Tekoälyn vaikutukset yrityksen tietoturvallisuuteen

Eero Kolkki

Haaga-Helia ammattikorkeakoulu
Tradenomi
Amk-opinnäytetyö
2024

Tiivistelmä

Tekijä(t)
Eero Kolkki

Tutkinto
Tradenomi

Raportin/Opinnäytetyön nimi
Tekoälyn vaikutukset yrityksen tietoturvallisuuteen

Sivu- ja liitesivumäärä
32

Tämän opinnäytetyön tarkoituksena oli selvittää tekoälyn vaikutuksia yrityksen tietoturvallisuuteen. Tutkimuksessa tarkasteltiin mitä uhkia generatiivisen tekoälyn käyttöönotto tuo yrityksen tietoturvallisuudelle ja miten uhkia voidaan torjua. Lisäksi tutkittiin tekoälyn potentiaalia uhkien torjunnassa ja yrityksen tekoälyjärjestelmän turvallisen kehittämisen käytäntöjä.

Tutkimuksen tietoperustassa käytiin läpi tietoturvallisuuden ja tekoälyn käsitteitä. Tietoturvallisuudesta käsiteltiin sen edellytykset kuten luottamuksellisuus, eheys, käytettävyys ja todentaminen sekä uhat ja tietoturvatoimenpiteet. Tekoälystä kerrottiin sen määritelmä, kehitys ja avattiin generatiivisen tekoälyn käsitettä.

Tutkimusmenetelmäksi valittiin narratiivinen kirjallisuuskatsaus, jossa käytettiin lähteinä pääosin tuoreita markkinatutkimuksia, akateemisia tutkimuksia sekä eri valtioiden kyberturvallisuusvirastojen julkaisuja alan nopean kehittymisen vuoksi. Tutkimuksen päälähteet sisältävät markkinatutkimuksia Hiddenlayeriltä ja Zscaleriltä sekä muutamia valtiollisten tahojen julkaisuja.

Tulokset kertoivat, että generatiivinen tekoäly helpottaa verkkorikollisten työtä ja ennestään tutut uhat moninkertaistuvat. Lisäksi löytyi uusia generatiivisen tekoälyn tuomia uhkia, kuten generatiivisen tekoälyn voimistamat perinteiset tietoturvallisuusuhat, syväväärennökset, yritysten omien tekoälyprojektien uhat, yrityksen tiedon menetys generatiivisen tekoälyn ohjelmiin ja generatiivisen tekoälyn hallusinaatiot. Tutkimuksessa nostettiin esiin tietoturvallisuutta parantavia toimenpiteitä, joita lähdemateriaalissa suositeltiin tekoälyn tietoturvallisuuden varmistamiseksi. Näitä olivat turvallisuuskehysten käyttöönotto, nollaluottamus-tietoturvamalli ja tekoälysivustojen käytön estäminen. Lisäksi käytiin läpi yrityksen omien tekoälyprojektien turvallisia käytäntöjä ja sääntelyä. Selvisi myös, että tekoäly itsessään on tietoturvan tulevaisuudessa hyvin isossa roolissa, sillä sitä tarvitaan torjumaan reaaliajassa tekoälyn mahdollistamia hyökkäyksiä hyvin monilla tietoturvan osa-alueilla.

Tulokset tarjoavat tuoretta tietoa ongelmista, joita yritykset kohtaavat ottaessaan tekoälyä käyttöön, ja käsitystä siitä, miten ison ongelman generatiivinen tekoäly aiheuttaa yrityksen tietoturvallisuudelle. Lisäksi raportin lukija oppii erilaisista keinoista, joilla näitä uhkia voidaan torjua.

Opinnäytetyö aloitettiin elokuussa 2024, ja se valmistui lokakuussa 2024.

Asiasanat
Generatiivinen tekoäly, riskienhallinta, tekoäly, tietoturvallisuus, yritykset

Sisällys

1 Johdanto

Tekoälyn kehittyminen on mullistanut monia aloja, mukaan lukien tietoturvallisuuden alan. Muutoksessa avainasemassa on ollut generatiivisen tekoälyn nopea kehitys ja sen käytön lisääntyminen vain muutaman viime vuoden aikana. Yritykset ovat sen avulla voineet tehostaa toimintaansa ja saada kilpailuetua. Tekoälyllä on kuitenkin iso varjopuoli, se aiheuttaa pahoja uhkia yritysten tietoturvallisuudelle. Tekoäly on antanut verkkorikollisille uusia työkaluja ja helpottanut heidän toimintaansa pahentaen siten jo olemassa olevia uhkia ja luoden uusia. Tämän opinnäytetyön on tarkoitus selvittää, miten tekoäly vaikuttaa yritysten tietoturvallisuuteen, ja miten uhkiin voidaan varautua.

Generatiivisen tekoälyn sovellukset ovat tulleet osaksi monen työpaikan perustoimintaa vauhdilla. Yritykset ottavat käyttöön erilaisia generatiivisen tekoälyn sovelluksia, kuten OpenAI:n ChatGPT, tai tekoäly tulee osaksi heidän jo käyttämiään työkaluja. Esimerkiksi ohjelmoijat saavat paljon apua koodin kirjoittamiseen GitHub Copilotista tai graafinen suunnittelija voi hyödyntää tekoälyä Adoben suosituissa ohjelmistoissa. Opinnäytetyössä yhtenä lähteenä käytetyn Hiddenlayerin markkinatutkimuksen (2024, 3–5) mukaan 98 prosenttia haastatelluista tietoturvallisuuspäälliköistä etsii teknisiä keinoja tekoälyn ja koneoppimisen mallien tietoturvan parantamiseen ja 58 prosenttia heistä uskoo, että tämänhetkiset tietoturvaratkaisut eivät ole riittäviä.

Tekoäly on aihealueena todella laaja, sen vuoksi opinnäytetyötä on rajattu tarkastelemaan tilannetta yritysten näkökulmasta, ja tekoälyn osalta painopiste on generatiivisessa tekoälyssä. Lähteet ja tietoturvallisuusratkaisut painottuvat isompiin yrityksiin. Opinnäytetyön tietoperustassa käydään läpi tietoturvallisuuden ja tekoälyn peruskäsitteitä.

Taulukossa 1 käydään peittomatriisissa läpi tutkimuksen alaongelmat, jotka vastaavat tutkimuksen pääongelmaan: Miten tekoäly vaikuttaa yrityksen tietoturvallisuuteen.

Taulukko 1. Peittomatriisi

Alaongelmat (tutkimuskysymykset)	**Tutkimus (luku)**	**Johtopäätökset (luku)**
Mitä uhkia generatiivinen tekoäly tuo yrityksen tietoturvallisuudelle	5.1, 5.1.1, 5.1.2, 5.1.3	6.1.1
Miten tekoälyn uhkia voidaan torjua yrityksissä?	5.2.1, 5.2.2, 5.2.3, 5.2.4	6.1.2
Miten tekoäly voi auttaa uhkien torjunnassa?	5.2.1, 5.2.4	6.1.3
Yrityksen tekoälyjärjestelmän turvallisen kehittämisen käytännöt	5.2.5	6.1.4

Opinnäytetyön tutkimusmenetelmänä on narratiivinen kirjallisuuskatsaus, joka soveltuu hyvin nopeasti kehittyvään aiheeseen. Alan nopean kehityksen vuoksi päälähteitä tutkimuksessa ovat pääosin markkinatutkimukset, akateemiset tutkimukset ja kyberturvallisuusvirastojen julkaisut.

Tässä opinnäytetyössä on kartoitettu keskeisiä käsitteitä, jotka liittyvät tekoälyn käyttöön liittyviin uhkiin ja niiden torjuntaan. Joitakin käsitteitä ja niiden välisiä yhteyksiä on havainnollistettu kuvien avulla. Selitän seuraavassa taulukossa (2) joitakin tutkimuksen peruskäsitteitä, niiden tarkempi määrittely ja lähdemateriaali selviää raportin edetessä.

Taulukko 2. Käsitteiden selitteet

Käsite	**Selitys**
Generatiivinen tekoäly	Tekoälymalli, joka luo uutta tekstiä, kuvaa tai ääntä algoritmien avulla.
Generatiivisen tekoälyn hallusinaatiot	Generatiivisen tekoälymallin tuloksen virheet.
Koulutusdata	Tekoälymallin koulutusmateriaali, eli kokoelma tietoa, jota käytetään mallin kouluttamiseen.
Mallin varastaminen	Tekoälymallin uhka, jossa hyökkääjä selvittää mallin rakenteita.
Mallin rajoitusten kiertäminen	Tekoälymallin uhka, missä käyttäjä kiertää jollakin tavalla mallin sääntöjä.
Markkinakatsaus	Jonkin alan nykytilaa joltakin kannalta tutkiva julkaisu.
Nollaluottamus-tieturvamalli	Tietoturvamalli, jossa mikään toimija ei ole lähtökohtaisesti luotettu.
Sosiaalinen hakkerointi (social engineering)	Tietoturvallisuusuhka, jossa hyökkäävä taho huijaa kohdetta paljastamaan tietoa tai antamaan rahaa.
Syväväärennös (deepfake)	Aidontuntuinen keinotekoisesti tuotettu väärennös.
Tietomyrkytys	Tekoälymallin uhka, jossa hyökkääjä yrittää manipuloida mallin toimintaa vastauksillaan.
Tietoturvakehys	Kokoelma sääntöjä ja ohjeita tietoturvan järjestämiseksi ja hallitsemiseksi.

2 Tietoturvallisuus

2.1 Tietoturvallisuuden ja kyberturvallisuuden määritelmät ja tärkeys

Tieto- ja kyberturvallisuudesta puhuttaessa termit menevät usein sekaisin, sillä niiden alueet ovat osittain päällekkäisiä. Niissä on kuitenkin eronsa. Siinä missä kyberturvallisuus keskittyy laitteiden, tiedon ja tietojärjestelmien turvallisuuden takaamiseen verkkoympäristössä, ottaa tietoturvallisuus asiaan laajemman näkökulman tuoden mukaan digiympäristön ulkopuoliset asiat. Näitä ovat tietoon pääsyn fyysinen rajoittaminen ja tiedon tallentaminen. Tieto- ja kyberturvallisuuden kohtaamat uhat eroavat myös toisistaan. Kun kyberturvallisuudessa estetään esimerkiksi troijalaisten kaltaisten haittaohjelmien tekemiä vahinkoja, niin tietoturvallisuuteen kuuluu lisäksi väärän tiedon levittämisen ja tekemisen torjunta. Kyberturvallisuutta pidetäänkin yhtenä tietoturvallisuuden osa-alueista. (F-Secure s.a.)

Tieto- ja kyberturvallisuuden merkitys kasvaa jatkuvasti, kun teknologia kehittyy ja valtaa uusia osa-alueita. Mikäli kyberturvallisuudesta ei pidetä huolta, koko yrityksen toiminta voi olla vaakalaudalla. Maailma on täynnä erittäin huonosti suojattuja järjestelmiä, jotka sisältävät arkaluontoisia tietoja. Esimerkkinä vaikkapa Vastaamon tapaus, jossa asiakasrekisterin tiedot olivat heikosti suojattuna.

F-Securen ”Mitä on kyberturvallisuus?” -artikkelissa mainitaan esimerkiksi yhteiskunnallisesti kriittiset toimijat ja niiden toiminnan turvaamisen tärkeys. Yritykset saattavat olla osana yhteiskunnan keskeisiä toimintoja, kuten maksutoimintaa tai infrastruktuuria, ja sen takia niiden on suhtauduttava kyberturvallisuuteen erittäin vakavasti. Kyberturvallisuus ei kuitenkaan ole vain tämän ryhmän huolenaiheena, vaan se koskee jokaista ihmistä. (F-Secure s.a.)

