitaalisessa toimintaympäristössä”. Tässä selvityksessä tukeudutaan tähän digitaalisen turvallisuuden määritelmään ja lisäksi sen synonyymina selvityksessä käytetään käsitettä digiturva.

2.2 Koneoppiminen

Tekoälyllä tarkoitetaan sellaisia koneiden ja järjestelmien kyvykkyyksiä ja ominaisuuksia, jotka on perinteisesti liitetty ihmisen älyyn, kuten päättelyiden tekeminen, suunnitteleminen tai oppiminen (13). Yksi tekoälyn osa-alue on koneoppiminen, jonka perusta on vahvasti tilastotieteen menetelmissä. Jotta järjestelmästä voidaan puhua koneoppivana, on sillä oltava kyky oppia löytämään lainalaisuuksia siitä datasta, jota järjestelmä analysoi.

Järjestelmän tavat oppia voidaan yleisesti jakaa kolmeen pääkategoriaan: ohjattu oppiminen (supervised learning), ohjaamaton oppiminen (unsupervised learning) sekä vahvistettu oppiminen (reinforcement learning). Ohjattu oppiminen soveltuu käyttötapauksiin, joissa tavoiteltu lopputulos on tiedossa. Konetta opetetaan opetusaineistolla, joka koostuu syöte-tavoite-pareista. Opettamisen jälkeen tavoitteena on tila, jossa kone osaa yleistää oppimiaan asioita ja tuottaa haluttuja tavoitteita myös sille entuudestaan tuntemattomilla syötteillä. Ohjaamattomassa oppimisessa toisin kuin ohjatussa oppimisessa, ei määritellä koneelle valmiiksi tavoitetilaa, vaan koneen annetaan omatoimisesti etsiä opetusaineistosta riippuvuuksia, lainalaisuuksia, rakenteita ja samankaltaisuuksia. Vahvistetussa oppimisessa haluttu lopputulos on ohjatun oppimisen tapaan tiedossa, mutta reittiä lopputulokseen ei ole yksinkertaista kuvailla matemaattisessa muodossa. Vahvistetussa oppimisessa konetta opetetaan positiivisen ja negatiivisen palautteen kautta, jolloin kone oppii suosimaan toimintaa, joka maksimoi positiivisen palautteen määrää (14) (15).

Digiturvallisuuden kontekstissa yksi esimerkki ohjatun oppimisen hyödyntämisestä on vaarallisen verkkoliikenteen tunnistaminen muun liikenteen joukosta. Edellytyksenä tälle on se, että joku on entuudestaan riittävässä määrin luokitellut ja opettanut koneelle millainen liikenne on sallittua ja mikä puolestaan ei, jotta koneen on mahdollista yleistää oppimaansa. Tilanteissa, joissa hyökkääjät käyttävät menetelmiä, jotka eivät ole ohjattuun oppimiseen tukeutuvalle järjestelmälle entuudestaan tuttuja, järjestelmän kyvykkyydet ei-halutun tai pahantahtoisen toiminnan tunnistamiseksi ovat siten hyvin rajalliset. Ohjaamattomalla oppimisella puolestaan voidaan tehdä havaintoja asioista, joita ei tunneta entuudestaan. Toisin kuin ohjatussa oppimisessa, ohjaamattomassa oppimisessa ei ole tarkalleen tiedossa sitä mitä ympäristöstä halutaan löytää. Tällöin esimerkiksi valvontajärjestelmä hälyttää sellaisista toiminta- ja käyttäytymismalleista, jotka ovat ristiriidassa valvottavassa ympäristössä tapahtuvan normaalin toiminnan kanssa. Ohjaamaton oppiminen on täten edellytys tunnistamaan sellaisia hyökkäyksiä, joita järjestelmä ei ole ennen havainnut, esimerkiksi nollapäivähyökkäyksiä. Vahvistetun oppimisen sovellutusalueita digiturvallisuudessa ovat puolestaan esimerkiksi tehokkaampien suojaus- ja hallintakeinojen kehittäminen hajautettuja palvelunestohyökkäyksiä vastaan tai penetraatiotestauksen automatisoiminen (16) (17) (18).

Ohjatun, ohjaamattoman ja vahvistetun oppimisen lisäksi koneoppimisen yhteydessä puhutaan usein syväoppimisesta (deep learning). Syväoppimisessa hyödynnetään monikerroksisia laskentamalleja, joita kutsutaan neuroverkoiksi. Monikerroksisuuden ja suuren parametrimäärän vuoksi syväoppimisen hyödyntäminen edellyttää sekä erittäin suuren määrän dataa että laskentatehoa (14) (19). Menetelmän avulla on kuitenkin mahdollista saada rakennettua perinteisiä koneoppimismenetelmiä tarkempia ennusteita ja parempaa havainnointikyvykkyyttä. Digitaalisen turvallisuuden parissa syväoppimista hyödynnetään muun muassa tunkeutumisen havainnointi- ja estojärjestelmien kehityksessä, haittaohjelmasuojauksessa sekä bottiverkkojen havaitsemisessa (20) (21).

Koneoppimisen luokasta tai alalajista riippumatta koneoppivan järjestelmän oppimisprosessin pääidea on se, että koneoppimismalliin syötetään opetusdataa, jonka perusteella malli oppii tekemään dataa koskevia päätelmiä ja havaintoja. Kun tässä selvityksessä käydään läpi asiantuntijahaastatteluista nousseita löydöksiä, ei tarkemmin eritellä mistä koneoppimisen alalajista missäkin tilanteessa on kyse. Käsitteellisesti puhutaan yleisesti vain koneoppimisesta.

2.3 Tekninen automaattinen valvonta

Teknisellä automaattisella valvonnalla ja digiturvan teknisellä valvonnalla tarkoitetaan prosesseja ja teknologioita, joiden avulla organisaatioiden tietojenkäsittely-ympäristöjä valvotaan uhkien ja haavoittuvuuksien löytämiseksi. Digiturvan teknisellä valvonnalla pyritään havaitsemaan pahantahtoista tai muuten riskejä sisältävää toimintaa, jolla voi olla haitallisia seurauksia valvottavalle tietojenkäsittely-ympäristölle ja siellä oleville toimijoille, tiedoille tai palveluille. Tällainen valvonta voi kohdistua esimerkiksi verkkoliikenteeseen, käyttäjiin ja identiteetteihin, päätelaitteisiin tai tietojärjestelmiin. Markkinoilla olevien valvontateknologioiden yhteydessä käytetään esimerkiksi lyhenteitä SIEM/SOAR (Security Information and Event Management / Security Orchestration, Automation and Response), MDR/XDR (Managed / Extended Detection and Response), ITDR (Identity Threat Detection and Response), CASB (Cloud Access Security Broker) ja IDS/IPS (Intrusion Detection/Prevention System). Valvonnasta organisaatioissa on yleensä vastuussa organisaation oma tai ulkopuoliselta toimittajalta palveluna hankittu tietoturvavalvomotoiminto (SOC). Valvonnan automaattisuudella puolestaan tarkoitetaan niitä tilanteita ja käyttötapauksia, joissa valvonta toteutetaan koneellisesti ja automatisoidusti ihmisen tekemän manuaalisen työn sijaan, esimerkiksi tekoälyä hyödyntämällä.

3 Koneoppimista koskeva lainsäädäntö

Tässä osiossa kuvataan juridisen yleiskuvan ymmärtämiseksi ylätasolla automaattisten ja oppivien järjestelmien toimintaan sovellettavaa normistoa, joka vaikuttaa koneoppivien teknisten valvontajärjestelmien käyttöön. Osion ei ole tarkoitus muodostaa tyhjentävää listausta kaikista automaattisiin ja oppiviin järjestelmiin sovellettavista normeista, vaan keskittyä näiden kannalta keskeisimpiin sisältöihin. Suosittelemme tutustumaan näihin normeihin yksityiskohtaisesti esimerkiksi silloin, kun koneoppivien järjestelmien käyttöönottoa suunnitellaan. Osiossa esitellyt normit muodostuvat pääasiassa kansallisen lainsäädännön sisältämistä säädöksistä, mutta tässä osiossa käsitellään tarvittavilta osin myös EU-tasoisia normeja. Ensiksi kuvataan selvityksen laatimisen aikana voimassa olevat tietosuojalainsäädäntöä koskevat normit, minkä jälkeen kuvataan EU-tasolla valmisteilla olevan tekoälyasetusehdotuksen olennainen sisältö. Tämän osion tiedot kuvaavat oikeustilaa toukokuun alussa 2023. Osion lopuksi käydään läpi asiantuntijahaastatteluissa esiin tulleita tekoälyä ja koneoppimista koskevaan lainsäädäntöön liittyviä huomioita.

3.1 Tietosuojalainsäädäntö

Mikäli koneoppivien teknisten valvontajärjestelmien toimintaan liittyy henkilötietojen käsittelyä, kuten keräämistä, tallentamista, muokkaamista tai tuhoamista, asettaa tietosuojalainsäädäntö käsittelylle reunaehtoja. Olennaisimmat tietosuojalainsäädännön vaatimukset sisältyvät EU:n yleiseen tietosuoja-asetukseen (GDPR) (22). Lisäksi kansallisella tasolla yksityisyyden suojasta työelämässä annettu työelämän tietosuojalaki (759/2004) (23) muodostaa normiperustan työ- tai virkasuhteessa tapahtuvalle henkilötietojen käsittelylle, sillä lain tarkoituksena on toteuttaa yksityiselämän suojaa ja muita yksityisyyden suojaa turvaavia perusoikeuksia työelämässä.

Tietosuoja-asetus on merkittävin henkilötietojen käsittelyä koskeva säädös ja se koskee myös koneoppivia järjestelmiä. Tietosuoja-asetuksen velvoitteet kohdistuvat henkilötietojen käsittelyyn osallistuviin osapuoliin, joita ovat rekisterinpitäjä tai yhteisrekisterinpitäjä sekä käsittelijä. Rekisterinpitäjällä tarkoitetaan tahoa, joka yksin tai yhdessä toisten kanssa määrittelee henkilötietojen käsittelyn tarkoitukset ja keinot. Yhteisrekisterinpitäjän tilanteessa on rekisterinpitäjiä vähintään kaksi, jotka yhdessä määrittävät käsittelyn tarkoitukset ja keinot. Käsittelijä puolestaan on taho, joka käsittelee henkilötietoja rekisterinpitäjän lukuun. Henkilötietoja koneoppivissa valvontajärjestelmissä käsittelevien

osapuolten tulee määrittää näistä rooleista omaan tilanteeseensa soveltuvat roolit. Automaattisten ja oppivien järjestelmien yhteydessä on käsillä tyypillisesti tilanne, jossa järjestelmän kehittäjä tarjoaa tuki- ja ylläpitopalveluita sekä suorittaa muita käsittelytoimia tilaajana toimivalle organisaatiolle.