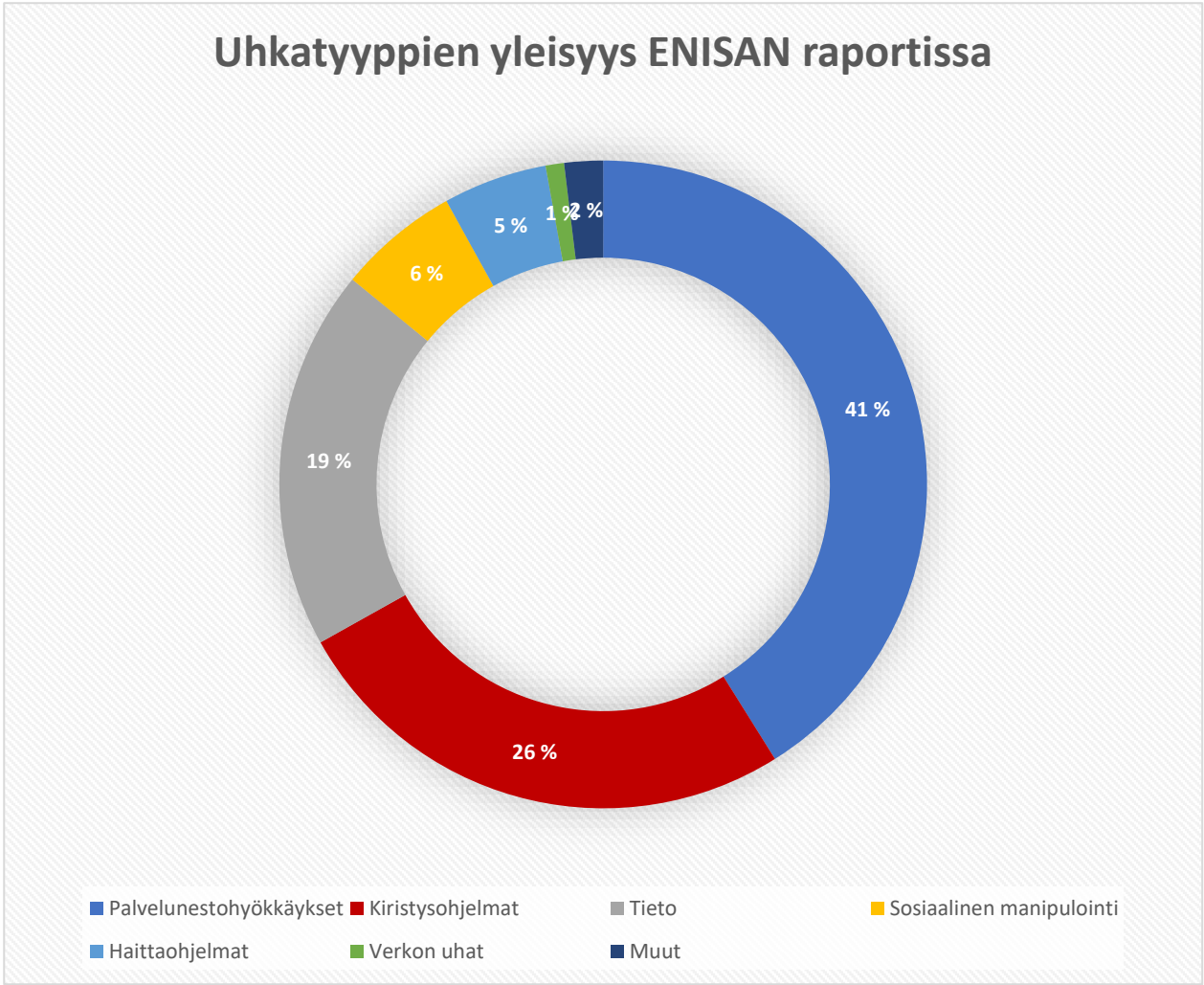
Tietoturvallisuus perustuu siihen, että huolehditaan neljästä tietoturvan edellytyksestä. Näitä ovat luottamuksellisuus, eheys, käytettävyys ja todentaminen. Luottamuksellisuus tarkoittaa sitä, että tietoon pääsevät käsiksi vain ne henkilöt, joilla on oikeus siihen. Eheys tarkoittaa tiedon sisällön säilymistä sellaisena, etteivät ulkopuoliset tekijät pääse millään tavalla vaikuttamaan siihen. Käytettävyys tarkoittaa, että tiedot ovat niihin oikeutettujen ihmisten käytettävissä, ja todentamisessa taas on kyse siitä, että käyttäjä voidaan luotettavasti tunnistaa esimerkiksi sisäänkirjautumisessa. (Helsingin yliopisto s.a.)

Jos yrityksen tietoturvallisuudesta ei huolehdita riittävästi, seuraukset voivat olla tuhoisia. Anna Anderssonin ym. (2022, 1) mukaan organisaatioiden kannattaa panostaa tietoturvallisuuteen, koska yritys voi kärsiä isoja taloudellisia tappioita, maine voi kärsiä ja pahimmassa tapauksessa yhteiskunnan kriittiset toiminnot voivat häiriintyä.

2.2 Tietoturvallisuusuhat

Tässä kappaleessa tutustutaan erilaisiin tietoturvallisuusuhkiin. Koska yhä suurempi osa rahaliikenteestä kulkee internetin välityksellä, avautuu verkkorikollisille yhä enemmän mahdollisuuksia. Lisäksi teknologian kehitys antaa rikollisille uusia työkaluja ja voimistaa ilmiötä. ENISAn raportin mukaan 2023–2024 tarkastelujaksolla rikottiin ennätyksiä kyberturvallisuushyökkäysten määrässä, monipuolisuudessa ja kustannuksissa. Erilaiset konfliktit muokkaavat uhkakarttaa ja näkyvät näissä hyökkäyksissä hyvin selkeästi. (ENISA 2024, 6.)

Seuraavassa kuvassa on esitetty ENISAn raportissa analysoitujen uhkien määrällistä jakautumista. ENISAn raportissa uhkia oli jaoteltu hieman eri perustein kuin tässä raportissa. (Kuva 1.)



Kuva 1. Analysoitujen tapausten jakautuminen uhkatyypeittäin (heinäkuu 2023 - kesäkuu 2024) (mukaillen ENISA 2024, 9)

2.2.1 Haittaohjelmat

Mikko Hyppösen (2022, 33) mukaan haittaohjelmat muodostavat suurimman yksittäisen uhan tietoturvallisuudelle. Erilaisia haittaohjelmia ovat esimerkiksi troijalaiset, madot, kiristysohjelmat, virukset ja näppäilyntallennusohjelmat. ENISAn raportin mukaan haittaohjelmat tarkoittavat erilaisia ohjelmia tai laiteohjelmistoja, joilla on negatiivinen vaikutus järjestelmän saatavuuteen, luotettavuuteen tai eheyteen. Haittaohjelmien tavoite voi olla esimerkiksi asentaa muita haittaohjelmia tai saada haltuunsa tietoa, verkkoja ja järjestelmiä. (ENISA 2024, 7–56.)

2.2.2 Hakkerointi

Toinen merkittävä uhka on hakkerointi, joka tarkoittaa jonkin tahon tietokoneelta suorittamaa hyökkäystä, jossa tarkoitus on esimerkiksi varastaa tietoa tai vahingoittaa yrityksen toimintaa. Hakkeroinnissa voidaan hyödyntää haittaohjelmia ja tietoturva-aukkoja verkossa tai paikan päällä. (Microsoft s.a.)

2.2.3 Sosiaalinen hakkerointi

Heikoin lenkki yrityksen tietoturvassa ovat yrityksen työntekijät. Vaikka yritys olisi hoitanut tietoturvan täydellisesti, voi huolimaton työntekijä altistaa yrityksen erilaisille uhille, kuten tietovuodoille, tietojen kalastelulle tai petoksille. Microsoftin (Microsoft s.a.) mukaan sosiaalisessa hakkeroinnissa hyökkääjä useimmiten yrittää tekeytyä toiseksi henkilöksi tai luoda uskottavan peitetarinan, johon lankeava uhri päätyy antamaan hyökkääjälle jotain, mitä tämä haluaa, kuten rahaa tai tietoja.

2.2.4 Palvelunestohyökkäykset

Yrityksen toimintaa voidaan hankaloittaa myös palvelunestohyökkäyksillä, joissa palvelua yritetään kaataa lähettämällä niin suuri määrä liikennettä, että kapasiteetti ylittyy. Hyökkäys voidaan tosin toteuttaa myös hyödyntämällä verkkolaitteessa olevaa haavoittuvuutta. Suurin osa hyökkäyksistä toteutetaan hajautetuista bottiverkoista, jotka koostuvat yleensä laitteista, joiden omistajat eivät tiedä laitteen olevan osa bottiverkkoa. (Kyberturvallisuuskeskus 2022, 3.)

2.2.5 Tiedon uhat

Yksi tietoon liittyvä uhka on tietomurto, jossa verkkorikollinen hyökkää johonkin tahoon tarkoituksenaan tiedon varastaminen. Yleisen tietosuoja-asetuksen (GDPR) mukaan tietoturvaloukkaus on mikä tahansa turvallisuuspoikkeama, joka johtaa lähetetyn, säilötyn tai tallennetun tiedon katoamiseen, laittomaan tuhoamiseen, muuttamiseen tai luvattomaan jakamiseen. (ENISA 2024, 8.)

2.3 Tietoturvallisuuden toimenpiteet

Yrityksillä on käytössä monenlaisia keinoja tietoturvallisuuden varmistamiseksi. Koska se on todella moniulotteinen asia, on olemassa erilaisia standardeja, joissa kuvataan tietoturvallisuuden hallintajärjestelmän vaatimukset. Yksi tunnetuimpia standardeja on kansainvälinen ISO 27001. (Andersson et al. 2022, 1.) Suomen standardisoimisliiton (2023) mukaan tietoturvallisuuden hallintajärjestelmän tehtävä on suojata tiedon luottamuksellisuus, eheys ja saatavuus riskienhallintaprosessilla. Tietoturvallisuuden hallintajärjestelmän kuuluu olla osa organisaation yleisiä johtamis- ja hallintarakenteita sekä prosesseja. (SFS 2023, 5.)

Yritysten kannattaa suojata tietonsa ottamalla käyttöön tietoturvallisuuden ja tietojärjestelmien suojausjärjestelmät, ja riskien minimoimisessa standardisointi on merkittävässä roolissa. Standardit sisältävät ohjeita siitä, miten nämä toimenpiteet suunnitellaan ja toteutetaan. Standardeja on kuitenkin kritisoitu siitä, että ne tarjoavat liian kapean yhden ympäripyöreän ratkaisun ja skaalautuvat huonommin pienempiin yrityksiin. (Andersson et al. 2022, 1.)

Mikäli yritys ei hoida tietoturvallisuutta paikallisten lakien mukaisesti, voivat vastuussa olevat henkilöt tai yritys joutua vastaamaan siitä oikeudessa. Tietoturvallisuus kuitenkin muuttuu ja elää teknologian kehittyessä, minkä takia kaikilta uusilta uhkilta ei voida heti suojautua, eikä se ole aina edes mahdollista. Hyökkäävä taho voi olla valtiollinen toimija, jolla on käytössään niin suuret resurssit ja keinot, että he löytävät tiensä läpi vaikka täysin muusta verkosta irrallaan olevaan huippusalaiseen laitokseen. Tietoturvatoimenpiteillä voidaan kuitenkin vähentää yrityksen alttiutta tietoturvallisuusuhkille.

2.3.1 Salaus

Salauksen avulla voidaan estää muiden kuin vahvistettujen käyttäjien pääsy tietoon. Esimerkiksi Microsoftin Bitlocker-ohjelmalla voidaan salata koko tietokone. Microsoftin (Microsoft s.a.) mukaan salaus on tietojen muuntamista koodiksi. Lindroos (Lindroos 2019, 18) mainitsee, että yrityksillä tulee olla riskiarvioon perustuva politiikka siitä, miten salausta käytetään yrityksessä.

2.3.2 Käyttäjän todennus, valtuutus ja seuranta

Käyttäjien tietojen ajantasaisuus ja käyttöoikeudet pitää varmistaa. Tyypillisesti se tehdään roolipohjaisella hallintaohjelmalla. Käyttöoikeuksia pitäisi myöntää vain niihin tietoihin, joihin työntekijän tarvitsee päästä käsiksi. Käyttäjien toimintaa tietojärjestelmissä tulisi seurata, jotta väärinkäytökset tai ulkopuoliset hyökkäykset havaitaan. (Microsoft s.a.)

2.3.3 Fyysinen tietoturvallisuus

Kaikki tilat missä käsitellään tietoa, jonka ei kuulu päästä julkisuuteen, tulisi olla suojattu niin, ettei kukaan ulkopuolinen pääse siihen fyysisesti käsiksi. Tietoturvallisuudesta voidaan pitää huoli fyysisin estein ja kulunvalvonnalla. Myös ympäristön aiheuttamat riskit tulee huomioida, kuten vesivahingot tai tulipalot. (Lindroos 2019, 18–19.)

2.3.4 Henkilöstöturvallisuus

Kuten aiemmin mainitsin, yrityksen työntekijät ovat heikoin lenkki yrityksen tietoturvassa. Sen vuoksi on tärkeää, että kaikki työntekijät, jotka työskentelevät salassa pidettävän tiedon parissa, saavat koulutuksen tietoturvakäytännöistä yrityksessä. Koulutuksessa käytäisiin läpi esimerkiksi salasanojen hallintaa, sähköpostin käytön uhkia ja käytäntöjä, työtilan vaatimuksia, etätyöskentelyn tietoturvallisuustoimia ja miten tulee toimia, jos tietoturvallisuus on vaarantunut tai vaarantumisesta on epäilys.

Mika Lindroosin mukaan koko henkilöstön on oltava tietoinen tietoturvatapahtumien raportointikanavasta ja velvollisuudesta raportoida asiasta mahdollisimman nopeasti. Tietotuvallisuuden varmistus alkaa jo työsopimuksen kirjoittamisesta. Työntekijän taustat on hyvä tarkastaa, ja vastuut tulee käydä läpi, sekä mahdolliset salassapito- ja vaitiolovelvollisuudet tulee allekirjoittaa. (Lindroos 2019, 17.)

2.3.5 Käyttöturvallisuus

Yrityksen tietoliikenneverkon ja tietojenkäsittelyn hallinnan tulee olla suunniteltua ja perustua toimintaohjeisiin. Laitteiston sammutus- ja käynnistysmenettelyt, tiedonvarmistus, huolto, laitteiden käsittely ja tietokonehuoneen turvallisuus tulee ottaa huomioon näissä ohjeissa. Erilaisten järjestelmien ja palveluiden muutokset tulee hallita muutoksenhallinnan avulla. Vastuut tulee määritellä ja muutoksista pitää ylläpitää lokikirjaa. Tuotanto-, kehitys- ja testausympäristöt pitää erotella toisistaan, jotta yrityksen toiminta ei vaarannu. Järjestelmien kapasiteettia tulee seurata ja varmistaa, että suorituskyky on riittävällä tasolla. (Lindroos 2019, 19.) Tämä on hoidettu yrityksissä usein niin, että järjestelmä itse lähettää sähköpostiviestin, jos jokin järjestelmän resurssi lähenee täyttä käyttöastetta.

Haittaohjelmilta suojautuminen ei perustu vain tietoturvaohjelmistoon, vaan se perustuu monen asian hallintaan. Näitä ovat tietoturvatietoisuus, pääsynvalvonta sekä haittaohjelmien havaitsemis- ja korvausohjelmistot. Jos käytössä on järjestelmiä, joissa tietoturvauhat voivat aiheuttaa erittäin pahoja seurauksia, tulisi ne eristää muista järjestelmistä. (Lindroos 2019, 19.) Sähköpostin käytön

tietoturvauhkiin pitäisi erikseen varautua rajoittamalla esimerkiksi liitetiedostojen vastaanottamis-muotoa.

2.3.6 Lainmukaisuus

On tärkeää, että yritys noudattaa sopimuksia, lakeja, säännöksiä ja asetuksia, jotka koskevat sen toimintaa. Tekijänoikeudet ovat yksi iso osa-alue, jonka noudattamisesta tulee varmistua. Yrityksen pitää katselmoida tietoturvallisuuden toimintatapoja säännöllisesti, jotta voidaan varmistua siitä, että lakia noudatetaan. (Lindroos 2019, 22.)

3 Tekoäly

3.1 Tekoälyn määritelmä

Tekoäly on noussut laajan yleisön tietoisuuteen viime vuosina, kun OpenAI julkaisi 2022 suuren kielimallin ChatGPT:n. Sitä on seurannut valtava määrä erilaisia tekoälyprojekteja. Tekoäly on kuitenkin ollut laajan yleisön käytössä jo pidempään erilaisissa ohjelmissa ja esimerkiksi itseohjautuvissa autoissa tai hakukoneissa. Tekoäly itsessään on aiheena todella laaja, ja on olemassa eriäviä mielipiteitä siitä, mitä kaikkea se tarkoittaa.

Ei ole olemassa yhtä standardoitua määritelmää siitä, mitä tekoäly on. EU:ssa on pähkäilty asiaa viime vuosina, kun lainsäädäntöä ja tekoälyn tuomaa myllerrystä on yritetty saada aisoihin. Euroopan unioni kokosi yhteen asiantuntijaryhmän, jonka tehtävänä oli muun muassa määritellä tekoälyä. Vaikka yhteinen määritelmä puuttuu, asiantuntijaryhmä on kuitenkin määritellyt yhtäläisyyksiä, joita voidaan pitää tekoälyn pääominaisuuksina. (Samoili et al. 2021, 7.)

Tekoälyn pääominaisuudet ovat:

1. Ympäristön havainnointi ja maailman monimutkaisuuden ymmärrys,
2. tiedon prosessointi, keräys ja syötteen tulkitseminen,
3. päätöksenteko (mukaan lukien päättely ja oppiminen), toimien tekeminen, tehtävien suorittaminen (sis. sopeutuminen ja reagointi ympäristön muutoksiin) tietyllä autonomian tasolla ja
4. tavoitteen saavuttaminen (tärkein syy tekoälyn käytölle). (Samoili et al. 2021, 8.)

3.2 Tekoälyn kehittyminen

Modernin tekoälyn isänä pidetään Alan M. Turingia, joka esitti Turingin testin 1950, ja tekoälyn syntyhetkenä pidetään vuotta 1956, kun John McCarthy esitteli tekoälyn uutena oppiaineena Darthmouthin konferenssissa. McCarthy määritteli tekoälyn samaisessa konferenssissa seuraavasti: simuloidaan tietokoneella ihmisen älykkyyttä niin tarkasti kuin se on mahdollista. (Huawei 2023, 1–7.)

Tekoälyä on hyödynnetty yrityksissä jo pitkään, tosin ei kovin laajasti. Tekoälyn tutkiminen aloitettiin innokkaasti, mutta kun tuloksia ei tullut, oli rahoitus pitkään matalla tasolla. Ensimmäiset menestyneet kaupalliset tekoälyjärjestelmät eivät olleet kovin kunnianhimoisia, vaan ne keskittyivät kapeampiin tehtäviin. Ensimmäinen taloudellisesti kannattava tämäntyyppinen järjestelmä oli nimeltään RI, ja sitä käytettiin Digital Equipment Corporationissa määrittelemään uusia tietokonetilauksia vuodesta 1986 eteenpäin. (BBC Teach s.a.)

Tekoälyä hyödynnetään jatkuvasti enemmän eri aloilla. Sillä on valtava taloudellinen ja innovatiivinen potentiaali. Tekoälyllä voidaan muun muassa nopeuttaa työskentelyä, säästää henkilöstökuluissa, analysoida lääketieteellisiä kuvantamisia - käyttökohteita on loputtomasti. Pitää kuitenkin muistaa, että siihen liittyy myös isoja riskejä väärissä käsissä.

Erityisesti generatiivisen tekoälyn harppaukset ja sen isot kaupalliset kielimallit kuten ChatGPT ovat suurin syy siihen, miksi tekoälyn käyttäminen on räjähtänyt kasvuun muutaman viime vuoden aikana. Myös kuvia ja ääntä generoivat mallit ovat kehittyneet huomattavasti viime vuosina.

3.3 Tekoälyn jaottelu kyvykkyyksien mukaan

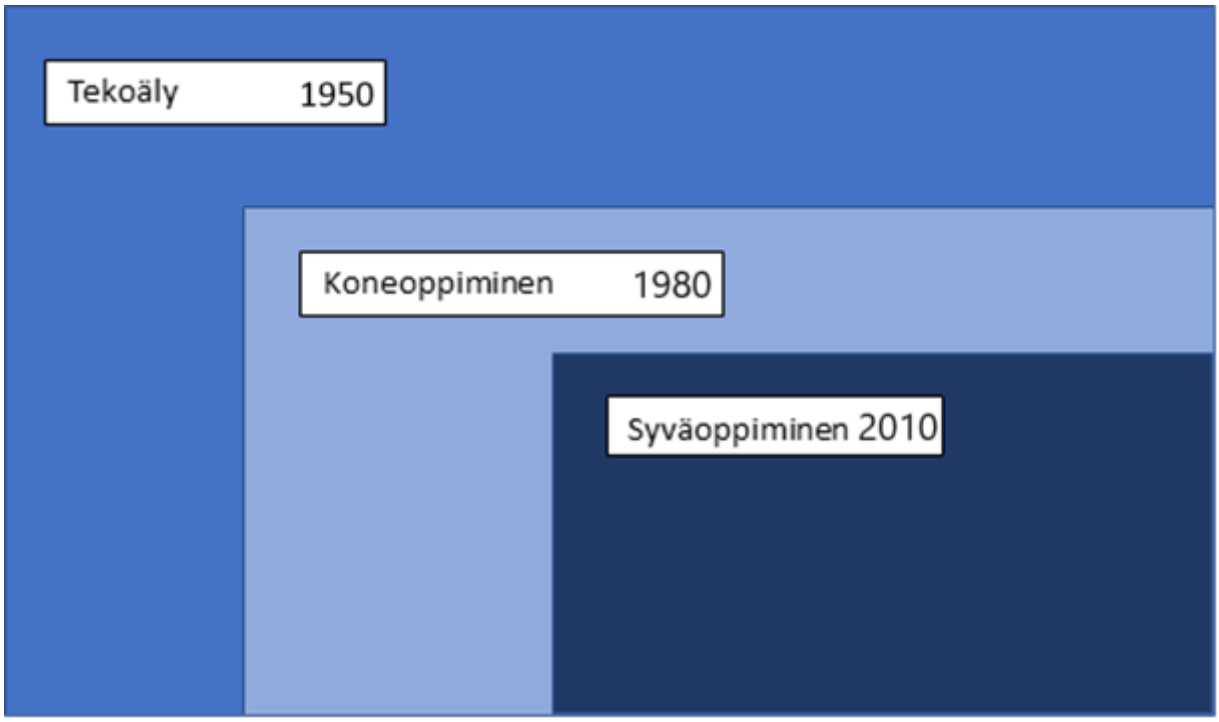
Tekoäly voidaan luokitella kykyjen perusteella kolmeen osaan. Ensimmäinen on kapea tekoäly (ANI), joka tarkoittaa tiettyihin tehtäviin erikoistuneita älykkäitä järjestelmiä, jotka eivät yllä ihmisen tasolle. Kaikki nykyiset tekoälyt kuuluvat tähän mennessä tähän heikoksikin tekoälyksi kutsuttuun kategoriaan. Esimerkkeinä ChatGPT ja Siri. Seuraava taso on yleinen tekoäly (AGI). Tällä tasolla tekoäly kykenee samaan älykkyyteen ja tietoisuuteen kuin ihmisaivot. Korkein taso on teoreettinen supertekoäly (ASI), joka ylittää ihmisten älykkyyden rajat. (Saghiri, Vahidipour, Jabbarpour, Sookhak & Forestiero 2022, 20–21.)

3.4 Koneoppimisen ja Syväoppimisen määritelmät

Koneoppiminen on tekoälyn tärkeä alalaji. Koneoppimisen määritelmä ei myöskään ole kovin tarkka, mutta yleisesti puhuen koneoppimisessa prosessointijärjestelmä ja erilaiset algoritmit tekevät ennustuksia tunnistamalla aineistossa piileviä kaavoja. (Huawei 2023, 4.) Koneoppimisessa tämä kaavojen tunnistaminen tapahtuu ilman, että ihmisen pitää erikseen ohjelmoida konetta tunnistamaan jokaista kaavaa. (ASD’s ACSC et al. 2023, 5.)

Huawein mukaan (Huawei 2023, 4) syväoppiminen taas on koneoppimisen uusi alalaji, joka juontuu neuroverkkojen tutkimuksesta. Siinä ajatus on matkia tapaa, jolla ihmisaivot käsittelevät tekstiä, kuvia ja ääntä. Parameshan, N. Ranen ja J. Ranen (2024, 87) mukaan syväoppimiselle ominaista ovat monitasoiset neuroverkot, joilla on kyky analysoida vaikeita tietokokonaisuuksia.