Tietosuoja-asetuksen olennaisen osan muodostavat henkilötietojen käsittelyä koskevat artiklassa 5 kuvatut periaatteet, joita tulee noudattaa henkilötietoja käsiteltäessä. Nämä periaatteet koskevat erityisesti käsittelyn lainmukaisuutta, henkilötietojen keräämisen käyttötarkoitussidonnaisuutta, tietojen asianmukaisuutta ja olennaisuutta, tietojen täsmällisyyttä ja päivittämistä, tietojen säilyttämistä vain tarpeen mukaisen ajan sekä tietojen turvallisuuden ja eheyden varmistamista. Näiden tietosuojaperiaatteiden noudattaminen asettaa reunaehtoja automaattisten ja oppivien järjestelmien toiminnassa tapahtuvalle henkilötietojen käsittelylle. Jos esimerkiksi koneoppivaa järjestelmää opetetaan henkilötiedoilla, tulee rekisterinpitäjän varmistua siitä, että henkilötietojen käsittely noudattaa lainmukaisuuden periaatetta, mikä tarkoittaa sitä, että käsittelytoimille on määriteltävä asianmukainen oikeusperuste tietosuoja-asetuksesta.

Tietosuoja-asetus asettaa myös rajoituksia henkilötietojen siirrolle Euroopan talousalueen ulkopuolelle. Tähän liittyviä seikkoja koneoppivien teknisten valvontaratkaisuiden kannalta sivutaan tarkemmin tämän selvityksen osiossa 5. Tietosuoja-asetuksessa kielletään lisäksi automaattisen päätöksenteon hyödyntäminen, mutta asetuksen artiklassa 22 asetetaan reunaehdot ja sallitut poikkeukset, joiden nojalla automaattinen päätöksenteko on sallittua. Asetuksessa automaattiseksi päätöksenteoksi määritellään päätöksenteko, jossa on kyse pelkästään automaattiseen henkilötietojen käsittelyyn perustuvasta päätöksenteosta ja jonka tuloksena tehtävillä päätöksillä on oikeusvaikutuksia tai muuten merkittäviä vaikutuksia rekisteröityyn. Jos koneoppivaa järjestelmää hyödynnetään henkilötietojen käsittelyä koskevaan automaattiseen päätöksentekoon, tulee rekisterinpitäjän määrittää käsittelyn perusteeksi jokin asetuksessa kuvatuissa poikkeusperusteista.

Suomessa kansallisella tasolla automaattista päätöksentekoa ja ratkaisumenettelyä koskien säädettiin keväällä 2023 voimaan tulleissa hallintolain ja tiedonhallintalain uudistuksissa rajoista ja vaatimuksista koskien automaattista päätöksentekoa hallintoasioissa (24) (25). Näiden seurauksena esimerkiksi tekoälypohjaista automaattista päätöksentekoa ei ole sallittua käyttää hallintoasioiden ratkaisemisessa. Näistä linjauksista on yleisesti tekoälyn ja automatisoidun päätöksenteon ympärillä käytävän keskustelun kannalta hyvä olla tietoinen, mutta koneoppivien digiturvan valvontaratkaisuiden käyttötapauksiin näillä ei ole suoraa vaikutusta, koska näiden järjestelmien mahdolliset päätökset esimerkiksi käyttäjätunnuksen tai palomuurin portin hetkellisestä sulkemisesta eivät ole luokiteltavissa hallintoasian ratkaisemiseksi.

Tietosuoja-asetuksen lisäksi suomalainen työelämän tietosuojalaki rajoittaa työ- ja virkasuhdetta koskevien henkilötietojen käsittelyä, kuten teknistä valvontaa työpaikalla sekä työntekijän sähköpostiviestin hakemista ja avaamista. Huomionarvoisena kohtana on työelämän tietosuojalain 21 §:n perusteella säädetty yhteistoimintamenettelyn sekä yhteistoimintalaissa tarkoitetun vuoropuhelun piiriin kameravalvontaa, kulunvalvontaa ja muuta teknistä valvontaa koskevat toiminnot sekä sähköpostin ja muun tietoverkon käyttö ja näitä koskevien tietojen käsittely. Tämän perusteella myös koneoppivien digiturvan valvontaratkaisuiden kohdalla tulee työnantajan varmistua lainsäädännön reunaehtojen toteutumisesta ja erityisesti määritellä valvonnan käyttötarkoitus ja siinä käytettävät tekniset menetelmät sekä viestiä näistä prosesseista työntekijöilleen. Tilanteissa, joissa koneoppivalla valvontaratkaisulla on mahdollisesti näkyvyyttä myös henkilöstön sähköpostiliikenteeseen, tulee huomioitavaksi työelämän tietosuojalain lisäksi myös laki sähköisen viestinnän palveluista (917/2014) (26). Edellä kuvattujen säädösten ohella tekoälyn ja koneoppimisen käytössä ja näiden käsittelemän datan kohdalla tulee huomioida myös sektorikohtaisen sääntelyn asettamat vaatimukset, kuten esimerkiksi potilastietojen hyödyntämiselle asetetut rajoitteet sosiaali- ja terveydenhuollossa.

3.2 Tekoälyasetusehdotus

Euroopan komissio on julkaissut ehdotuksen tekoälyasetuksesta, jonka tarkoituksena on varmistaa ihmisten ja yritysten turvallisuus ja perusoikeuksien toteutuminen tekoälyn hyödyntämisen yhteydessä (27). Sääntelyn on tarkoitus tulla voimaan ja sovellettavaksi lähivuosina. Asetusehdotuksessa tekoälyjärjestelmät luokitellaan riskikategorioihin sen perusteella, aiheutuuko niistä ei-hyväksyttävä riski, suuri riski, vähäinen riski vai minimaalinen riski. Tekoälyjärjestelmään sovellettavat reunaehdot määräytyisivät osaltaan riskikategorian perusteella.

Tekoälyjärjestelmät, joista aiheutuvaa riskiä ei pidetä hyväksyttävänä, kiellettäisiin kokonaan. Näitä olisivat muun muassa tekoälyjärjestelmät tai -sovellukset, joilla manipuloidaan ihmisen käyttäytymistä vapaan tahdon vastaisesti sekä sosiaalisten tekijöiden perusteella ihmisiä viranomaisten tarkoituksiin pisteyttävät järjestelmät. Suuririskisten tekoälyjärjestelmien tulisi täyttää tiettyjä seuraavassa kappaleessa kuvattuja vaatimuksia ennen kuin niitä olisi sallittua saattaa markkinoille EU:ssa.

Suuririskiksi katsottaisiin muun muassa tekoälyjärjestelmät, joita hyödynnetään i) kriittisissä infrastruktuureissa, joissa ne voisivat vaarantaa ihmisten hengen ja terveyden, ii) lainvalvonnassa, erityisesti, jos tekoäly saattaisi loukata ihmisten perusoikeuksia, iii) työllistämisen ja henkilöstöhallinnon yhteydessä ja iv) keskeisessä asemassa olevien yksityisten ja julkisten palveluiden yhteydessä. Suuririskisten tekoälyjärjestelmien markkinoille asettamisen edellytykseksi asetettaisiin vaatimuksia koskien muun muassa datan hallintaa,

dokumentointia, tietojen säilyttämistä, läpinäkyvyyttä, ihmisen suorittamaa valvontaa, luotettavuutta, tarkkuutta ja turvallisuutta. Lisäksi järjestelmille tulisi toteuttaa vaatimustenmukaisuuden arviointi. Asetusehdotuksessa asetetaan näiden ohella vaatimuksia muun muassa asianmukaisten seurantatoimenpiteiden käyttöönottoon ja ongelmatilanteista ilmoittamiseen.

Eräille tekoälyjärjestelmille, joiden käytöstä katsotaan aiheutuvan vähäinen riski, asetettaisiin läpinäkyvyyttä koskevia velvoitteita, jotta niiden käyttäjät ymmärtäisivät olevansa vuorovaikutuksessa koneen kanssa. Asetusehdotuksella ei kuitenkaan puututtaisi sellaisten tekoälyjärjestelmien käyttöön, joista aiheutuvaa riskiä henkilöiden oikeuksille tai turvallisuudelle pidetään minimaalisena tai olemattomana. Selkeästi suurin osa Euroopan unionin alueella käytettävistä tekoälyjärjestelmistä ovat tällä hetkellä minimaalisen tai olemattoman riskin kategoriassa (28). Koneoppivat digiturvan valvontaratkaisut asettuvat nykyisillä kyvykkyyksillään ja käyttötarkoituksillaan kahteen jälkimmäiseen kategoriaan. Tilanne voisi kuitenkin muuttua, jos näihin järjestelmiin alkaisi tulla mukaan ominaisuuksia, joilla työntekijöitä luokiteltaisiin esimerkiksi ”hyviin” ja ”huonoihin”, ja jotka toimisivat perusteina järjestelmien tekemille päätöksille.

3.3 Tavoitteena tekoälyn hyödyntämistä palveleva lainsäädäntö

Tätä selvitystä varten tehdyissä haastatteluissa moni asiantuntija painotti sitä, että koneoppimista koskevassa säädösvalmistelussa tulisi juridiikkaa tuntevien lisäksi olla riittävässä määrin mukana koneoppimista tuntevia asiantuntijoita. Tämän koettiin edesauttavan sitä, että lainsäädäntö palvelisi ja tukisi tekoälyn ja koneoppimisen hyödyntämistä mahdollisimman tehokkaasti ja riskilähtöisesti. Yhtenä huolena esiin nousi se, ettei Euroopan unioni nyt valmisteilla olevalla tekoälyasetuksella rajoittaisi ja hankaloittaisi tekoälyn käyttöä ja sitä koskevaa kehitystä liikaa, jotta välimatka muualla maailmassa tapahtuvaan teknologiseen kehitykseen ja tutkimukseen ei kasvaisi liian suureksi. Lisäksi toivottiin, että EU:n tekoälyasetus ja muu aihepiiriä koskeva sääntely ei muodostuisi yhtä tulkinnanvaraiseksi kuin on tilanne esimerkiksi GDPR:n osalta. Edellä kuvattujen GDPR:n rajoitteiden myötä EU-tason tietosuojasääntelyn voidaan todeta asettavan merkittäviä reunaehtoja henkilötietojen hyödyntämiselle automaattisten ja oppivien järjestelmien käytössä, eikä tietosuojasääntely kaikissa tilanteissa vastaa automaattisten ja oppivien järjestelmien hyödyntämisen realiteetteja. Hankalaksi koettiin myös se, miten lainsäädännön ja ohjauksen on mahdollista pysyä näin nopeasti kehittyvän alan mukana. Esimerkiksi koneoppivia järjestelmiä koskevien sertifiointiskeemojen käyttöönottamisessa EU-tasolla uskotaan kuluvan vielä useita vuosia, vaikka voimakasta tarvetta näille olisi jo nyt.