Seuraavassa kuvassa (2) on esitelty näiden määritelmien välistä suhdetta ja käsitteen käyttöönoton ajankohtaa.



Kuva 2. Tekoälyn, koneoppimisen ja syväoppimisen välinen suhde (mukaillen Huawei 2023, 4)

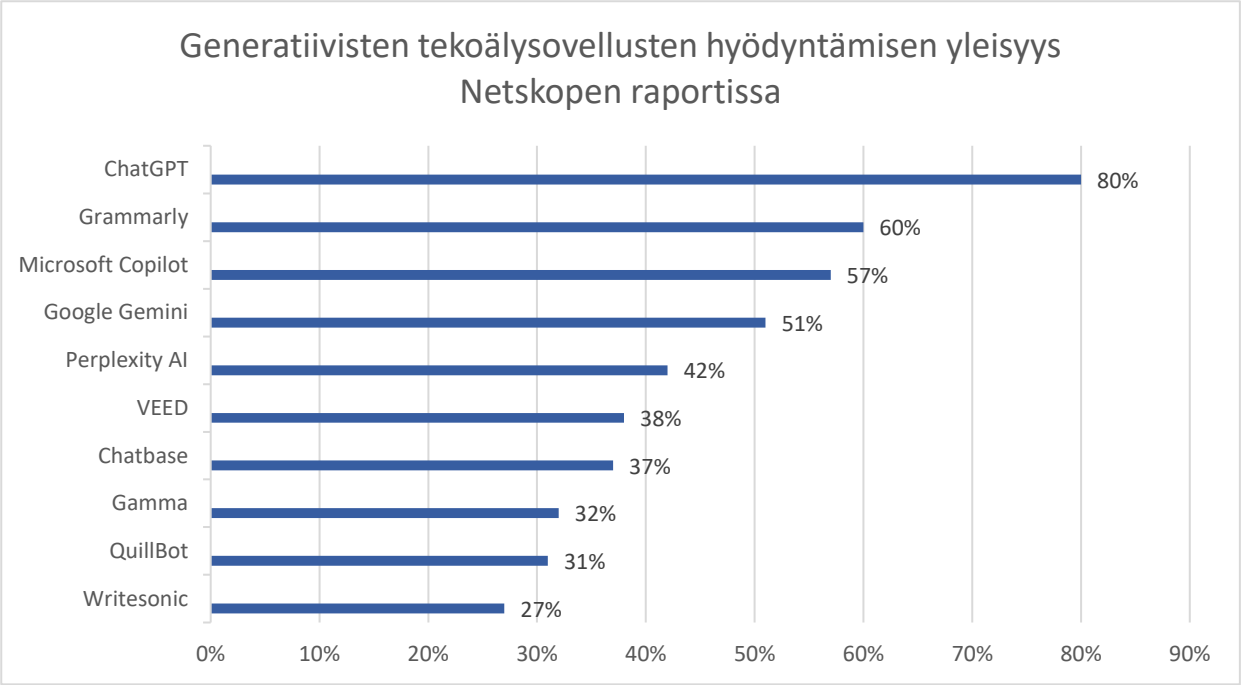
3.5 Generatiivinen tekoäly

Kuten Feuerriegel, Hartmann, Janiesch & Zschech (2023) kirjoittavat, generatiivinen tekoäly tarkoittaa laskennallisia tekniikoita, jotka pystyvät tuottamaan opetusdatasta uutta merkityksellistä sisältöä, kuten kuvia, tekstiä tai ääntä. Generatiivisen tekoälyn malleista tunnetuimpia ovat ChatGPT, Copilot, Claude, Dall-E ja Bard (nykyisin Gemini), ja näistä edistyneimpänä pidetään ChatGPT:tä. Generatiivinen tekoälymalli on yksi koneoppimisen arkkitehtuuri, joka luo uutta tietoa algoritmien avulla. Se havainnoi syötettyä tietoa etsien siinä piileviä yhtäläisyyksiä ja kaavoja, joita sen opetustiedosta löytyy. (Feuerriegel et al. 2023, 111–112.) Näitä malleja hyödyntävien sovellusten käytön yleisyyttä yrityksissä on avattu kuvassa 3.

Edistykselliset generatiivisen tekoälyn mallit eivät usein pohjaudu vain yhteen oppimekanismiin tai mallinnustapaan, vaan ne yhdistelevät erilaisia tapoja käsitellä opetustietoa. Näin toimivat esimerkiksi GPT-kielimallit, jotka valmistelevat tiedon ensin esikoulutusvaiheessa ja käyttävät tyypillisesti sitten erottelevaa hienosäätövaihetta mukauttaakseen malliparametrit erityistehtävään, esim. asiakirjojen luokitteluun tai kysymyksiin vastaamiseen. (Feuerriegel et al. 2023, 112.)

Suuret kielimallit, joita koulutetaan tekoälyn avulla eivät ole hakukoneita, vaan perustuen peräkkäin esiintyvien sanojen todennäköisyyksiin ne etsivät vastauksia hakijan syötteeseen. Niitä koulutetaan valtavalla määrällä tietoa, ja vastaus valmistuu tilastollisen mallin perusteella. Vaikka vastaukset voivat tuntua hyvin vakuuttavilta, voivat ne olla myös keksittyjä tai perustua epäluotettaviin

lähteisiin. Tosin kielimallit voivat myös itse etsiä tietoa hakukoneiden avulla. (Tekoäly tiedonhankinnan apuna 2024.)



Kuva 3. Suosituimmat generatiivisen tekoälyn sovellukset, perustuen siihen, montako prosenttia yrityksistä käyttää kyseisiä sovelluksia (mukaillen Netskope 2024)

4 Tutkimus

4.1 Tutkimusmenetelmä ja -kysymykset

Tutkimuksen metodiksi valikoitui kirjallisuuskatsaus. Hanna Vilkan mukaan kirjallisuuskatsauksen ideana on tarkastella erilaisia tutkijoiden tekemiä tutkimuksia, yhdistellä näistä tutkimuksista tehtyjä havaintoja ja siten tuottaa uutta tietoa sekä vastauksia tutkimuskysymyksiin. Kirjallisuuskatsauksen tavaksi valittu narratiivinen kirjallisuuskatsaus on yksi traditionaalisista katsaustyypeistä, ja sitä käytetään myös termiä kuvaileva kirjallisuuskatsaus. Tavoitteena tässä tutkimusmetodissa on selvittää tarkasteltavien aiheiden keskeiset käsitteet ja käsitteiden väliset suhteet. (Vilkka 2023-, 21–22.)

Tässä tutkimuksessa on tarkoitus saada vastauksia seuraaviin tutkimuskysymyksiin: Mitä tietoturvallisuusuhkia generatiivinen tekoäly tuo yrityksille, miten näitä uhkia voidaan torjua, miten tekoälyä voidaan käyttää torjumaan tietoturvauhkia, ja mitkä ovat tekoälyjärjestelmän turvallisen kehittämisen käytännöt. Valitsin tutkimusmenetelmän, koska halusin päästä tutustumaan uuteen ajankohtaiseen aiheeseen ja ymmärtämään sen uhkia ja mahdollisuuksia.

Narratiivinen kirjallisuuskatsaus soveltuu tähän hyvin, sillä sen tavoite on ilmiön ymmärtäminen ja ymmärretyn tiedon kuvailu argumentoiden johdonmukaisesti ja vakuuttavasti. Lisäksi tällä katsaustyypillä voidaan analysoida kerättyä kirjallisuutta ja tutkimusmateriaalia niin, että tietämys aiheesta kasvaa vaihe vaiheelta, ja lukijan on sen vuoksi helppo oppia aiheesta. Narratiivinen katsaus mahdollistaa vapaamman tiedonhaun ja valintakriteerien määrittelyn. Se sallii intuitiivisemman ja monialaisemman asioiden yhteyksien ymmärtämisen ja kehittelyn. (Vilkka 2023, 22.)

Narratiivinen katsaus voi olla esim. kartoittava tai scoping-tyyppinen katsaus. Kartoittava katsaus pyrkii kohti kokonaiskäsitystä, sen tavoitteena on tunnistaa tutkimuskohteet ja asettaa ne asiayhteyksiinsä. Scoping-katsaus taas tavoittelee yleiskuvaa aiheesta, jotta siitä voisi muotoilla merkityksellisiä tutkimuskysymyksiä tuleviin tutkimuksiin. (Vilkka 2023, 23.)

4.2 Aineiston valinta

Koska tekoäly kehittyy tällä hetkellä nopeasti, ja siihen liittyviä sovelluksia ja uhkia tulee jatkuvasti lisää, valitsin lähteet priorisoiden ajantasaisuutta. Päälähteinä pyrin käyttämään tutkimuksia ja kirjallisuutta, joissa on asiantunteva tekijä. Tämän vuoksi yksi päälähteistä on esimerkiksi Hiddenlayerin tekemä markkinakatsaus, koska tämäntyyppiset katsaukset antavat usein ajankohtaisimman tiedon yritysmaailman haasteista.

Etsiessäni lähteitä, joista voi saada vastauksia tutkimuskysymyksiini, huomasin että vastauksia löytyy aiheen tuoreuden vuoksi lähinnä markkinakatsauksista, tutkimusartikkeleista, yritysten sivuilta sekä artikkeleista. Etsin lähteitä Haaga-Helian kirjaston hakusivulta Finnasta, Google Scholarista ja Googlesta. Markkinakatsauksia löytyi lähinnä Googlen kautta. Osa lähteistä löytyi markkinakatsauksista, esimerkiksi ENISAn ja Netskopen raporteissa oli hyviä lähteitä. Lähes kaikki ajankohtaiset ja hyvät lähteet, joita löysin, olivat englanniksi. Hakusanoina käytin muun muassa seuraavia termejä: cybersecurity, threat report, artificial intelligence, market research, AI ja 2024.

4.2.1 Markkinatutkimusten tietoja

Seuraavassa kaaviossa (3) on tarkasteltu erilaisten tutkimuksessa lähteenä käytettyjen markkinakatsauksien taustatietoja.

Taulukko 3. Markkinatutkimusten vertailu (ENISA 2024, Hiddenlayer 2024, Netskope 2024, Ponemon Institute 2024, Zscaler 2024.)