Toisaalta kansallisen lainsäädännön, kuten esimerkiksi työelämän tietosuojalain, nähtiin olevan hyvällä tasolla, koska sen katsottiin mahdollistavan teknisen valvonnan toteuttamisen, kunhan tarvittavista yhteistoimintamenettelyistä on huolehdittu. Haastatteluissa nousi esiin myös näkemys, jonka mukaan teknisiä havainnointi- ja valvontakyvykkyyksiä velvoittava lainsäädäntö olisi tervetullutta ja tärkeässä asemassa yleisen havainnointikyvyn parantamisen ja laajenemisen kannalta. Saman suuntaisia kehotuksia annetaan Euroopan unionin uuden NIS2-kyberturvallisuusdirektiivin (2022/2555) johdanto-osan resitaalissa 51 (29). Siinä todetaan, että ”jäsenvaltioiden olisi kannustettava käyttämään mitä tahansa innovatiivista teknologiaa, myös tekoälyä, jonka käyttö voisi parantaa kyberhyökkäysten havaitsemista ja ehkäisemistä ja mahdollistaa resurssien tehokkaamman kohdentamisen kyberhyökkäyksiin”. Kannustimet tekoälyn ja koneoppimisen hyödyntämiselle digiturvaa edesauttavissa käyttötapauksissa ovat näin olemassa, mutta työtä lainsäädännön yhdenmukaistamisessa ja selkeyttämisessä on vielä jäljellä, jotta harmaalla alueella toimiminen voitaisiin minimoida ja yhteiset pelisäännöt olisivat kaikille selvät.

4 Koneoppimisen käytöstä digiturvan teknisessä valvonnassa saatavat hyödyt

Tässä osiossa käydään läpi koneoppimisesta digiturvan valvonnassa saatavia hyötyjä havannointikyvyn kannalta sekä pohditaan ihmisen roolia valvontatoiminnoissa näiden järjestelmien rinnalla.

4.1 Havainnointikyvyn tehostuminen

Asiantuntijahaastatteluiden perusteella koneoppimisesta saatava keskeisin hyöty digiturvan teknisessä valvonnassa on sen avulla saavutettava tehokkaampi ja kokonaisvaltaisempi havainnointikyky. Dataa kerätään eri lähteistä, kuten päätelaitteilta, palvelimilta, verkkolaitteilta ja identiteetin- ja pääsynhallintaratkaisuista niin paljon, että sen analysoiminen ilman koneoppimista olisi hyvin vaikeaa, ellei jopa mahdotonta. Koneoppiminen mahdollistaa tämän datan reaaliaikaisen läpikäymisen sekä siten ihmisasiantuntijoiden ajankäytön tehokkaamman kohdentamisen. Lisäksi koneoppimiseen tukeutuvat valvontaratkaisut mahdollistavat hälytysten tuottamisen aiempaa hiljaisemmista vihjeistä. Reaaliaikaisen valvontakyvykkyyden tehostumisen ohella koneoppimisen avulla on mahdollista löytää loki- ja historiatiedoista sellaista, joka aiemmin on jäänyt piiloon.

Tulokset koneoppimisen hyödyntämisestä myös niin sanottujen nollapäivähyökkäysten (zero-day attack) havaitsemisessa ovat olleet lupaavia. Nollapäivähyökkäyksillä tarkoitetaan sellaisten haavoittuvuuksien hyödyntämistä, joiden paikkaamiseksi kyseisestä ohjelmakoodista vastuussa olevat kehittäjät eivät ole vielä ehtineet julkaista korjaavaa päivitystä. Kyvykkyys nollapäivähyökkäyksien havaitsemiseen perustuu siihen, että erilaisilla hyökkäystyypeillä opetettu malli oppii tunnistamaan hyökkäykselle tyypillisiä vaiheita, piirteitä ja indikaattoreita, vaikka nämä eivät toistuisikaan täysin identtisesti. Täten koneoppiva valvontajärjestelmä voi esimerkiksi merkitä uuden käyttäytymistään muuttaneen haittaohjelman vaaralliseksi, mihin perinteisellä tiiviste- ja tunnistevertailulla ei pystytä (30).

4.2 Sääntö- ja tunnistepohjaisuudesta käyttäytymisen analysointiin

Yksi digiturvaan ja koneoppimiseen liittyvä käsite on UEBA (User & Entity Behavioral Analytics), jolla viitataan käyttäjien ja muiden entiteettien, kuten sovellusten ja päätelaitteiden, käyttäytymisen analysointiin. Digiturvan valvontaratkaisuiden UEBA-kyvykkyyksien tarkoituksena on löytää seurattavien identiteettien ja kohteiden käytöksestä poikkeamia, jotka voisivat viitata haitalliseen tai pahantahtoiseen toimintaan organisaation tietojenkäsittely-ympäristössä. Jotta poikkeava ja epätyypillinen toiminta voidaan tunnistaa normaalista toiminnasta, täytyy UEBA-ratkaisun ensin ymmärtää kyseisen valvottavan ympäristön normaali lähtötaso (baseline). Tähän tarvitaan koneoppivaa valvontajärjestelmää, joka seuraa ympäristöä ja oppii siten, mikä ympäristössä on sinne kuuluvaa tavallista toimintaa ja mikä puolestaan on epätyypillistä käytöstä, joka edellyttää reagointia (31).

Perinteisillä sääntö- ja tunnistepohjaisilla digiturvaratkaisuilla kuten haittaohjelmasuojauksella tai tunkeutumisenestojärjestelmällä voidaan löytää ainoastaan entuudestaan tunnettuja riskitekijöitä. Näitä voivat olla esimerkiksi mahdollisia riskejä sisältäviksi merkityt IP-osoitteet, tietoliikennepakettien otsikkotiedot tai tiedostojen tiivisteet. Sääntö- ja tunnistepohjaiset ratkaisut ovat siten reaktiivisia proaktiivisten sijaan: niillä reagoidaan sellaisiin tapahtumiin ja tiedostoihin, joiden tiedetään aiemman kokemuksen ja kerätyn tiedon pohjalta olevan vaarallisia tai pahantahtoisia (20). Koneoppimiseen tukeutuvilla UEBA-ratkaisuilla voidaan puolestaan löytää myös sellaisia valvontaympäristössä olevia haitallisia toimia, jotka eivät ole tuttuja entuudestaan. Tyypillinen esimerkki UEBA-kyvykkyyksistä on mahdollisesti vääriin käsiin joutuneen identiteetin tunnistaminen, kun havaitaan, että identiteetin suorittamat toimet eivät vastaa sitä, miten kyseinen identiteetti on tyypillisesti toiminut. Tällaista toimintaa voivat olla vaikkapa poikkeavan suuret tiedostojen latausmäärät tai kirjautumiset sellaisesta maantieteellisestä paikasta, josta käyttäjä normaalisti ei ole kirjautunut sisään palveluihin.

Käyttäytymisen seurantaan pohjautuvan havainnointikyvykkyyden lisäksi toinen haastatteluissa esiin tullut keskeinen piirre, ja tavallaan tehostuneen havainnointikyvyn hinta verrattuna perinteisiin sääntö- ja tunnistepohjaisiin järjestelmiin, on koneoppivien järjestelmien tuottamien niin sanottujen väärien positiivisten hälytysten suhteellisen suuri määrä. Väärällä positiivisella tarkoitetaan hälytystä, jossa käyttäjien normaali toiminta on tulkittu virheellisesti pahantahtoiseksi toiminnaksi. Sääntö- ja tunnistepohjaisten valvontajärjestelmien tuottamat hälytykset sisältävät korkean tason indikaation meneillään olevasta pahantahtoisesta tai haitallisesta toiminnasta, koska ne perustuvat jo entuudestaan tunnettuja riskejä sisältävien tapahtumien havainnoimiseen. Koneoppivien järjestelmien

kyvystä havaita entuudestaan tuntemattomia riskitekijöitä seuraa, että hälytyksiä saattaa aiheutua myös tapahtumista, joissa mistään vaarallisesta tai kielletystä toiminnasta ei ole kyse.

Koneoppimista hyödyntävien valvontaratkaisuiden perinteisiin sääntö- ja tunnistepohjaisiin ratkaisuihin verrattuna tehokkaammasta havainnointikyvykkyydestä huolimatta haastatteluissa tuli esille, että perinteisille valvontaratkaisuille on edelleen tarvetta, eikä niitä kannata kaikissa käyttötapauksissa lähteä väkisin tai itsearvoisesti korvamaan koneoppivilla järjestelmissä. Entuudestaan tunnettujen hyökkäystyyppien ja haitallisten sovellusten havainnoimisessa sääntö- ja tunnistepohjaiset ratkaisut ovat edelleen tehokkaita ja organisaatioiden kyberpuolustusta vahvistavia työkaluja. Paras lopputulos havainnoinnin kannalta saavutetaan, kun vanhoja ja olemassa olevia kyvykkyyksiä täydennetään uusilla koneoppimiskyvykkyyksillä.

4.3 Ihmisen rooli

Asiantuntijahaastatteluissa korostettiin, että koneoppivat valvontajärjestelmät eivät tee ihmistä tarpeettomaksi, mutta ne muuttavat ihmistyön luonnetta. Toisaalta haastatteluissa todettiin, että koneoppivan valvontajärjestelmän käyttöönotto voi olla järjestelmää operoivalle tietoturvatyöntekijälle niin näkymätön muutos, että työntekijä ei välttämättä aina edes tiedosta olevansa tekemisissä koneoppivan ratkaisun kanssa. Eräs haastateltu asiantuntija kuvasikin koneoppivian valvontajärjestelmän käyttöönottoa tietyissä tilanteissa ”näkymättömäksi prosessiksi, joka lopputuloksenaan lisää näkyvyyttä (havainnointikykyä)”.

Edellä mainittujen väärien positiivisten hälytysten kohdalla ihmisellä on keskeinen rooli koneoppivan valvontajärjestelmän opettamisessa tuottamaan entistä tarkempia ja parempia havaintoja. Käytännössä tämä tapahtuu esimerkiksi niin, että ihminen merkitsee valvontajärjestelmän käyttöliittymän kautta väärät positiiviset hälytykset vääriksi ja siten viestii koneoppimismallille tarpeesta muokata perusteita seuraavien hälytysten tuottamiselle. Koneoppimismallin jatkuvan ja dynaamisen opettamisen vaihtoehtona on kertaalleen opetettu staattinen malli, mutta uhkien ja hyökkäystapojen jatkuvasti muuttuessa ja kehittyessä on dynaaminen lähestymistapa digiturvan valvonnan käyttötapauksissa suositeltavampi.

Koska koneoppivat järjestelmät ovat ihmistä tehokkaampia käymään läpi suuria datamassoja, on näiden järjestelmien ensisijainen hyöty ensimmäisen tason analyysivaiheen suorittaminen ja relevantin tiedon suodattaminen epäoleellisesta ihmisille jatkoanalyysiä ja toimenpiteitä varten. Haastateltujen asiantuntijoiden enemmistön mukaan koneoppivien valvontajärjestelmien nykyisellä kypsyystasolla ensimmäisen tason (Tier 1) päätöksentekoa on mahdollista ulkoistaa koneoppivalle algoritmille, mutta tästä

eteenpäin säilyy ihmisen rooli päätöksentekoketjussa ennallaan. Haastatteluissa tuotiin esiin, että koneoppivien järjestelmien kyvyt tehdä riski- tai vaikutustenarviointia ovat edelleen suppeat, minkä takia ihmisen panosta tarvitaan esimerkiksi laajamittaista palveluiden pysäyttämistä koskevien päätösten tekemisessä. Monimutkaisiin häiriötilanteesiin liittyy paljon epävarmuutta, jonka ratkaisemiseksi ja oikeiden jatkotoimenpiteiden päättämiseksi ihmiskokemus on edelleen välttämätöntä.