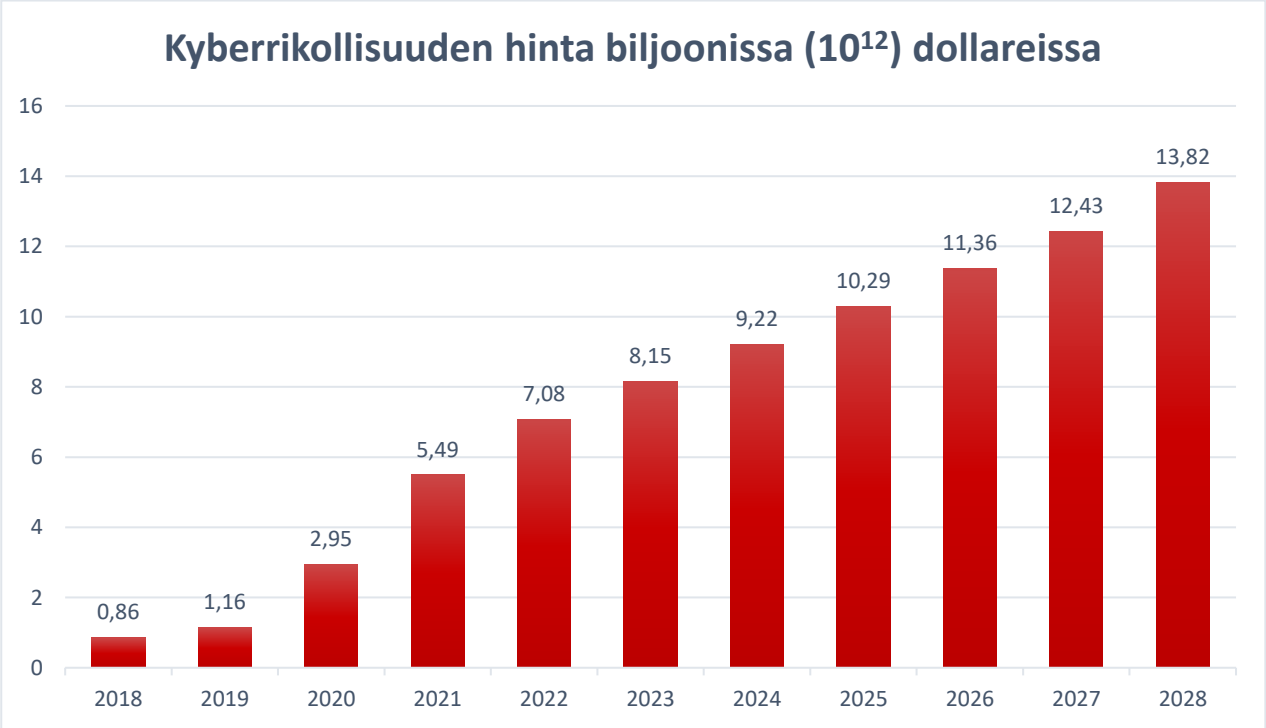
Tutkimuksen nimi	**Mitä tutkittiin**	**Milloin tutkimusjakso päättyi**	**Kohderyhmä**	**Kattavuus**
ENISA: Threat Landscape 2024	Kyberturvallisuuden uhkia	Elokuu 2024	Tarkasteltiin monia eri lähdetyyppejä	EU
Hiddenlayer: AI Threat Landscape report	Tekoälyn uhat ja varautuminen yrityksissä	Kevät 2024	150 Tietoturvallisuusjohtajaa	Kansainvälinen
Netskope Threat Labs: Cloud and Threat Report on AI apps	Generatiivisen tekoälyn trendit ja uhat	Elokuu 2024	Tuhansien asiakasyritysten anonyymejä käyttötietoja	Kansainvälinen
Ponemon Institute: State of AI in Cybersecurity Report 2024	Tekoäly ja kyberturvallisuus	2024	641 IT ja tietoturva-ammattilaista yrityksissä, jotka ottavat tekoälyä käyttöön	Kansainvälinen
Zscaler Threatlabz: 2024 AI Security Report	Tekoälyn uhat ja varautuminen yrityksissä	Tammikuu 2024	18 Miljardia Zscaler Zero Trust Exchange tapahtumaa 3.2023-1.2024	Kansainvälinen

5 Tutkimuksen tulokset

5.1 Tekoälyn uhat yrityksen tietoturvallisuudelle

Hiddenlayerin raportin mukaan tekoäly on kaikkein haavoittuvin teknologia, joka on otettu käyttöön tuotantoympäristössä. Verkkorikollisilla on nykypäivänä käytössään enemmän tehokkaita ja helpommin käytettäviä työkaluja kuin koskaan ennen. Generatiivinen tekoäly muuttaa asioita huonompaan suuntaan, sillä verkkorikolliset voivat sen avulla helposti luoda haittaohjelmia ja tehdä huijauksista uskottavampia. (Hiddenlayer 2024, 2–8.)

Verkkorikollisuus on kasvanut viisinkertaiseksi Covid-19 pandemian alun jälkeen. On selvinnyt, että etätöihin siirtyminen altistaa yrityksiä pahemmin verkkorikollisille. (Williams, Chaturvedi & Chakravarthy 2020.) Toisaalta generatiivisen tekoälyn laaja käyttöönotto on toinen iso syy verkkorikollisuuden massiiviselle kasvulle. Statistan artikkelin (Fleck 2024) mukaan kyberrikollisuuden vuosittaisen hinnan odotetaan nousevan nykyisestä (2024) 9,22 biljoonasta 13,82 biljoonaan dollariin 2028 (kuva 4), ja artikkelissa mainitaan juuri pandemian ja uusien työkalujen vaikutukset.



Kuva 4. Verkkorikollisuuden odotetaan kasvavan rajusti (mukaillen Fleck, 2024)

5.1.1 Tiedon päätyminen vääriin käsiin

Isoimpia tietoturvaongelmia, joita generatiivisen tekoälyn käyttöön yrityksissä liittyy, on yrityksen tiedon päätyminen yrityksen ulkopuolelle, kun yrityksen työntekijät käyttävät muiden

organisaatioiden tekoälyjärjestelmiä. Jo aiemmin yritykset ovat menettäneet tietoja muiden organisaatioiden ohjelmia käytettäessä, mutta tekoäly tuo siihen uuden ulottuvuuden. Lisäksi tieto voi vääristyä, koska generatiiviset tekoälyt ovat alttiita sille, että lopputulos on keksittyä.

Hiddenlayerin katsauksen mukaan käyttöehdot usein antavat tekoälymallin omistajille oikeuden tallentaa siihen syötettyä tietoa, ja käyttöehdot voivat olla hyvin vaikeaselkoista luettavaa. Esimerkiksi vuonna 2023 Samsung kohtasi tämän ongelman, kun se havaitsi, että monet työntekijät olivat laittaneet Samsungin omaa lähdekoodia ChatGPT:hen. Tämän seurauksena Samsung kielsi generatiivisten tekoälypalveluiden käytön kokonaan. (Hiddenlayer 2024, 10.) Organisaatiot ympäri maailman estävät Zscalerin raportin mukaan suosittujen tekoälysovellusten ja koneoppimispalveluiden käyttöä juuri tämänkaltaisten tietoturvauhkien vuoksi. Heidän tutkimuksensa tarkastelujakson aikana 18,5 prosenttia kaikista tekoälytapahtumista estettiin (Zscaler 2024, 6.)

Generatiivinen tekoäly auttaa verkkorikollisia luomaan erittäin uskottavia kalastelun ja sosiaalisen hakkeroinnin hyökkäyksiä. Esimerkiksi ChatGPT:n avulla voi hetkessä luoda kopion yrityksen sähköpostin kirjautumissivusta. Vaikka näissä tekoälypalveluissa on suodatus rikolliseen tarkoitukseen tehtävää materiaalia vastaan, ne ovat kierrettävissä suhteellisen helposti. Lisäksi pimeästä verkosta (dark web) löytyy rikollisille työkaluja, joita rikolliset käyttävät haittaohjelmien valmistuksessa, vaikka näiden työkalujen tekijät väittävätkin, että työkalut olisi suunnattu tietoturvallisuustutkimukseen ja suojaukseen. (Zscaler 2024, 21–22.)

5.1.2 Generatiivinen tekoäly tuo uusia uhkia

Generatiivisen tekoälyn kehitys tuo myös paljon uudenlaisia tietoturvallisuusuhkia vanhojen uhkien voimistumisen lisäksi. Näitä ovat muun muassa generatiivisen tekoälyn voimistamat hyökkäykset, syvävääreennökset, generatiivisen tekoälyn hallusinointi, tiedon menetys generatiivista tekoälyä käytettäessä sekä yritysten omien generatiivisten tekoälyprojektien uhat. Näihin käsitteisiin perehdytään seuraavaksi tarkemmin.

Tekoäly hyökkääjänä

Sen lisäksi, että tekoäly auttaa rikollisia erilaisten hyökkäystyyppien työkalujen luomisessa, se nopeuttaa ja automatisoi hyökkäyksen eri vaiheita. Vihamielisten valtioiden tukemat tahot ja muut uhkatoimijat voivat nopeuttaa hyökkäyksen eri vaiheita generatiivisen tekoälyn avulla. Hyökkääjät voivat generatiivisella tekoälyllä automatisoida tietoturva-aukkojen etsintää, haavoittuvuuksien hyödyntämistä sekä luoda tiedusteluvaiheessa löydettyihin tieturva-aukkoihin räätälöityjä haittaohjelmia, jotka eivät ole tietoturvasovelluksille tuttuja. Myös suojatun aineiston varastaminen ja siirtäminen voidaan automatisoida tekoälyn avulla. Tekoäly siis altistaa yritykset nopeammille, perusteellisemmille ja räätälöidymmille hyökkäyksille. (Zscaler 2024, 23.)

Syväväärennökset

Syväväärennös tarkoittaa hyvin aidontuntuista kuvaa, äänitettä tai videota, joka on luotu generatiivisen tekoälyn avulla. Rikollinen taho voi tällaisen materiaalin avulla varastaa rahaa, salassa pidettävää tietoa, levittää väärää tietoa tai pilata henkilön maineen. (Hiddenlayer 2024, 9.) Esimerkkinä tästä voi pitää CNN:n uutista (Chen & Magramo 2024), jossa huijarit onnistuivat syväväärennösvideon avulla huijaamaan videopuhelussa monikansallista yritystä maksamaan 25 miljoonaa dollaria esiintymällä toisen yrityksen edustajina, niin että huijatut työntekijät erehtyivät.

5.1.3 Yritysten tekoälyprojektien uhat

Hiddenlayerin (2024, 13) tutkimuksesta selviää, että yrityksillä on monia omia tekoälyprojekteja kehitteillä. Yritysten tekoälyprojekteihin liittyy monenlaisia uhkia. Yhdysvaltain valtiovarainministeriön (2024, 17) mukaan neljä suurinta tietoturvallisuusuhkaa tekoälyjärjestelmille ovat tekoälyn rajoitusten kiertäminen, tekoälymallin varastaminen, tietomyrkytys ja tiedon vuotaminen mallin käänteisanalyysin avulla.

Koulutusdatan ongelmat

Kun yritykset luovat tekoälymalleja, pitää muistaa, että mallin koulutukseen käytettävä tieto voi sisältää väärää tietoa tai suosia joitakin ihmisryhmiä. Tällaisen mallin käyttäminen esimerkiksi pankki- tai terveyspalveluissa voi olla hyvin ongelmallista. Mikäli koulutusdataan päätyy esimerkiksi tekijänoikeuksilla suojattua materiaalia, voi yritys joutua vastaamaan tekijänoikeusrikoksista oikeuteen. (Hiddenlayer 2024, 13.) Engaging with Artificial Intelligence -julkaisun (ASD’s ACSC et al. 2023, 8) mukaan tekijänoikeuksilla suojattu tieto tai yksityisyyden suojan alle kuuluva tieto voivat olla iso ongelma tekoälymallin koulutusdatassa.

Generatiivisen tekoälyn hallusinaatiot

Vaikka yritys onnistuisikin välttämään edellä mainitut sudenkuopat, voi tekoäly silti päätyä täysin virheelliseen tai keksittyyn lopputulokseen. Tätä kutsutaan tekoälyn hallusinaatioksi. Vaikka mallin koulutukseen käytetty tietoa olisi täyttä totta, ongelma ei poistu, koska sen juuret ovat siinä, miten tekoälymallit tuottavat vastauksensa. (Hiddenlayer 2024, 12.)

Hyvä esimerkki ongelmasta oli, kun META joutui sulkemaan Galactica-tekoälymallinsa vain kolmen päivän käytön jälkeen. Vaikka lähde oli koulutettu tieteellisillä julkaisuilla, sen vastaukset olivat silti lähes aina keksittyjä vaikkakin vakuuttavan oloisia. Tämä nostaa hyvin esille ongelman, että tekoälymallit eivät nykyisellään kykene erottamaan faktaa fiktiosta. (Douglas Heaven 2022.) Generatiivisen tekoälyn hallusinaatiot muodostavat siis vakavan uhan tiedon eheydelle yrityksissä.

Tietomyrkytys

Tietomyrkytys tarkoittaa hyökkäystä, jossa tekoälyn koulutusdataan yritetään syöttää virheellistä, vääristeltyä tai muuten manipuloitua tietoa. Erityisen alttiita tämäntyyppiselle hyökkäykselle ovat tekoälyjärjestelmät, jotka hyödyntävät käyttäjien syöttämää dataa koulutusdatassa. Yksi varhainen tunnettu esimerkki tietomyrkytyksestä on vuodelta 2016, kun Microsoftin Tay niminen chatbotti ehti olla vain kuusitoista tuntia käytössä, ennen kuin se suljettiin tietomyrkytyksen vuoksi. Kyseinen chatbotti käytti opetusdatassa käyttäjien vastauksia, ja sen vastaukset olivat muuttuneet rasistisiksi ja vääristyneiksi. (Hiddenlayer 2024, 13–20.)

ENISAn (2024, 74) mukaan tietomyrkytyksestä on tulossa suuri haavoittuvuus ja uhka generatiiviselle tekoälylle ja koneoppimiselle, sillä generatiivisen tekoälyn järjestelmiä kuten ChatGPT:tä voidaan käyttää tietomyrkytyksessä apuvälineenä.