Haastatellut asiantuntijat painottivat myös sitä, että koneoppiviin valvontaratkaisuihin ei tule suhtautua kuin taianomaisiin hopealuoteihin, jotka itsestään ratkaisisivat kaikki digiturvaa koskevat haasteet. Kuten muidenkin IT-projektien kohdalla, on myös koneoppivia ratkaisuja käyttöönotettaessa tärkeä määritellä tavoitteet ja käyttötapaukset, joihin näillä ratkaisuilla halutaan löytää apuja. On myös tärkeää huolehtia näiden ratkaisujen ylläpitoon ja niiden tuottaman tiedon jatkokäsittelyyn tarvittavasta resursoinnista. Vaikka koneoppiva valvontajärjestelmä kykenisi nostamaan esiin kuinka tarkkoja tai aiemmin piiloon jääneitä havaintoja tahansa, niin tämän hyöty jää hyvin vähäiseksi, jos organisaatiolla ei ole osaava henkilöstöä, joka käynnistäisi jatkotoimenpiteitä hälytysten perusteella.

5 Koneoppimisen hyödyntämisen riskit

Koneoppimiseen, kuten kaikkeen teknologiaan, liittyy myös rajoitteita ja riskejä. Haastatteluissa korostui se, että koneoppivia valvontajärjestelmiä käytettäessä ja käyttöönotettaessa on hyvä ymmärtää ja hyväksyä se, että koneoppivilla teknologioilla on rajansa ja ne voivat toimia tietyissä tilanteissa väärin tai virheellisesti. Tämän takia riskilähtöinen lähestymistapa on tärkeää myös koneoppimisen kohdalla. Kun riskejä on pohdittu, tunnistettu ja arvioitu jo etukäteen, on mahdollisten tulevien yllätysten määrä vähäisempi ja niihin on tällöin osattu varautua paremmin.

Koneoppimiseen liittyviä riskejä ja niiden ehkäisymenetelmiä on yksityiskohtaisesti kuvattu vuonna 2021 julkaistussa Traficomin julkaisussa Tekoälyn soveltamisen kyberturvallisuus ja riskienhallinta (6). Tämän osion tarkoituksena ei ole toisintaa kyseisen julkaisun havaintoja, vaan nostaa esiin tätä selvitystä varten haastateltujen asiantuntijoiden näkemyksiä koneoppiviin valvontateknologioihin liittyvistä keskeisistä riskeistä. Nämä riskit ovat osin päällekkäisiä Traficomin julkaisussa esiin tuotujen kanssa, mutta osittain ne myös täydentävät aiempia löydöksiä.

5.1 Hyökkäykset koneoppimismallia vastaan

Koneoppimisesta saatavat hyödyt ja koneoppimismallin oikeanlainen toiminta perustuvat pitkälti siihen, kuinka hyvin malli on onnistuttu opettamaan. Oppimisen ja mallin opetuksessa käytetyn koulutusdatan laadulla on ratkaiseva merkitys. Jos koulutusdata on määrältään hyvin suppeaa, kasvavat väärien havaintojen, vinoumien ja painotusvirheiden riskit. Koneoppivien valvontajärjestelmien opetusdatan laatuun liittyvä esimerkkiriskiskenaario on tilanne, jossa valvontajärjestelmää opetetaan ympäristössä, jossa hyökkääjä on jo sisällä. Tällöin on mahdollista, että valvontajärjestelmä tulkitsee hyökkääjän läsnäolon normaaliksi kyseiseen ympäristöön kuuluvaksi toiminnaksi, eikä osaa tehdä tilanteesta asianmukaisia hälytyksiä.

Toisaalta koneoppimismallia vastaan voidaan hyökätä myös tarkoituksella. Jos hyökkääjä tuntee, kuinka puolustus- ja valvontateknologian koneoppimismalli toimii, voi hyökkääjä tietoisesti pyrkiä naamioimaan toimintansa sellaiseksi, jonka malli kokee normaaliksi ja siten sulautua joukkoon (4). Koneoppimismallia voidaan myös esimerkiksi yrittää myrkyttää toimimaan väärin. Myrkyttäminen on koneoppimismallin opetusvaiheeseen liittyvä

alakategoria koneoppimismallia itseään vastaan kohdistuville hyökkäyksille, jolla pyritään vaikuttamaan mallin toimintaan hyökkääjälle suotuisalla tavalla (32). Myrkytyshyökkäyksessä malliin syötetään dataa, jonka tavoitteena on saada malli oppimaan jotakin, mitä sen ei pitäisi oppia. Valvontajärjestelmien tapauksessa tällä voitaisiin pyrkiä esimerkiksi siihen, että valvontajärjestelmä näyttäisi toimivan oikein, mutta siinä olisi takaovi, joka mahdollistaisi hyökkääjän toiminnan ilman, että järjestelmä hälyttäisi siitä. Koneoppismallin opetusdatan laatuun ja tahallisiin myrkyttämisyrityksiin liittyvien riskien lisäksi riskialtis tilanne voi olla myös sellainen, jossa jokin valvontaympäristön lähdejärjestelmä alkaa tuottaa hyvin poikkeuksellista dataa, jollaista valvontajärjestelmän koneoppimisalli ei ole kohdannut aiemmin. Tämä saattaa saada valvontajärjestelmän toimimaan arvaamattomalla ja vaikeasti ennakoitavalla tavalla.

5.2 Läpinäkyvyys, valvontajärjestelmien valvonta ja luottamus

Koneoppimismalleihin kohdistuvien mahdollisten hyökkäysten ja muiden riskien vuoksi olisi tärkeää, että mallin oikeanlaista toimintaa kyettäisiin valvomaan. Tämä voi olla vaikeaa, koska koneoppimista syvällisesti ymmärtäviä asiantuntijoita ei ole läheskään kaikkien organisaatioiden palveluksessa. Lisäksi monet markkinoilla olevat koneoppivat valvontaratkaisut ovat niin sanottuja black box -ratkaisuja, eli järjestelmän käyttäjällä ei ole näkyvyyttä järjestelmien sisäiseen toimintalogiikkaan. Asiantuntijahaastatteluiden vastauksissa muistutettiin myös siitä, että koneoppimisalgoritmien toimintalogiikka ei ole ihmiselle helppoa, vaikka mallit olisivatkin julkisesti kaikkien saatavilla. Yhdessä haastattelussa käytettiin vertauksena sitä, miten ihminen kykenee ymmärtämään ja hahmottamaan maailmaa neliulotteisesti (aika mukaan lukien), mutta algoritmin käsittelemä ulottuvuuksien määrä voi olla huomattavasti suurempi. Erot ihmisen ja algoritmin välillä tiedon jäsentämisessä voivat olla siten hyvin perustavanlaatuisia, mikä voi entisestään lisätä epäilyksiä koneoppivia järjestelmiä kohtaan. Haastatteluissa painotettiin, että koneoppivia järjestelmiä koskevissa riskianalyyseissä tulisi huomioida myös se, mitä kaikkea valvonta-algoritmien toiminnasta ei tiedetä.

Asiantuntijahaastatteluissa nousi esille kysymys siitä, kuinka paljon päätöksentekoa voidaan ulkoistaa ”mustille laatikoille”, joiden päätöksenteon perusteita emme tunne tai joita emme ymmärrä. Kysymyksellä on vahva yhteys selvityksen johdannossa esitettyyn Future of Life Institute -järjestön vetoomukseen tekoälyn kehittämisen laittamisesta tauolle, jotta emme päädy tilanteeseen, jossa toimintaamme ohjaa teknologia, jonka päätösten perusteita emme voi ymmärtää. Toisaalta mustiin laatikkoihin perustuvasta lähestymistavasta heräsi osassa haastatteluista epäilys siitä, kuinka paljon koneoppimista

hyödyntäviä ominaisuuksia markkinoilta saatavissa teknisissä valvontaratkaisuissa lopulta on ja kuinka paljon ne koneoppimisen sijaan tukeutuvat esimerkiksi hyvin hienojakoiseen sääntöpohjaisuuteen.

Näille kysymyksille ja epäilyksille saatiin myös vasta-argumentteja. Haastatteluissa todettiin, etteivät ihmisetkään yleensä avaa kaikkia toimintansa ja päätöstensä perusteita, tai ole niistä aina edes tietoisia. Miksi tällöin koneoppivilta järjestelmiltä pitäisi odottaa jotakin enemmän? Hyödynnämme paljon muutakin teknologiaa, jonka toimintaa emme syvällisesti ymmärrä. Harva meistä tietää, miten lentokoneet lopulta toimivat, mutta luotamme siihen, että koneen suunnittelun ja rakentamisen asiantuntijoilla on tarvittavat tiedot koneiden oikean toiminnan varmistamiseksi. Ihmisen ja koneen päätöksentekoprosessien rinnastaminen ei ole kuitenkaan ongelmatonta ja oma kysymyksensä on, onko niitä edes mahdollista tai tavoiteltavaa arvioida keskenään samoista lähtökohdista? Ihmiseltä voidaan tiedustella perusteita hänen tekemilleen ratkaisuille, vaikka ihminen ei välttämättä olisikaan tietoinen kaikista päätöksiin vaikuttaneista tekijöistä. Hyväksymmekö sen, että emme voi ymmärtää koneoppivien ja tekoälyjärjestelmien sisäistä logiikkaa, vai pitäisikö meidän vaatia näiltä läpinäkyvyyttä nykyistä enemmän?

Koska valvontateknologioiden käyttäjillä ei välttämättä ole mahdollisuuksia tai riittäviä resursseja ja osaamista käyttämiensä teknologioiden oikeanlaisen toiminnan valvomiseen, on luottamus avainasemassa. Pelkän uskon varaan teknologioiden valvontaa ei ole suositeltavaa jättää. Koneoppivia valvontajärjestelmiä hankittaessa olisi toimittajalta ja valmistajalta hyvä edellyttää todisteita siitä, millaisia valvontaan ja testaamiseen liittyviä hallintakeinoja teknologian oikeanlaisen toiminnan varmistamiseksi käytetään. Eräässä valvontaan ja sen laiminlyöntiin liittyvässä haastatteluissa esiin tullut näkökulma oli liian hyvin ja laadukkaasti työnsä tekevä koneoppiva valvontajärjestelmä. Tällöin riskinä on, että luottamus järjestelmän toimintaan ja kykyihin voi kasvaa niin suureksi, että toiminnan valvontaa ja ihmisten omaa arviointikykyä ryhdytään laiminlyömään, koska järjestelmän uskotaan toimivan kaikissa tilanteissa oikein.

5.3 Tietosuojariskit

Koneoppimismallien oikeanlaisen toiminnan valvontaan liittyvien kysymysten lisäksi merkittävän riskialueen muodostavat tietosuojaan liittyvät kysymykset. Koneoppivilla digiturvan valvontaratkaisuilla valvotaan usein henkilötietojen sääntöjenmukaista ja turvallista käsittelyä. Jossain vaiheessa tällainen valvontajärjestelmä on yhtä kriittinen suojattavia kohde kuin ne järjestelmät, joita sillä on tarkoitus suojata ja valvoa. Henkilötiedot kasaantuvat myös valvontajärjestelmään.

Osiossa 3 todettiin, että nykyinen tietosuojaa koskeva sääntely ei kaikissa tilanteissa vastaa niihin realiteetteihin, joissa automaattisia ja oppivia järjestelmiä pyritään hyödyntämään. Monet koneoppivia digiturvan valvontajärjestelmiä toimittavista suurimmista yrityksistä tulevat Euroopan ulkopuolelta ja aina ei ole tietoa siitä mihin näiden järjestelmien keräämä lopulta päätyy. Ongelma on siis sama kuin muidenkin pilvipalveluiden kohdalla. Asiantuntijahaastatteluissa huomautettiin, että tämä johtaa tilanteeseen, jossa julkisen hallinnon organisaatioiden on mahdotonta varmasti sanoa tai tietää, mitä analyysiä asiakasorganisaatioiden omistamaan tietoon lopulta kohdistuu ja missä tämä analyysi tehdään. Esimerkiksi yhdysvaltalaiset Patriot Act ja CLOUD Act mahdollistavat tietyissä tilanteissa turvallisuusviranomaisten pääsyn amerikkalaisten teknologiayhtiöiden tallentamaan tietoon. Yksittäisen organisaation neuvottelumahdollisuudet tietosuojaan liittyvissä kysymyksissä ovat suurten globaalien teknologiatoimittajien kanssa vähäiset. Käytännössä näiden yritysten palveluehtoihin on suostuttava, jos niiden tarjoamia teknologioita on tarkoitus käyttää, mikä tarkoittaa samalla mm. tietosuojariskien hyväksymistä.

5.4 Resilienssi ja vastuiden määrittely

Asiantuntijahaastatteluissa painotettiin sitä, että koneoppivien valvontajärjestelmien, kuten minkä tahansa muunkin organisaation käytössä olevan tietojärjestelmän kohdalla, on tärkeää muistaa toipumismenettelyjen ja resilienssin tärkeys, koska epäonnistumisten ja virhetilanteiden mahdollisuus on aina olemassa. Tämä tarkoittaa etukäteen laadittavia suunnitelmia siitä, miten toimia häiriötilanteissa esimerkiksi silloin, kun koneoppiva valvontajärjestelmä ei ole käytettävissä tai miten koneoppiva valvontajärjestelmä voidaan ajaa hallitusti alas. Suunnitelmien tekemisen lisäksi olisi tärkeää myös harjoitella ennakkoon toimintaa häiriötilanteissa.

Häiriötilanteisiin varautumisen lisäksi toinen suositeltava riskienhallintakeino on vastuiden määrittely. Mitkä ovat teknologian kehittäjän, omistajan ja käyttäjän vastuut eri tilanteissa? Mitkä ovat vastuut jos valvontateknologian tekemät väärät havainnot tai automatisoidut toimenpiteet johtavat haitallisiin lopputuloksiin? Vaikka vastuunjaon muuttamisesta suurten globaalien teknologiatoimittajien kanssa ei läheskään aina ole mahdollista neuvotella, on nämä vastuut siitä huolimatta hyvä dokumentoida ja sisäistää.

5.5 Hyödyntämättä jättämisen riski

Tässä osiossa on esitelty mahdollisesti suurelta tuntuva joukko koneoppiviin valvontaratkaisuihin liittyviä riskejä. Nämä riskit on hyvä tiedostaa ja arvioida tapauskohtaisesti, mutta niiden hallintakeinoksi riskien täydellinen välttäminen, eli koneoppivien ratkaisuiden käyttämättä jättäminen, ei ole suositeltava lähestymistapa. Haastatteluissa nimittäin

korostettiin sitä, että koneoppivien valvontaratkaisuiden käyttämättä ja hyödyntämättä jättäminen on itsessään myös huomattava digiturvariski. Uudet koneoppivat tekniset valvontateknologiat ovat keskeinen väline ja työkalu varauduttaessa moderneihin digitaalisen turvallisuuden uhkiin. Asetelmaa tekoälyyn pohjautuvien hyökkäyksellisten ja puolustuksellisten turvallisuusteknologioiden välillä on kutsuttu myös asevarustelukilvaksi, jonka pitkän aikavälin seurauksia on vielä vaikea ennakoida (33). Myös Traficomin Tekoälyn mahdollistamat kyberhyökkäykset -selvityksessä todettiin automaatiota ja tekoälyteknologioita tarvittavan puolustusratkaisuissa, jotta koko ajan kehittyviin kyberhyökkäyksiin olisi mahdollista vastata (7).

Koneoppiviin teknologioihin liittyvät riskit on tiedostettava ja niiden varalle on oltava riskienhallintasuunnitelmat. Koneoppivien valvontateknologioiden käyttämättä jättäminen voi kuitenkin johtaa organisaatioon kohdistuvien uhkien kannalta vielä riskialttiimpaan tilanteeseen, mistä syystä nämä ratkaisut on hyvä pitää organisaation keinovalikoimassa. On myös hyvä huomata, että käytännössä koneoppimiskyvykkyyksien sivuuttaminen ja käyttämättä jättäminen uusien digiturvateknologioiden kohdalla on nykypäivänä lähes mahdotonta, sillä niin laajalle nämä ominaisuudet ovat markkinoilta saatavilla oleviin ratkaisuihin ulottuneet.

6 Koneoppivien valvontajärjestelmien hankinta

Suuri osa kaupallisia digiturvateknologioita kehittävistä ja tuottavista yrityksistä mainostaa tällä hetkellä tuotteitaan koneoppimiseen pohjautuvilla ominaisuuksilla ja kyvykkyyksillä. Tämän seurauksena organisaation lähtiessä hankkimaan uutta digiturvateknologiaa, kuten valvontaratkaisua, on lähes väistämätöntä, että hankinta kohdistuu ainakin osittain koneoppimista hyödyntävään järjestelmään. Tässä osiossa tarkastellaan tekijöitä, joita koneoppivan teknisen valvontajärjestelmän hankinnassa on hyvä ottaa huomioon.

6.1 Riski- ja käyttötapauslähtöisyys

Yleinen syy tietoturvateknologiahankintojen ja -käyttöönottoprojektien epäonnistumiselle on se, että organisaatiossa päädytään hankkimaan jokin ominaisuuksiltaan kattavalta näyttävä ratkaisu ilman tarkempaa selvittämistä siitä, miten se palvelee organisaation tarpeita, käyttötapauksia ja kokonaisarkkitehtuuria riippuvuussuhteineen. Tästä syystä koneoppivan valvontateknologian hankintaprosessia käynnistettäessä on asiantuntijahaastatteluiden mukaan keskeisenä ohjaavana suuntaviivana hyvä pitää riski- ja käyttötapauslähtöisyyttä. Käytännössä tämä tarkoittaa esimerkiksi sitä, että ennen tarjolla olevien teknisten ratkaisujen kartoittamista on analysoitava, mitkä ovat juuri niitä uhkia, joita vastaan organisaation tulisi kyetä suojautumaan ja millaisia kyvykkyyksiä organisaatiolla on jo entuudestaan näiltä suojautumiseksi? Uhkien tunnistamisen sekä riski-, nykytila- ja riippuvuusanalyysien lisäksi on hyvä selvittää ja kartoittaa organisaation muutosvalmiutta. Onko esimerkiksi loppukäyttäjien kannalta käytössä sellaisia hyvin toimivia prosesseja, joita uudella teknologialla ei haluta rikkoa? Soveltuisiko jokin tietty teknologia muita paremmin nykyisiin prosesseihin? Kun analyysit on tehty, voidaan aloittaa markkinakartoitus. Tavoitteena on löytää ratkaisuja, joilla voidaan hallita tunnistettuja riskejä ja jotka palvelevat haluttuja käyttötapauksia. Nämä tarpeet sisällytetään hankinta-asiakirjojen vaatimuksiin.

Koneoppivien valvontaratkaisuiden vertailu keskenään ei ole aina helppoa ja sitä hankaloittaa myös toimittajien käyttämien terminologioiden erot. Osa toimittajista voi puhua esimerkiksi ”korrelaatiosta” kun toiset puhuvat ”fuusiosta”, vaikka näillä saatetaan viitata hyvin samankaltaisiin ominaisuuksiin. Ennen laajamittaisen hankinnan tekemistä

ja käyttöönottoprojektin käynnistämistä on hyvä pyrkiä mahdollisuuksien mukaan hyödyntämään esimerkiksi Proof of Concept -tyyppisiä harjoituksia (ts. koekäyttöä), joissa eri teknologiota ja niiden sopivuutta ostavan organisaation käyttötapauksiin arvioidaan pienessä mittakaavassa, mutta riittävällä tarkkuudella.

6.2 Asiantuntemuksen tärkeys

Asiantuntijahaastatteluissa painotettiin koneoppimisen mahdollisuuksia ja reunaehtoehtoja tuntevien asiantuntijoiden mukana olon tärkeyttä hankintaa toteuttavassa tiimissä ja organisaatiossa, jotta odotukset uudesta teknologiasta saatavia hyötyjä kohtaan olisivat mahdollisimman realistiset. Koneoppimisen asiantuntemusta tarvitaan sekä käyttötapausten että hankinnan laajuuden määrittelyssä, mutta myös toimittajien antamien vastausten ymmärtämisessä ja arvioimisessa. Osiossa 5 todettiin, että useat koneoppimista hyödyntävät ratkaisut, mukaan lukien valvontateknologiat, ovat sisäiseltä toimintalogiikaltaan mustia laatikoita, joiden sisälle loppukäyttäjä ei näe. Tällöin korostuu tarve ymmärtää, miten teknologiatoimittajat itse valvovat myymiensä teknologioiden toimintaa. Hankintaprosessin yhteydessä on hyvä pyytää teknologiatoimittajia toimittaan selvennystä siitä, miten he itse testaavat ja suojaavat omia koneoppimismallejaan ja pyrkivät varmistumaan siitä, että mallit toimivat odotetulla tavalla. Tämä edellyttää ostajalta ja hankintaa mahdollisesti tukevilta konsulteilta koneoppimiseen liittyvää erityisosaamista, jotta valvontatoimien riittävyyttä ja asianmukaisuutta voidaan uskottavasti arvioida.

6.3 Pilvi- ja konesaliympäristöt

Mallien oppimiseen tarvittavan suuren datamäärän vuoksi monet markkinoilla olevat koneoppivat valvontaratkaisut perustuvat tiedon keräämiseen pilveen ja pilvilaskennan hyödyntämiseen. Useammassa haastattelussa esitettiin valvontateknologioiden kehitykseen ja saatavuuteen liittyvä huoli, jonka mukaan näitä teknologioita rakentavat yritykset panostavat entistä enemmän pilvikyvykkyyksiin perinteisten paikallisten konesalikyvykkyyksien jäädessä vähemmälle huomiolle. Jos organisaation tietojenkäsittelyinfrastruktuuri nojaa voimakkaasti paikallisiin konesaliympäristöihin pilvipalveluiden sijaan, voi riskinä olla, että pilvipalveluna tuotettavan valvonnan tasoista valvontaa ei ole paikallisesti mahdollista tuottaa tai ainakin se on hyvin hankalaa.