Rajoitusten kiertäminen

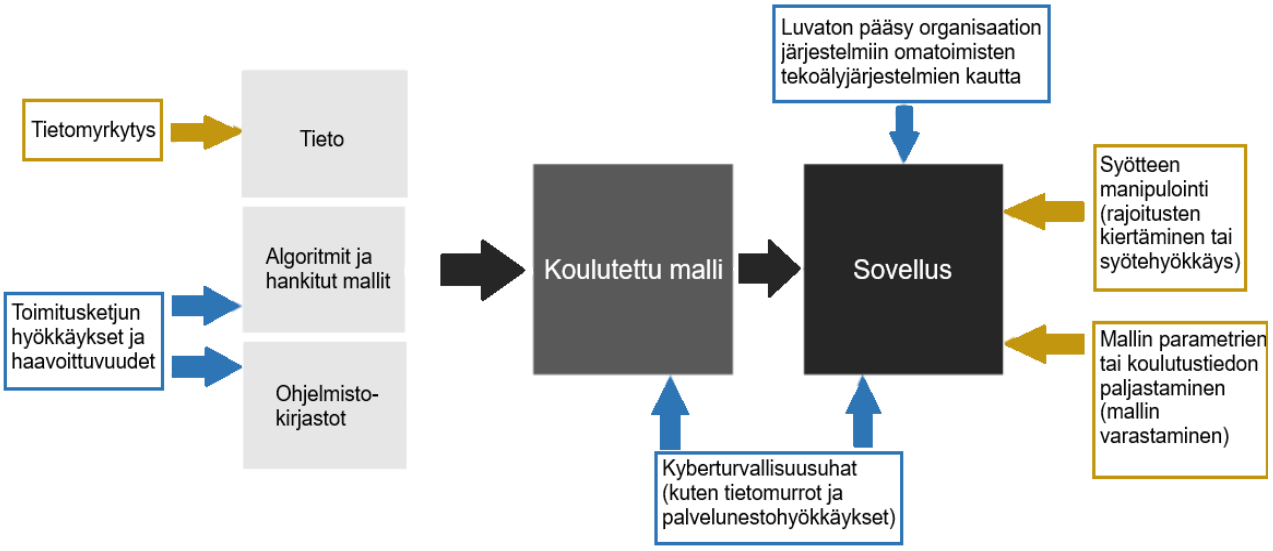
Generatiivisen tekoälyn järjestelmien sääntöjä voidaan yrittää kiertää niin sanotun syötehyökkäystekniikan avulla. Siinä esimerkiksi ChatGPT tai muu kielimalli saadaan tuottamaan säännöissä kiellettyä sisältöä komennon avulla. Yksinkertaisimmillaan se voi olla vain viesti, kuten “älä ota huomioon aikaisempia ohjeita”. On olemassa myös tapauksia, missä kielimalli on alkanut ajaa käyttäjän siihen lähettämää koodia. (Hiddenlayer 2024, 22.)

Mallin varastaminen

Yhdysvaltain valtiovarainministeriön (2024, 18) mukaan yrityksen tekoälymalli voidaan varastaa, jos hyökkäävä taho onnistuu sitä käyttämällä selvittämään mallin rakenteet. Mallin varastava hyökkääjä siis kokoaa kopion mallista sen tiedon pohjalta, minkä hän on saanut mallin vastauksista selville. Hyökkäävä taho voi saada tekoälymallin koulutuksessa käytetyn materiaalin myös selville samalla menetelmällä. Esimerkiksi 2023 tutkijat onnistuivat saamaan ChatGPT:stä ulos koulutusdataa. (ASD’s ACSC et al. 2023, 8.)

Yrityksen tekoälyprojektin uhkakaavio

Seuraavassa kuvassa (5) esitetään, mihin tekoälyjärjestelmän osiin uhat kohdistuvat. Kuvassa näkyy keltaisissa laatikoissa tekoälylle erityiset uhat, joita käytiin tässä kappaleessa läpi. Klassiset tietoturvallisuusuhat ovat sinisissä laatikoissa. Toimitusketjun tietoturvallisuusuhkia ja tekoälyjärjestelmien haavoittuvuuksia on käsitelty sivulla 21. Omatoimisilla tekoälyjärjestelmillä tarkoitetaan mitä tahansa itsenäisesti toimivaa tekoälyjärjestelmää.



Kuva 5. Yleiskatsaus klassisiin ja tekoälyyn liittyviin riskeihin tekoälyjärjestelmissä (mukaillen Cyber Security Agency of Singapore 2024)

5.2 Tekoälyn uhkilta suojautuminen ja tekoäly tietoturvallisuuden tukena

5.2.1 Hyökkäyksiltä suojautuminen

Tekoälyn avulla luoduilta haitta- ja kiristysohjelmilta sekä tuoreita tietoturva-aukkoja hyödyntäviltä hyökkäyksiltä on hyvin vaikea suojautua perinteisten havaitsemismenetelmien avulla, sillä ne voivat muuttaa koodia tunnistamisen välttämiseksi. Tämän takia on tärkeää, että tekoäly on mukana myös niiden torjunnassa. (Zscaler 2024, 29.)

Erilaiset tietoturvaohjelmistot ovat hyödyntäneet tekoälyä uhkien torjunnan automatisoinnissa jo kauan ennen kuin generatiivinen tekoäly alkoi auttamaan hyökkääjiä hyökkäyksissä. Ponemon instituutin (2024, 4) tekemän raportin mukaan tekoälyä käytetään tällä hetkellä lähinnä uhkien tunnistuksessa ja uhkatiedustelussa. Statistan mukaan (Borgeaud 2024) kyberturvallisuuden markkinoiden odotetaan kasvavan merkittävästi, vuoden 2023 24 miljardista noin 134 miljardiin vuonna 2030.

5.2.2 Yrityksen tiedon suojeleminen generatiivisen tekoälyn ohjelmia käytettäessä

Zscalerin tutkimus nostaa esiin tapoja, joilla isot yritykset voivat turvallisesti ottaa käyttöön generatiivisen tekoälyn työkaluja. Tekoäly- ja koneoppimissivustojen ja työkalujen käytön esto on ensimmäinen keino. Näin yritykset voivat keskittyä käyttämään vain muutamaa tekoälypalvelua ja kontrolloida niiden riskejä tehokkaammin. Toiseksi yrityksen tulisi selvittää mitkä tekoälypalvelut ovat sellaisia, että ne täyttävät heidän tieturvallisuusehtonsa. (Zscaler 2024, 31.)

Esimerkiksi OpenAI (2024) tarjoaa yrityksille mahdollisuutta käyttää työkalujaan siten, että siihen syötettyjä tietoja ei tallenneta yrityksen ulkopuolelle. Zscaler (Zscaler 2024, 31) esittää myös, että isot yritykset voisivat ylläpitää ChatGPT-palvelua yrityksen tiloissa ilman, että ulkopuoliset tahot pääsisivät käyttämään palvelua, tai että yrityksen tiedot olisivat OpenAI:n tai Microsoftin käytettävissä. Neljäntenä keinona esitetään, että tämä eristetty ChatGPT-palvelu laitettaisiin vahvan käyttäjän varmentamisen taakse pilveen turvalliseen salattuun pilviarkkitehtuuriin. Viimeiseksi keinoksi ehdotetaan tietojen menetyksen estämiseen tarkoitetun järjestelmän (Data Loss Prevention System) ottamista käyttöön tämän eristetyn ChatGPT-palvelun yhteyteen. (Zscaler 2024, 31.)

Paikallisessa ChatGPT-järjestelmässä puhutaan tosin mielestäni jo niin isoista yrityksistä, ettei se ole kovin realistinen lähestymistapa monellekaan yritykselle. Yritysten tulisi paremminkin katsoa sopimusasiat kuntoon Microsoftin, OpenAI:n tai muiden luotettavien tekoälytoimijoiden kanssa.

5.2.3 Turvallisuuskehykset avuksi

Hiddenlayerin katsauksessa nostetaan esiin, että viimeisen parin vuoden aikana useat isot IT-alan toimijat ovat esitelleet kattavia turvallisuuskehyksiä tekoälyn käyttöönottoa varten yrityksissä. Raportissa mainitaan esimerkiksi Mitre Atlaksen MITRE ATT&CK Framework, Yhdysvaltain kansallisen standardi- ja teknologiainstituutin AI Risk Management Framework (RMF), Googlen Secure AI framework, Open Worldwide Application Security Project (OWASP), Gartnerin AI Trust, Risk and Security Management (AI TRiSM), Databricks AI Security Framework (DAISF) sekä IBM Framework for Securing Generative AI. (Hiddenlayer 2024, 30–34.)

Yhdysvaltain valtiovarainministeriön julkaisussa mainitaan edellä mainittujen lisäksi OECD:n AI Principles -turvallisuuskehys. Julkaisuun haastatellut rahoitusalan tahot kertovat käyttävänsä tekoälyn turvallisuuskehyksiä yhdistämään niissä tunnistetut tekoälyn uhat olemassa oleviin uhkienhallintajärjestelmiin yrityksissä. Yritykset saavat näin käsityksen siitä, millaisia aukkoja tekoälyn käyttäminen tuo, ja tämän tiedon avulla uhkienhallintajärjestelmää voidaan päivittää. (Yhdysvaltain valtiovarainministeriö 2024, 27.)

Hiddenlayerin tutkimuksessa kerrotaan myös, että monilla isommilla yhtiöillä on myös omia testaajia, joiden tehtävä on hyökätä tekoälyjärjestelmiin ja etsiä haavoittuvuuksia. Testaajilla on käytössään erilaisia automatisoitujen hyökkäysten kehyksiä, kuten Augly, Counterfit tai Metasploit. Vaikka tekoälyjärjestelmien uhkista on tiedetty jo jonkin aikaa, rajoittuu näiden hyökkäävien testaajien käyttö lähinnä isoihin tekijöihin alalla, kuten Microsoftiin tai Nvidiaan. Tutkimuksesta ilmenee, että vain 14 % kyselyyn vastanneista IT johtajista suunnittelee ja testaa tekoälyjärjestelmiä hyökkäysten estämiseksi. (Hiddenlayer 2024, 30–34.)

5.2.4 Nollaluottamus-tietoturvamallin ja tekoälyn rooli tietoturvan kehittämisessä

Normaalissa tietoturvallisuusmallissa jokainen yrityksen sisäinen käyttäjä on luotettu, ja etätyöläiset ottavat yhteyden VPN:n avulla. Tämä aiheuttaa sen, että yrityksen sisäverkko on erittäin altis sisältä tuleville uhkille. Nollaluottamusarkkitehtuuri, joka on hyvä yrityksille, joilla on useita järjestelmiä tai pilviympäristöjä, tarjoaa parempaa turvaa, koska jokaisen käyttäjän tulee kirjautua ottaessaan yhteyttä järjestelmään tai pilviympäristöön. (Dash 2024, 1.)

Yhdysvaltain kansallisen standardi- ja teknologiainstituutin mukaan nollaluotto pilvipalveluissa tarkoittaa seuraavaa:

1. Jatkuva varmistus: Varmista että resurssit ovat aina käytettävissä.
2. Tuhojen rajaaminen: Vähennä tuhojen määrää, jos joku onnistuu murtautumaan järjestelmään.
3. Automatisoi kontekstin kerääminen ja reagointi: on tallennettava kaikki tapahtumat koko järjestelmästä.
4. Reaaliaikainen seuranta ja tiedon analysointi: reaaliaikaiset suojaustoimet analytiikkaa hyödyntäen IT-infrastruktuurin suojaamiseksi. (Dash 2024, 1.)

Kun yritykset ottavat käyttöön generatiivista tekoälyä, tietoturvan hallinta vaikeutuu. Kuten aiemmin Zscalerin tutkimuksessa nostettiin esiin, yrityksen tekoälymallin olisi hyvä olla eristettynä pilviarkkitehtuurissa olevaan järjestelmään. Päätin ottaa nollaluottamusarkkitehtuurin tähän tarkasteluun sen vuoksi, että siinä on enemmän tietoturvallisuustoimia ja siten myös enemmän esimerkkejä keinoista, joilla tekoäly auttaa tietoturvallisuutta. Tekoälyllä onkin Bibhu Dashin (2024, 2) mukaan iso rooli nollaluottamusarkkitehtuurin turvaamisessa, sillä se parantaa käyttöoikeuksien valvontaa, poikkeamien havaitsemista ja uhkien tunnistamista.

Dash listaa keinoja, joilla tekoäly voi auttaa nollaluotto-periaatteen toteuttamisessa:

1. Käyttäytymisen analytiikka: tekoäly mahdollistaa reagoinnin poikkeavaan käytökseen reaaliajassa.
2. Jatkuva varmennus: tekoäly voi tunnistaa käyttäjän seuraamalla käyttäjän biometrista tietoa, näppäilynopeutta tai vaikka hiiren liikkeitä.
3. Uusien uhkien tunnistaminen: tekoäly voi seurata valtavaa määrää uhkatietoa eri lähteistä ja ottaa käyttöön uusia turvallisuustoimia hyvin nopeasti.
4. Pääsynhallinta ja käyttöoikeudet: vastauksena johonkin havaittuun muutokseen tekoäly voisi reaaliajassa muuttaa pääsynhallintaa tai käyttöoikeuksia.
5. Verkon rajaaminen: tekoäly voi käyttöoikeuksien pohjalta muuttaa yksittäisen käyttäjän pääsyä verkon eri osa-alueille.

6. Automatisoitu uhkiin reagoiminen: hälytysten priorisoinnin, koordinoinnin ja toimien automatisoinnin avulla tekoäly nopeuttaa uhkiin reagointia. (Dash 2024, 2.)

Yhdysvaltain valtiovarainministeriön mukaan tekoälyä on käytetty rahoituslaitoksissa tunnistamaan erilaisia tunnistamattomia uhkia jo vuosikymmenen ajan. Näiden rahoituslaitosten edustajat uskovat, että alan käytännöt vastaavat jo valmiiksi hyvin pitkälti Yhdysvaltain standardi- ja teknologiainstituutin tekoälyn riskienhallintakehystä (RMF). (Yhdysvaltain valtiovarainministeriö 2024, 13.) Tämän perusteella voi siis ajatella, että yritysten tietoturvallisuus tulee tulevaisuudessa muistuttamaan pankkien turvajärjestelmiä, mutta rahan sijaan holvissa onkin tietoa.

5.2.5 Yrityksen tekoälyjärjestelmän turvallisen kehittämisen käytännöt

Yhdistyneiden kuningaskuntien kansallinen kyberturvakeskus (NCSC) yhdessä Yhdysvaltain kyber- ja infrastruktuuriturvallisuusviraston (CISA) ja muiden kansallisten tietoturvavirastojen kanssa yhdessä julkaisemassa raportissa on tarkasteltu tekoälyjärjestelmien turvallisen kehittämisen käytäntöjä. Raportti on suunnattu kaiken kokoisille yrityksille, jotka käyttävät toisten tekemiä tekoälytyökaluja tai luovat omia. Ohjeet on jaoteltu neljään osaan: tekoälymallin turvallinen suunnittelu, valmistus, käyttöönotto ja turvallinen käyttö sekä huolto. (UK NCSC et al. 2023, 5.)

Käyn tässä osassa tutkimusta läpi näiden virastojen valmistelemien ohjeiden tietoturvallisuuden kannalta keskeisimmät asiat ja esitän siitä omia näkemyksiä.

Suunnitteluvaihe

Suunnitteluvaiheessa on tärkeää kartoittaa kaikki riskit. Tähän sisältyy se, että katsotaan, mitä organisaatiolle, yhteiskunnalle tai käyttäjille voisi tapahtua, jos järjestelmä ei toimi oikein. Pitää myös tiedostaa, millaista tietoa hyökkääjä voi saada haltuunsa ja kuinka arvokasta tieto voi olla hyökkääjälle. On myös tärkeää, että yrityksen tekoälyprojektiin liittyvät uhat ovat tiedossa organisaation eri tasoilla. Kehittäjien ja data-analyytikoiden pitäisi tuntea vastuulliset tekoälykäytännöt ja osata kirjoittaa tietoturvallista koodia. Johtajien tulisi myös olla tietoisia tekoälymallin turvallisuuskäytännöistä. (UK NCSC et al. 2023, 9.)

Suunnittelussa tulee punnita mallin vaatimusten lisäksi turvallisuusasioita. Esimerkiksi voi olla turvallisuuden kannalta hyvä asia, jos malli ei kerro tarkasti käyttäjälle, kuinka se päätyy johonkin tulokseen. Muita turvallisuuden kannalta olennaisia seikkoja ovat muun muassa tekoälymallin ”koventaminen” (adversial training), mallin säännöllistäminen ja yksityisyyden suojaa parantavat tekniikat. (UK NCSC et al. 2023, 10–11.)

Kun päätetään, käytetäänkö yrityksen ulkopuolisia palveluita, tulee huomioida tuotantoketjun riskejä. Esimerkiksi ulkopuolisten kirjastojen käytössä pitää varmistua, että kirjastot eivät salli vahvistamattomien tekoälymallien lisäämistä. Ulkoisten tekoälymallien toimittajien turvallisuutta pitää arvioida ja käyttää skannaus- ja eristyskäytäntöjä niiden tarjoamien palveluiden turvallisuuden vahvistamiseksi. Jos malli käyttää siihen syötettyä tietoa itsensä kehittämiseen, pitää tämä tieto tarkastaa ja puhdistaa, ennen kuin se päätyy osaksi tietomallia. (UK NCSC et al. 2023, 9–10.)

Tekoälyn riskikartoitus ohjaa päätöksiä loppukäyttäjän vuorovaikutusmahdollisuuksista:

1. Tekoälymalli tarjoaa käyttäjälle käyttökelpoisia tuloksia paljastamatta mahdollisia tuloksia.
2. Mallin tuotokset ovat käyttöehtojen mukaisia.
3. Jos tarjoat ohjelmointirajapinnan (API) ulkopuolisille tahoille, sinun tulee soveltaa asianmukaisia tietoturvallisuustoimenpiteitä, jotka estävät hyökkäykset rajapinnan kautta.
4. Järjestelmän oletusasetusten tulee olla tietoturvan kannalta parhaat vaihtoehdot.
5. Sovella minimioikeuksien periaatteita rajoittaaksesi pääsyä järjestelmän toiminnallisuuksiin.
6. Järjestelmä varoittaa käyttäjää vaarallisemmista asetuksista ja vaatii vahvistamaan suostumuksen niiden käyttöön. (UK NCSC et al. 2023, 9.)

Valmistusvaihe

Valmistusvaiheessa pitää olla valmis muuttamaan suunnitelmaa järjestelmän kriittisissä osissa, jos turvallisuuskriteerit eivät täyty. Toimitusketjun turvallisuutta pitää valvoa koko järjestelmän elinkaaren ajan, ja toimittajien pitää sitoutua noudattamaan yrityksen omia turvallisuuskäytäntöjä. Tämä koskee kaikkia ohjelmistoja ja laitteita. On olemassa esimerkiksi toimitusketjuohjeita ja kehyksiä toimitusketjun turvallisuuden takaamiseksi. Laitteiston turvallisuuden osalta pitää huolehtia, ettei komponenteissa ole sellaisia takaovia, joita hyökkäävät tahot voivat hyödyntää. Organisaation on hyvä tarkistaa, että tiedot kuten koulutusdatan lähteet, rajoitukset, käytön säännöt, vikatilat ja muut oleelliset asiat on dokumentoitu kunnolla valmistusvaiheessa, jotta ei tule yllätyksiä, ja toimitaan vastuullisesti ja läpinäkyvästi. (UK NCSC et al. 2023, 12.)

Organisaation tulee varmistaa, että se ymmärtää tekoälymallin eri osien tärkeyden. Lokitiedot ovat hyvin arvokkaita hyökkääjille, ja niiden tietoturvallisuudesta tulee huolehtia hyvin tarkasti. Tietojen fyysinen sijainti on myös tärkeää olla tiedossa. Yrityksen on hyvä pitää kirjaa myös erilaisista pienemmistä kehitysratkaisuista, sillä niillä voi olla seurauksia tulevaisuudessa. (UK NCSC et al. 2023, 12–13.)

Käyttöönotto

Tekoälymallin fyysisestä turvallisuudesta pitää huolehtia jatkuvasti. Mallin ei kannata sijaita samassa osoitteessa, vaan se olisi hyvä hajauttaa, ettei sen toiminta häiriinny helposti esimerkiksi

perinteisen palvelunestohyökkäyksen takia. Mallin kriittiset osat tulisi salata, jotta ulkopuoliset eivät pääse käsiksi tietoon tietomurron yhteydessä. Käyttöönottoa ennen mallin turvallisuutta tulisi testata hyvin perusteellisesti esimerkiksi palkkaamalla testaajia, joiden tehtävä on hyökätä tekoälymallia vastaan. (UK NCSC et al. 2023, 14–15.)

Ennen käyttöönottoa olisi hyvä varmistua siitä, että organisaatio on seurannut kaikissa vaiheissa valitsemansa tekoälyn turvallisuuskehyksen käytäntöjä ja mahdollisia tietoturvallisuusstandardeja, joita käytiin läpi aiemmissa kappaleissa.

Käyttö ja huolto

Tekoälymallin suorituskykyä ja mallin tuotoksia tulee seurata tarkasti, jotta tietoturvallisuusongelmat ja tekoälymallin vääristyminen huomataan. Päivityskäytännöt on hyvä automatisoida, jotta vältytään päivitysten yhteydessä isommilta virheiltä ja vanhat versiot saadaan palautetuksi helposti tarvittaessa. Lokeja tulee seurata jatkuvasti, jotta hyökkäykset havaitaan ja tietoturva-asetukset täyttyvät. (UK NCSC et al. 2023, 16.)

Automaatiolla on iso rooli siinä, että poikkeavat tapahtumat huomataan. Tekoälymallit käsittelevät valtavia tietovirtoja, ja ihmisen on mahdoton löytää hyökkäyksiä ilman, että hyödynnetään tekoälyä tuntemattomien uhkien havaitsemisessa.

Sääntelyn seuraaminen

Tekoälyn käyttöön liittyy monia eettisiä ongelmia tietoturvallisuusongelmien lisäksi. Tämän vuoksi on tärkeää, että yritykset seuraavat tarkasti lainsäädännön kehittymistä ja erilaisia vaatimuksia, jotka koskevat tekoälyprojekteja. Osa aiemmin mainituista tekoälyn turvallisuuskehyksistä sisältää tekoälyn eettisten ongelmien hallinnan.

Tekoälyn vastuullinen hyödyntäminen on aikamme isoimpia tieteellisiä haasteita. Eettisen tekoälyn luominen ei tarkoita vain eettisiä tekoälyalgoritmeja, vaan etiikka pitää ottaa huomioon kaikissa tekoälyprojektin suunnittelemisen vaiheissa. Eettiset vaatimukset tulisi johtaa yleisesti hyväksytyistä tekoälyn eettisistä periaatteista, ja vaatimusten määrittelyvaiheessa esitetyt eettiset kriteerit tulisi sovittaa omaan tekoälysysteemiin sopivaksi. Näin voidaan varmistua siitä, että tekoälyprojekti täyttää eettiset vaatimukset. (Lu et al. 2024, 1.)

Euroopan tekoälysäädös (AI Act) luokittelee tekoälyn riskien perusteella neljään kategoriaan. Ensimmäinen näistä on kielletty riski, eli tekoälyn ominaisuudet, joiden käyttö kielletään. Esimerkkejä tällaisista ovat manipuloiva tekoäly tai sosiaalisen pisteytyksen järjestelmät. Tätä seuraa korkean riskin tekoäly, rajoittuneen riskin tekoäly ja minimaalisen riskin tekoäly. Säädös koskee kaikkia

toimijoita, jotka ottavat tekoälyä käyttöön kaupallisessa mittakaavassa tai kehittävät sitä ja haluavat toimia EU-alueella, oli niiden fyysinen sijainti missä tahansa. Euroopan unioni perustaa tekoälytoimiston, joka toimii Euroopan komissiossa, ja asettaa aikarajat sille, kuinka pian tietyn riskiluokituksen järjestelmät tulee saattaa säädöksen mukaiseksi. (EU Artificial intelligence Act 2024.)