Toisaalta vaikka tietyt teknologiatoimittajat suuntautuisivat voimakkaasti kohti pilvipalveluita perinteisen konesalimaailman sijaan, jää konesaliympäristöihin tukeutuvien organisaatioiden markkinassa jäljelle jääneille toimittajille suhteellisesti suurempi asiakaskunta. Tästä syystä on todennäköistä, että koneoppimiseen tukeutuvia valvontaratkaisuja on tarjolla niin kauan kuin asiakkaita on myös pilvipalveluiden ulkopuolella. Lisäksi on

huomioitava, että suljetussa ympäristössä, joka ei ole auki julkiseen internetiin, riskitaso erityisesti kyberhyökkäyksille on matalampi kuin ympäristöissä, jotka ovat kytkeytyneet internetiin. Tällöin pilvilaskentaan tukeutuviin palveluihin verrattuna mahdollinen matalampi valvontakyvykkyys voi olla riskienhallintanäkökulmasta hyväksyttävämpää.

6.4 Kustannusnäkökulmat

Pelkkien lisenssikustannusten lisäksi tulee koneoppivien valvontateknologioiden kustannuksissa huomioida esimerkiksi käyttöönottoon ja ylläpitoon liittyvät kustannukset. Jotta investointi uuteen teknologiaan tuottaisi sillä tavoiteltua arvoa, on mahdollisesti palkattava tai koulutettava henkilöstöä tai ostettava asiantuntijatyötä. Lisäksi asiantuntijahaastatteluissa pidettiin tärkeänä arvioida sitä, kuinka hyvin uusi teknologia sopii organisaatiolla entuudestaan olevaan tietojenkäsittelyekosysteemiin ja minkä verran kompleksisuutta ja sitä kautta lisäkustannuksia se mahdollisesti aiheuttaa.

Kustannuksia arvioitaessa ja suunniteltaessa on hyvä ottaa huomioon myös datan tallennustilaan liittyvät kustannukset. Valvontajärjestelmien toiminta perustuu esimerkiksi päätelaitteilta, palvelimilta, palomuureilta ja kirjautumisratkaisuilta kerättyyn suureen määrään dataa, jonka säilyttäminen ja prosessointi ei useinkaan ole palveluntoimittajien puolelta ilmaista. Koneoppiviin järjestelmiin perustuvan digiturvan valvontapalvelun kustannukset saattavat myös kasvaa, jos palvelun hinta on sopimusmallissa sidottu hälytysten määrään; koneoppivat järjestelmät tuottavat usein perinteisiä sääntö- ja tunnistepohjaisia ratkaisuja enemmän hälytyksiä.

Koneoppivien valvontateknologioiden, kuten muidenkin digiturvainvestointien, rahallisia kustannushyötyjä on hyvin vaikea laskea tarkasti, koska syyn ja seurauksen vetäminen jonkin digiturvaan investoimatta jättämistä koskevan päätöksen ja toteutuneen riskin välille ei ole koskaan yksiselitteistä. Rahallista arvoa mittaavien liiketoimintaesimerkkilaskelmien sijaan soveltuvampi lähestymistapa on lähestyä asiaa koneoppivien valvontateknologioiden havainnointikykyä tehostavien kyvykkyyksien näkökulmasta ja siitä, miten paljon niiden avulla parantunutta organisaation digiturvan tasoa arvostetaan. Koneoppimiseen liittyvistä riskeistä asiantuntijahaastatteluissa todettiin, että koneoppivien valvontajärjestelmien käyttämättä jättäminen on suurempi riski kuin niiden käyttäminen. Samoin voitaneen todeta koneoppivien teknologioiden kustannuksista. Vaikka nämä järjestelmät saattavatkin lisätä kustannuksia, on niiden avulla saavutettu parantunut valvontakyky usein lisäkustannuksen arvoinen.

6.5 Käyttöönoton ensiaskeleet

Kun sopiva ratkaisu on löytynyt ja käyttöönottoa aloitetaan, painotettiin haastatteluissa, että uuteen koneoppimiseen tukeutuvaan valvontajärjestelmään tulee ensin suhtautua kuin testijärjestelmään. Tämä johtuu siitä, että käyttöönottovaiheessa oleva järjestelmä ei ole vielä oppinut tuntemaan valvottavaa ympäristöä, eikä kaikkien sen tuottamien hälytysten tarkkuuteen voida aluksi luottaa kovin vahvasti. Koneoppimismallin oppimisjakson pituus on hyvin järjestelmä-, ympäristö- ja käyttötapauskohtaista. Joissain tapauksissa oppiminen voi tuottaa merkittävää lisäarvoa jo muutamassa viikossa, mutta joskus se voi viedä pidempään. Ohjenuorana voidaan pitää sitä, että mitä enemmän ja pidempään dataa ympäristöstään koneoppimismallin annetaan kerätä, sen tarkemmaksi kehittyy järjestelmän kyky tunnistaa ympäristössä olevia oikeita riskin sisältäviä poikkeamia. Samalla laskee väärien positiivisten hälytysten määrä.

Uutta koneoppivaa valvontaratkaisua käyttöönotettaessa on uudesta teknologiasta ja sen kyvykkyyksistä hyvä informoida avoimesti myös henkilöstöä. Osin tätä käsitellään yhteistoimintaelimissä (ks. osio 3.1), mutta muutakaan aihetta koskevaa viestintää ei pidä laiminlyödä. Osa järjestelmistä voidaan esimerkiksi konfiguroida tuottamaan varoituksia loppukäyttäjille mahdollisia riskejä sisältävästä toiminnasta. Ennen tällaisten ominaisuuksien käyttöönottoa on tärkeää, että loppukäyttäjät tietävät, miten tällaisissa varoitustilanteissa tulee toimia ja mitä heiltä odotetaan.

7 Koneoppivien valvontajärjestelmien tulevaisuus ja tutkimus

Asiantuntijahaastatteluissa keskusteltiin myös koneoppivien valvontajärjestelmien tulevaisuudennäkymistä sekä näiden ratkaisujen tutkimuksen ja kehityksen tilasta. Tässä osiossa käydään läpi keskeiset havainnot näistä keskusteluista.

7.1 Kyvykkyyksien jalostumisesta kohti automatisoitua SIEMiä

Asiantuntijahaastatteluissa ennakoitiin nykyisten koneoppivien digiturvan valvontaratkaisuiden kyvykkyyksien jalostuvan ja kehittyvän edelleen. Esimerkiksi haavoittuvuuksien tunnistamisen ja paikallistamisen lisäksi koneoppiminen voi mahdollistaa haavoittuvien kohteiden automaattisen päivittämisen, jossa huomioidaan päivitysten riippuvuussuhteet ja niistä koituvat mahdolliset seuraukset palveluiden toiminnalle. Lisäksi koneoppimisen myötä digiturvan valvontaa saadaan ulotettua verkkotasolta syvemmälle sovellus- ja datan käsittelyn tasolle. Erään haastateltavan mukaan teknologiatoimittajien nykyiset mainoskalvot saattavat olla realistinen visio siitä, mitä näiden ratkaisuiden kyvykkyydet voivat käytännössä olla muutaman vuoden kuluttua.

Merkittävä myös digiturvan valvontaan vaikuttava kehityspolku on API-ekosysteemin kasvaminen, jolla tarkoitetaan eri palvelujen välistä tietojen vaihtoa ja hyödyntämistä rajapintojen kautta. Digiturvan valvonnan kohdalla rajapintojen myötä perinteinen yksisuuntainen lokien kerääminen valvottavista palveluista esimerkiksi SIEM-järjestelmään (Security Information and Event Management eli tietoturvatiedon ja -tapahtumien hallinta) korvautuu kaksisuuntaisella dialogilla valvottavien ja valvovien palveluiden välillä, mikä mahdollistaa havainnointikyvyn lisäksi automaattiset vastatoimet. Toiseksi kehityskuluksi nähtiin valvontaratkaisujen keskittyminen ja yhteensulautuminen, joka mahdollistaisi esimerkiksi sisäistä ja ulkoista verkkoa valvovien järjestelmien yhdistymisen sekä siirtymän paikallisesta näkyvyydestä kohti laajempaa ja kattavampaa näkyvyyttä ja tilannekuvaa. Koneoppimisen koettiin tuovan apua myös organisaatioiden oman digiturvakyvykkyyden ja -kypsyyden mittaamiseen ja arviointiin, tietoturvadokumentaation

tarkistamiseen, tekniseen testaamiseen, sekä työkaluja palveluiden, käyttäjien ja muiden resurssien sijainnista riippumatonta luottamussuhdetta painottavien Zero Trust -periaatteiden (34) käytännön toteutuksiin.

Osa haastatelluista koki koneoppimisen ja tekoälyn aiheuttavan sähköön tai internetiin verrattavissa olevan peruuttamattoman murroksen. Digiturvan valvonnassa käytännön esimerkkinä tällaisesta murroksesta voisi olla haastatteluissa ”SIEMin tappajaksi” luonnehdittu koneoppimispohjainen ratkaisu, joka toisi mukanaan siirtymän nykyisestä valvontadataa ihmisten hyödynnettäväksi keräävästä ja pureskelevasta välikerroksesta uskottavaksi, automaattiseksi ja nykyistä kustannustehokkaammaksi poikkeamankäsittelyksi. Tämä kehitys vaikuttaisi merkittävästi siihen, miten SOC-palveluita (Security Operations Center eli tietoturvavalvomo) tuotetaan ja kulutetaan. Haastatteluissa pitkän aikavälin koneoppimisen optimoiduksi tavoitetilaksi digiturvan valvonnassa asetettiin itseoppiva ja itse itseään päivittävä valvontajärjestelmä, joka palvelisi toimintaympäristössään sekä turvallisuuden että yksityisyydensuojan toteutumista. Panostukset koneoppimisen ja digiturvan tutkimukseen ovat keskeisessä roolissa tähän tavoitetilaan pääsemiseksi.

7.2 Reaalimaailman data ja yhteistyön merkitys

Koneoppivia valvontajärjestelmiä koskevan akateemisen tutkimuksen kannalta merkittävä haastatteluissa esille tullut ongelma ja tutkimuksen laatuun vaikuttava tekijä on reaalimaailmaan sovellettavissa olevien tietoaineistojen puute. Tutkimustulokset jäävät helposti teoreettisiksi, jos mallien toimintaa päästään testaamaan ja analysoimaan tutkimusolosuhteissa vain laboratoriodatalla organisaatioiden oikeassa toiminnassa syntyneen datan sijaan. Tämä puolestaan heikentää tutkimustulosten sovellettavuutta jokapäiväisen toiminnan tilanteisiin. Erityisen arvokasta aineistoa akateemiselle tutkimukselle digitaalisen turvallisuuden kehittämisen näkökulmasta olisivat oikeisiin hyökkäystilanteisiin liittyvät loki- ja sensoritiedot, jotka vain harvoin päätyvät kohdeorganisaation ja viranomaisten ulkopuolelle.

Toinen haastatteluissa esiin tullut tutkimuksia koskeva huomio liittyi koneoppimisen tutkimiseen niissä tilanteissa, joissa oppimisdataa on saatavilla rajoitetusti. Koneoppivien valvontajärjestelmien havaintotarkkuus on sitä tarkempi mitä enemmän näillä järjestelmillä on käytössään opetusdataa. Tutkimuksen kannalta kiinnostavaa sekä käytännön käyttötapauksia hyödyttävää olisi pyrkiä tutkimaan enemmän sitä, miten koneoppimismalli saadaan tuottamaan haluttuja lopputuloksia optimoidussa ajassa sellaisissa tilanteissa, joissa dataa on saatavissa vain rajoitettu määrä. Tällainen tutkimus palvelisi myös osiossa 6 esitettyä huolta siitä, että perinteisiin konesaliympäristöihin ei mahdollisesti tulevaisuudessa ole saatavilla pilviympäristöihin verrattavissa olevia valvontakyvykkyyksiä.

Lisäksi haastatteluissa mainittiin esimerkkinä lisäpanostuksia kaipaavasta tutkimusalueesta koneoppimismallien myrkyttämisen (ks. osio 5.1) sekä sen seurausten ja vastatoimien tutkiminen. Tämä tutkimusalue koettiin riskitasoltaan sellaiseksi, että se edellyttäisi entistä vahvempaa yhteistyötä yliopistojen ja korkeakoulujen sekä elinkeinoelämän ja julkisen sektorin välille, koska koneoppivia ratkaisuja kehittävillä yrityksillä yksinään ei välttämättä ole riittävää liiketaloudellista kannustinta tämän alueen tutkimiseen.

Koneoppimisen ja tekoälyn saralla toivottiin asiantuntijahaastatteluissa kokonaisuudessaan lisää sektorit ylittävää yhteistyötä. Sektorirajat ylittävä yhteistyö koettiin edellytyksenä laaja-alaisen tekoälyä ja koneoppimista koskevan osaamisen synnyttämiseksi sekä tulevaisuuden suunnan määrittämiseksi. Tätä osaamista tarvitaan näiden teknologioiden mukanaan tuomien mahdollisuuksien tehokkaaseen, mutta samalla ihmisoikeudet ja yksityisyydensuojan huomioivaan hyödyntämiseen sekä koneoppimiseen ja tekoälyyn liittyvien riskien hallitsemiseen.

8 Lopuksi

Tässä selvityksessä tarkasteltiin koneoppimisen käyttöä, roolia ja merkitystä digiturvan automaattisessa teknisessä valvonnassa. Selvityksen tulokset ja havainnot perustuivat ensi sijassa digiturvan ja koneoppimisen asiantuntijoiden haastatteluihin, jotka edustivat elinkeinoelämää, julkista sektoria ja tutkimuslaitoksia. Haastatteluissa esiin tulleita näkökulmia täydennettiin tutkimusartikkeleista sekä esimerkiksi tietoturvayhtiöiden julkaisuista kerätyillä tiedoilla ja näkemyksillä. Seuraavaksi käydään läpi haastatteluista nousseet tärkeimmät havainnot sekä pohditaan miten koneoppivien digiturvan valvontaratkaisuiden kehittämistä, käyttöä ja hyödyntämistä voitaisiin edistää.

8.1 Yhteenveto

Koneoppimisen käytön keskeisimmäksi hyödyksi digiturvan teknisessä valvonnassa haastateltavat kokivat parantuvan ja tehostuvan valvonta- ja havainnointikyvyn, koska tämä teknologia mahdollistaa entistä suurempien datamäärien keräämisen ja analysoinnin. Koneoppimiseen pohjautuvilla valvontaratkaisuilla on mahdollista havaita entuudestaan tuntemattomia riskitekijöitä ja nollapäivähyökkäyksiä, kun taas perinteiset sääntö- ja tunnistepohjaiset ratkaisut soveltuvat ainoastaan entuudestaan tunnettujen uhkien reaktiiviseen havaitsemiseen. Entuudestaan tuntemattomien uhkien havaitsemisen kyky perustuu käyttäjien ja muiden ympäristössä toimivien entiteettien käyttäytymisen valvontaan sekä poikkeamien tunnistamiseen sellaisesta käytöksestä, jonka koneoppiva valvontajärjestelmä on oppinut normaaliksi. Asiantuntijat kuitenkin painottivat sitä, että perinteisillä sääntö- ja tunnistepohjaisilla valvontaratkaisuilla on edelleen paikkansa ja paras lopputulos digiturvan teknisessä valvonnassa saadaan käyttämällä sekä sekä vanhoja että koneoppimisen mukanaan tuomia uusia kyvykkyyksiä rinnakkain.

Tehostuneen valvontakyvyn ja entuudestaan tuntemattomien uhkien havainnointikyvykkyyden kääntöpuolena koneoppivat valvontajärjestelmät voivat tuoda mukanaan merkittävän määrän vääriä hälytyksiä. Näiden määrää on kuitenkin mahdollista pienentää jatkamalla valvontajärjestelmän taustalla olevan koneoppimismallin opettamista käyttöönoton jälkeen. Muina koneoppiviin valvontajärjestelmiin liittyvinä keskeisinä ongelmina ja riskeinä asiantuntijat toivat esiin hyökkäykset koneoppimismalleja vastaan, läpinäkyvyyden puutteen sekä tietosuojariskit. Esimerkkinä hyökkäyksestä koneoppimismallia vastaan on yritykset myrkyttää oppimisdataa tavoitteena opettaa malli toimimaan

väärin mahdollista hyökkääjää palvelevalla tavalla. Läpinäkyvyyden osalta ongelmaksi koettiin näkyvyyden puute koneoppivien valvontajärjestelmien sisäiseen toimintalogiikkaan, minkä takia on hyvin vaikea tietää mihin järjestelmien tekemät päätökset tarkalleen perustuvat. Nykyinen tietosuojalainsäädäntö ei myöskään ole täysin yhteensopiva tekoälyn hyödyntämisen kanssa. Tietosuojaan liittyen haastatteluissa nousi esiin myös näkemys, jonka mukaan julkisten organisaatioiden ei ole mahdollista käytännössä tietää mihin heidän datansa lopulta päätyy ja millaista analyysia siihen kohdistuu, jos he hyödyntävät toiminnassaan globaaleja pilvipohjaisia digiturvan valvontateknologioita.

Koneoppiviin valvontajärjestelmiin liittyvien riskien hallinnan lähtökohdaksi nähtiin sen tiedostaminen ja tunnistaminen, että tämänkään teknologian käyttö ei ole riskitöntä. Lisäksi esille tuotiin häiriötilanteisiin varautumisen, resilienssin sekä vastuiden määrittelyn tärkeys. Tietosuojan osalta osa riskeistä on tällä hetkellä vain hyväksyttävä, jos järjestelmiä haluaa toiminnassaan hyödyntää. Haastatteluissa korostettiin myös koneoppivien valvontajärjestelmien oikeanlaisen toiminnan valvonnan tärkeyttä, eli valvontajärjestelmiä itseään tulee myös valvoa. Yhtenä keinona tämän toteuttamiseksi on vaatia koneoppivia valvontajärjestelmiä toimittavilta yrityksiltä dokumentaatiota ja todisteita siitä miten he itse pyrkivät valvomaan kehittämänsä järjestelmän oikeaoppista toimintaa. Vaikka koneoppiviin valvontajärjestelmiin liittyy useita riskejä, on hyvä pitää mielessä haastateltavien muistutus siitä, että näiden järjestelmien käyttämättä jättäminen on itsessään myös suuri riski. Koneoppivat valvontajärjestelmät ovat oleellisia työkaluja modernien digiturvauhkien torjunnassa ja niiden kyvykkyyksiä on mahdotonta korvata muilla keinoilla.

Koneoppivien teknisten valvontajärjestelmien hankintaprosessissa on muistettava sama käyttötapauslähtöisyys kuin minkä tahansa muunkin teknologian hankinnassa. Kustannusrakenteen osalta tulisi lisenssikulujen lisäksi huomioida ylläpitokulut, koneoppivan valvontajärjestelmän yhteensopivuus organisaation nykyisen teknologiaympäristön kanssa, asiantuntijoiden palkkaamiseen, kouluttamiseen tai ostamiseen liittyvät kulut sekä mahdolliset suurempaan hälytysmäärään tai tallennustilaan liittyvät palvelukulut. Haastatteluissa nousi esiin huoli siitä, että kaupallisten koneoppivien valvontaratkaisuiden kehitys keskittyy tällä hetkellä liiaksi pilviympäristöihin, eikä vastaavia kyvykkyyksiä ole välttämättä jatkossa saatavilla suljettuihin konesaliympäristöihin.

Tulevaisuuden kehityskulkuina nähtiin koneoppivien valvontaratkaisujen yleinen havainnointikyvyn jalostuminen. Digiturvan valvontatehtäviä ja sen markkinoita keskeisesti muuttava kehitysaskel olisi sellaisen valvontajärjestelmän kehittäminen, joka uskottavasti ja luotettavasti voisi automatisoidusti hoitaa suuren osan nykyisistä ihmisvoimin tehtävistä valvomotehtävistä. Tässä pisteessä ei kuitenkaan vielä olla. Haastateltujen asiantuntijoiden mukaan koneoppivat valvontajärjestelmät helpottavat tietoturvavalvomotyöntekijöiden arkea ensimmäisen tason analyysissa sekä suodattamalla olennaisen valvontatiedon ihmisasiantuntijoiden käyttöön. Monimutkaisempien sekä

laajempivaikutteisten päätösten ja arviointien tekemisessä ihminen on edelleen keskeisessä roolissa, eikä näitä tehtäviä voida nykyisillä koneoppivien valvontateknologioiden kyvykkyyksillä ulkoistaa koneoppivalle järjestelmälle.

Koneoppivia valvontaratkaisuja koskevan tutkimuksen osalta haastatteluissa esitettiin toive organisaatioille reaalimaailman datan antamisesta yliopistojen ja korkeakoulujen käyttöön parempien sekä käytäntöön soveltuvampien kyvykkyyksien ja ominaisuuksien tutkimiseksi ja kehittämiseksi. Lisäksi haastatteluissa painotettiin sektoreiden rajat ylittävän yhteistyön tärkeyttä, jotta voidaan parhaalla mahdollisella tavalla kehittää ja tutkia tekoälyjärjestelmiä, jotka ovat sekä tehokkaita, turvallisia että ihmisoikeudet ja yksityisyydensuojan huomioon ottavia.

8.2 Jatkosuositukset

Koneoppimisen käyttöä digiturvan teknisessä valvonnassa koskevan selvitystyön perusteella on nähtävissä, että koneoppivat valvontaratkaisut ovat keskeisiä puolustustyökaluja nykyaikaisessa digiturvaympäristössä. Näiden avulla on mahdollista havaita sellaisia uhkia, joihin perinteisillä valvontajärjestelmillä ei ole kykyä. Lisäksi näillä järjestelmillä on mahdollista automatisoida osa tietoturvavalvomotyöstä, mikä mahdollistaa ihmisasiantuntijoiden ajankäytön keskittämisen vakavimpien havaintojen tutkimiseen ja selvittämiseen. Koneoppiviin järjestelmiin liittyy tiettyjä ominaisriskejä, mutta myös monet yleiset teknologioiden lainalaisuudet pätevät näihin, kuten käyttötapaus- ja riskilähtöisyyden sekä vastuiden määrittelyn tärkeys.

Asiantuntijahaastatteluiden perusteella on tunnistettavissa kaksi merkittävää jatkokehityskokonaisuutta koneoppivien valvontajärjestelmien tehokkaammaksi hyödyntämiseksi sekä näiden tutkimuksen ja kehityksen edistämiseksi:

1. Koneoppivia järjestelmiä koskevaa ohjeistusta tulee tarkentaa ja tehdä käytännönläheisemmäksi.
2. Tekoälyä ja digitaalista turvallisuutta koskevan sektoreiden tulee tiivistää yhteistyötään.

Ensimmäisen kohdan osalta koneoppivien valvontaratkaisuiden ja myös muiden tekoälyyn pohjautuvien teknologioiden käyttöä ja hyödyntämitä vaikeuttaa tiettyjen juridisten pelisääntöjen epäselvyys ja tulkinnanvaraisuus, kuten osin nykyinen tietosuojalainsäädäntö. Koneoppivien valvontaratkaisuiden mahdollisimman tehokkaan hyödyntämisen kannalta olisi tärkeää, että näitä järjestelmiä käyttäville tai niiden käyttöä harkitseville organisaatioille olisi tarjolla mahdollisimman käytännönläheistä ohjeistusta siitä, millä periaatteilla näitä järjestelmiä voidaan ottaa käyttöön ja mitä käytön aikana tulisi osata

ottaa huomioon. Esimerkkejä tällaisista ohjeista voisivat olla entistä konkreettisemmat ohjeet pilvipohjaisten valvontapalvelujen käytöstä tai koneoppivien valvontajärjestelmien hankintaohjeistus, joka voisi sisältää esimerkiksi vaatimuslistauksia tällaisille järjestelmille. Oleellisen ja ajantasaisien ohjeistuksen tuottamista vaikeuttaa kuitenkin tekoäly- ja koneoppimisteknologian kehityksen nopeus. Ohjeita pitäisi pystyä päivittämään jatkuvasti, jotta ne vastaisivat teknologian ajantaisaista tilannetta.

Lisätoimia tarvitaan asiantuntijahaastatteluissa toivotun sektorien välisen yhteistyön tiivistämiseksi tekoälyn ja digitaalisen turvallisuuden tutkimuksen, kehittämisen, ennakoinnin ja suunnittelun alueilla. Keinoina tähän voisivat olla nykyisten yhteistyöfoorumeiden toiminnan viestinnän tehostaminen, mahdollisten uusien ryhmien perustaminen tai yhdistäminen sekä eri sektoreiden välisten tutkimushankkeiden tukeminen. Automaattiset ja oppivat järjestelmät ja niihin liittyvät kehityskulut ovat tärkeitä huomioida myös tulevissa julkisen hallinnon digitaalisen turvallisuuden kehittämisohjelmissa ja -hankkeissa.

Lähteet

1. **Gartner.** Gartner Survey Finds CEOs Cite AI as the Top Disruptive Technology Impacting Industries. [Online] 17. 5 2023. https://www.gartner.com/en/newsroom/press-releases/2023-05-17-gartner-survey-finds-ceos-cite-ai-as-the-top-disruptive-technology-impacting-industries

2. **Business Insider.** In the battle between AI & Metaverse, CEOs choose AI. [Online] 5. 7 2022. https://www.businessinsider.in/tech/enterprise/in-the-battle-between-ai-metaverse-ceos-choose-ai/articleshow/92674686.cms

3. **Future of Life Institute.** Pause Giant AI Experiments: An Open Letter. [Online] 23. 3 2023. https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

4. **World Economic Forum.** Future Series: Cybersecurity, emerging technology and systemic risk. [Online] 2022. https://www3.weforum.org/docs/WEF_Future_Series_Cybersecurity_emerging_technology_and_systemic_risk_2020.pdf.

5. **KPMG.** KPMG Cyber trust insights 2022. [Online] 2022. https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2022/10/kpmg-cyber-trust-insights-2022.pdf.

6. **Vähä-Sipilä, A.;Marchal, S. ja Aksela, M.** Tekoälyn soveltamisen kyberturvallisuus ja riskienhallinta (Traficomin tutkimuksia ja selvityksiä 9/2021). [Online] 2021. https://www.traficom.fi/sites/default/files/media/publication/Teko%C3%A4lyn%20soveltamisen%20kyberturvallisuus%20ja%20riskienhallinta.pdf.

7. **Aksela, M.;ym.** Tekoälyn mahdollistamat kyberhyökkäykset (Traficomin tutkimuksia ja selvityksiä 30/2022). [Online] 2022. https://www.traficom.fi/sites/default/files/media/publication/TRAFICOM_Teko%C3%A4lyn_mahdollistamat_kyberhy%C3%B6kk%C3%A4ykset%202022-12-12_web.pdf.

8. **Valtiovarainministeriö.** Kansallinen tekoälyohjelma AuroraAI. [Online] https://vm.fi/tekoalyohjelma-auroraai.

9. **OECD.** Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI (OECD Digital Economy Papers, No. 349). [Online] 2023. https://doi.org/10.1787/2448f04b-en.

10. **Euroopan komissio.** Luotettavaa tekoälyä koskevat eettiset ohjeet (Tekoälyä käsittelevä korkean tason asiantuntijaryhmä). [Online] 2019. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

11. **OECD.** OECD Policy Framework on Digital Security: Cybersecurity for Prosperity. [Online] 2022. https://doi.org/10.1787/a69df866-en.

12. **Valtiovarainministeriö.** Julkisen hallinnon digitaalinen turvallisuus (Valtiovarainministeriön julkaisuja 2020:23). [Online] 2020. http://urn.fi/URN:ISBN:978-952-287-857-1.

13. **Euroopan parlamentti.** Mitä tekoäly on ja mihin sitä käytetään? [Online] 4. 9 2020. https://www.europarl.europa.eu/news/fi/headlines/society/20200827STO85804/mita-tekoaly-on-ja-mihin-sita-kaytetaan.

14. **Tuominen, H. ja Neittaanmäki, P.** [Online] 2019. https://tim.jyu.fi/view/kurssit/tie/tiep1000/tekoalyn-sovellukset/kirja.

15. **Solin, A.** Tekoäly oppii. [Online] https://www.eduskunta.fi/FI/vaski/JulkaisuMetatieto/Documents/EDK-2021-AK-405320.pdf.

16. **Technative.** Why Unsupervised Machine Learning is the Future of Cybersecurity. [Online] 9. 9 2021. https://technative.io/why-unsupervised-machine-learning-is-the-future-of-cybersecurity/.

17. **Crowdstrike.** Machine learning (ML) & cybersecurity. How is ML used in cybersecurity? [Online] 14. 9 2022. https://www.crowdstrike.com/cybersecurity-101/machine-learning-cybersecurity/.

18. 18. **Winder, P.** Automating Cyber-Security with Reinforcement Learning. [Online] https://winder.ai/automating-cyber-security-with-reinforcement-learning/.

19. **Thompson, N. C.;ym.** The Computational Limits of Deep Learning (MIT Initiative on the Digital Economy Research Brief 2020 Vol. 4). [Online] 2020. https://ide.mit.edu/wp-content/uploads/2020/09/RBN.Thompson.pdf.

20. **Macas, M.;Wu, C. ja Fuertes, W.** A survey on deep learning for cybersecurity (Computer Networks 212). [Online] 2022. https://doi.org/10.1016/j.comnet.2022.109032.

21. **Pawlicki, M.;Kozik, R. ja Choras, M.** A survey on neural networks for (cyber-) security and (cyber-) security of neural networks (Neurocomputing 500, 1075-1087). [Online] 2022. https://doi.org/10.1016/j.neucom.2022.06.002.

22. **Euroopan parlamentti ja neuvosto.** Asetus (EU) 2016/679 luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta ja direktiivin 95/46/EY kumoamisesta (yleinen tietosuoja-asetus). [Online] 2016. https://eur-lex.europa.eu/legal-content/FI/TXT/PDF/?uri=CELEX:32016R0679.

23. **Laki yksityisyyden suojasta työelämässä.** 759/2004. [Online] https://www.finlex.fi/fi/laki/ajantasa/2004/20040759.

24. **Hallintolaki.** 434/2003. [Online] 2003. https://www.finlex.fi/fi/laki/ajantasa/2003/20030434.

25. **Tiedonhallintalaki.** 906/2019. [Online] https://www.finlex.fi/fi/laki/ajantasa/2019/20190906.

26. **Laki sähköisen viestinnän palveluista.** 917/2014. [Online] https://www.finlex.fi/fi/laki/ajantasa/2014/20140917.

27. **Euroopan parlamentti ja neuvosto.** Ehdotus asetukseksi tekoälyä koskevista yhdenmukaistetuista säännöistä (tekoälysäädös) ja tiettyjen unionin säädösten muuttamisesta COM (2021) 206 2021/0106 (COD). [Online] https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0007.02/DOC_1&format=PDF.

28. **European Comission.** Regulatory framework proposal on artificial intelligence. [Online] https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

29. **Euroopan parlamentti ja neuvosto.** Direktiivi (EU) 2022/2555 toimenpiteistä kyberturvallisuuden yhteisen korkean tason varmistamiseksi kaikkialla unionissa, asetuksen (EU) N:o 910/2014 ja direktiivin (EU) 2018/1972 muuttamisesta sekä direktiivin (EU) 2016/1148 kumoamisesta (NIS 2). [Online] https://eur-lex.europa.eu/legal-content/FI/TXT/PDF/?uri=CELEX:32022L2555.

30. **Y., Guo.** A review of Machine Learning-based zero-day attack detection: Challenges and future directions. Computer Communications 198, 175-185. [Online] 2022. https://doi.org/10.1016/j.comcom.2022.11.001.

31. **Logpoint.** What is User and Entity Behavior Analytics? A complete guide to UEBA, how it works, and its benefits. [Online] 15. 9 2020. https://www.logpoint.com/en/blog/ueba-user-and-entity-behavior-analytics/.

32. **Oprea, A. ja Vassilev, A.** Adversial Machine Learning. A Taxonomy and Terminology of Attacks and Mitigations (NIST AI 100-2e2023 ipd). [Online] 2023. https://doi.org/10.6028/NIST.AI.100-2e2023.ipd.

33. **Yampolskiy, R.** AI is the Future of Cybersecurity, for Better and for Worse. *Cybersecurity. Insights You Need from Harvard Business Review (s. 141–145).* Boston : Harvard Business Review Press, 2019.

34. **Rose;ym.** Zero Trust Architecture (NIST Special Publication 800-207). [Online] 2020. https://doi.org/10.6028/NIST.SP.800-207.

VALTIOVARAINMINISTERIÖ
FINANSMINISTERIET

VALTIOVARAINMINISTERIÖ
Snellmaninkatu 1 A
PL 28, 00023 VALTIONEUVOSTO
Puhelin 0295 160 01
vm.fi

ISSN 1797-9714 (pdf)
ISBN 978-952-367-641-1 (pdf)

Elokuu 2023