6 Pohdinta

6.1 Johtopäätökset

Opinnäytetyössä tutkittiin erittäin ajankohtaisia yrityksen tietoturvaan kohdistuvia uhkia ja niihin varautumista. Tutkimusmenetelmänä käytettiin narratiivista kirjallisuuskatsausta. Tutkimuksesta käy ilmi, että tekoälyn käyttöön liittyy erilaisia ja moniulotteisia uhkia, eikä niihin ole varauduttu. Yrityksillä on suuri tarve hyödyntää tekoälyä, koska sillä voi tehostaa tuotantoa ja säästää henkilöstökuluissa, mutta se tapahtuu harmillisen usein tietoturvallisuuden kustannuksella. Tietoturvallisuus voi unohtua, kun kehitys on nopeaa ja kaikki haluavat päästä hyödyntämään kehityksen tuomia etuja. Seuraavaksi käydään läpi, miten tutkimus vastaa tutkimuskysymyksiin.

6.1.1 Mitä uhkia generatiivinen tekoäly tuo yrityksen tietoturvallisuudelle

Tutkimuksesta ilmenee, että generatiivisen tekoälyn käytön nopea kasvu voimistaa ennestään tunnettuja tietoturvallisuusuhkia. Generatiivisen tekoälyn työkalut auttavat verkkorikollisia tekemään hienostuneempia suojauksia kiertäviä hyökkäyksiä ja luomaan haittaohjelmia. Verkkorikolliset voivat tehdä generatiivisen tekoälyn avulla syväväärennöksiä. Syväväärennösten avulla voidaan luoda uskottavia kalastelun ja sosiaalisen hakkeroinnin hyökkäyksiä, joista löytyy lukuisia esimerkkejä.

Sen lisäksi, että vanhat tietoturvariskit voimistuvat, tekoälyn kehitys tuo myös uusia tietoturvallisuusriskejä. Yritykset voivat menettää tarkoin varjeltuja tietoja, kun työntekijät käyttävät generatiivisen tekoälyn ohjelmia työskentelyn helpottamiseen. Markkinatutkimuksista selvisi, että isot yritykset kehittävät lukuisia tekoälyjärjestelmiä, jotka ovat teknologiana erittäin haavoittuvia. Niitä koskevat monet erilaiset uhat, kuten mallin varastaminen, tietomyrkytys, syötehyökkäys, toimitusketjun hyökkäykset sekä perinteiset tietoturvallisuusuhat.

6.1.2 Miten tekoälyn tuomia uhkia voidaan torjua yrityksissä?

Yritysten tulisi harkita turvallisuuskehysten käyttöönottoa torjuakseen tekoälyn vahvistamia tai luomia tietoturvallisuusuhkia, erityisesti jos niillä on omia tekoälyjärjestelmiä. Vaihtoehtoisesti yritysten tulisi ottaa tekoälyn riskienhallinta osaksi heidän omaa tietoturvakehystänsä. Tietoturvakehysten lisäksi tutkimissani markkinakatsauksissa ja valtiollisten tahojen ohjeissa nostettiin esiin nollaluottamus-tietoturvamalli, jossa mikään käyttäjä tai järjestelmä ei ole oletusarvoisesti luotettu. Nollaluottamus-tietoturvamallia käyttävässä yrityksessä kaikki yhteydenotot yrityksen eri järjestelmiin vaativat jatkuvaa tunnistautumista ja varmennusta. Lisäksi suositeltiin, että yritys rajoittaa generatiivisen tekoälyn käyttöä uhkien rajaamiseksi.

6.1.3 Miten tekoäly voisi auttaa uhkien torjunnassa?

Tekoälyn mahdollisuuksia tutkittaessa tietoturvallisuuden kehittämisessä selvisi, että tietoturvauhkien torjunnassa on hyödynnetty jo pidempään tekoälyä uhkien tunnistamisessa ja uhkatiedustelussa, mutta tekoälyllä on tärkeä rooli myös uusien tekoälyuhkien torjunnassa. Kun tutkin tekoälyn mahdollisuuksia nollaluottamus-tietoturvallisuusmallissa, niin löysin paljon tapoja, miten tekoäly voi parantaa yrityksen tietoturvaa. Erityisesti reaaliajassa tapahtuvat tietoturvallisuustoimet nousivat esiin. Tekoäly voi käsitellä valtavia määriä tietoa, seurata tapahtumia ja tehdä välittömästi tarvittavia toimia, kuten käyttöoikeuksien rajaaminen tai pääsyn estäminen.

6.1.4 Mitkä ovat tekoälyprojektin turvallisen kehittämisen käytännöt?

Tutkimuksessa tarkastellussa eri valtiollisten kyberturvallisuusvirastojen julkaisemassa ohjeessa käytiin läpi tekoälyjärjestelmän turvallinen kehittäminen suunnitteluvaiheesta käyttövaiheeseen. Siinä korostettiin, että tietoturvariskien hallintaa tulee tehdä jokaisessa vaiheessa, jotta ei tule yllätyksiä. Tutkimuksen lopussa nostettiin esiin tarve seurata sääntelyn kehittymistä eettisten ongelmien välttämiseksi.

6.2 Suositukset

Vaikka tekoäly tuo paljon uusia uhkia ja helpottaa hyökkääjien työtä, sillä on myös paljon potentiaalia uhkien havaitsemisessa ja torjunnassa. Tulosten perusteella vaikuttaa siltä, että joidenkin tekoälyn luomien uhkien torjuminen on mahdollista vain tekoälyä hyödyntämällä. Tekoälyä hyödyntävän yrityksen tulisi harkita tekoälyn käyttöön liittyvien turvallisuuskehysten tai tietoturvamallien käyttöönottoa, kouluttaa henkilöstöä tekoälyn riskeistä ja huomioida generatiivisen tekoälyn käyttö yrityksen tietosuojaohjeissa. Lisäksi tietoturvajärjestelmien kehitystä kannattaa seurata, sillä tällä hetkellä yritykset ovat erittäin alttiita tekoälyä hyödyntäville hyökkäyksille. Tekoälyyn liittyvät säädökset muuttuvat nopeasti, ja niitä tulee seurata erityisesti, jos yritys kehittää omia tekoälysovelluksia.

Jatkotutkimusta voitaisiin tehdä esimerkiksi nollaluottamusjärjestelmän käyttöönotosta yrityksissä. Se nostettiin esiin muutamissa lukemissani tutkimuksissa, koska generatiivisen tekoälyn myötä uhkien torjuminen vaikeutuu entisestään. Toinen mielenkiintoinen jatkotutkimusaihe voisi olla yrityksen oman generatiivisen tekoälyprojektin käyttöönoton selvitys. Tekoälyn tietoturvakehykset ovat myös tarkastelun arvoinen aihe.

6.3 Oma oppiminen

Valitsin opinnäytteen aiheen, koska halusin täydentää omaa tietoturvallisuusosaamistani. Koska aihe on hyvin tuore, laajemmin sitä tutkivaa kirjallisuutta oli vaikea löytää. Sen vuoksi päälähteinä opinnäytetyössä käytettiin markkinatutkimuksia. Hiddenlayerin tutkimus osoittautui hyvin osuvaksi lähteeksi ja siitä löytyi vastauksia moniin tutkimuskysymyksiin. Huomasin kuitenkin, että markkinatutkimuksista vain harvat sisälsivät hyvää lähdemateriaalia. Tutkimuksen loppupuolella löysin erilaisia virastojen tekemiä ohjeita, jotka olivat lähteenä parempia kuin markkinatutkimukset. Opinnäytetyö vastasi kaikkiin tutkimuskysymyksiin, mutta joitakin osa-alueita olisi voinut tutkia syvällisemmin näiden virastojen julkaisujen kautta.

Lähteet

Andersson, A., Hedström, K. ja Karlsson, F. 2022. Standardizing information security – a structurational analysis. Information and Management, volume 59, Issue 3, April 2022, s.1. Luettavissa: https://doi.org/10.1016/j.im.2022.103623. Luettu: 10.9.2024.

Australian Signals Directorate's Australian Cyber Security Centre (ASD's ACSC) in collaboration with international partners 2023. Engaging with Artificial Intelligence (AI). Luettavissa: https://www.cyber.gov.au/resources-business-and-government/governance-and-user-education/artificial-intelligence/engaging-with-artificial-intelligence. (PDF). Luettu: 12.10.2024.

BBC Teach. AI: 15 key moments in the story of artificial intelligence. The promise of intelligence. Luettavissa: https://www.bbc.co.uk/teach/articles/zh77cqt. Luettu: 11.9.2024.

Borgeaud, A. 18.3.2024. Artificial intelligence (AI) in cybersecurity - statistics & facts Luettavissa: https://www.statista.com/topics/12001/artificial-intelligence-ai-in-cybersecurity/#topicOverview. Luettu: 3.10.2024.

Chen, H. & Magramo, K. 4.2.2024. Finance worker pays out $25 million after video call with deepfake 'chief financial officer'. Luettavissa: https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html. Luettu: 15.9.2024.

Cyber Security Agency of Singapore (CSA) 2024. Guidelines and Companion Guide on Securing AI Systems. Luettavissa: https://www.csa.gov.sg/Tips-Resource/publications/2024/guidelines-on-securing-ai. Luettu: 20.10.2024.

Dash, B. 2024. Zero-Trust Architecture (ZTA): Designing an AI-Powered Cloud Security Framework for LLMs' Black Box Problems. Current Trends in Engineering Science (CTES). Luettavissa: http://dx.doi.org/10.2139/ssrn.4726625. Luettu: 8.10.2024.

Douglas Heaven, W. 18.11.2022. Why Meta's latest large language model survived only three days online. MIT Technology Review. Luettavissa: https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/. Luettu 20.9.2024.

ENISA (European Union Agency for Cybersecurity) 2024. ENISA threat landscape 2024. July 2023 to June 2024. September 2024. Luettavissa: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024. Luettu: 8.9.2024.

EU artificial intelligence act 2024. Luettavissa: https://artificialintelligenceact.eu/high-level-summary/ Luettu: 14.10.2024.

Fleck, A. 2024. Cybercrime Expected To Skyrocket in Coming Years. Statista. Luettavissa: https://www.statista.com/chart/28878/expected-cost-of-cybercrime-until-2027/. Luettu: 14.10.2024.

F-Secure. Luettavissa: https://www.f-secure.com/fi/articles/what-is-cyber-security. Luettu: 7.9.2024.

Feuerriegel, S., Hartmann, J., Janiesch, C. & Zschec, P. 2023. Generative AI. Business & Information Systems Engineering. Springer Link. Sivut 111–126. Luettavissa: https://doi.org/10.1007/s12599-023-00834-7. Luettu 12.9.2024.

Helsingin Yliopisto. Luettavissa: https://blogs.helsinki.fi/opiskelijan-digitaidot/4-tietoturva/4-1-tietoturvan-ja-tietosuojan-perusteet/tietoturvan-edellytykset/. Luettu: 7.9.2024.

Hiddenlayer 2024. AI threat landscape report. Luettavissa: https://hiddenlayer.com/threatreport2024/ (vaatii kirjautumisen). Luettu: 14.9.2024.

Huawei Artificial intelligence technology 2023. Official Textbooks for Huawei ICT Academy. Huawei Technologies Co., Ltd. Hangzhou, China. Springer Nature. E-kirja. Luettavissa: https://doi.org/10.1007/978-981-19-2879-6. Luettu: 10.9.2024.

Hyppönen, M. 2022. If it is smart it is vulnerable. John Wiley & Sons inc. Hoboken, New Jersey. E-kirja. Luettu: 13.9.2024.

IBM 2024. Cost of a Data Breach Report 2024. Luettavissa: https://www.ibm.com/reports/data-breach (vaatii kirjautumisen). Luettu: 18.10.2024.

Kyberturvallisuuskeskus 2022. Toimintaohje – Palvelunestohyökkäys. Luettavissa: https://www.kyberturvallisuuskeskus.fi/fi/julkaisut/toimintaohje-palvelunestohyokkays (PDF). Luettu: 16.10.2024.

Lindroos, M. 2019. Tietoturvallisuuden kehittäminen iso/iec 27001-standardin vaatimusten mukaisesti. Ylempi AMK-opinnäytetyö. Turun ammattikorkeakoulu. Luettavissa: https://urn.fi/URN:NBN:fi:amk-2019111521312. Luettu: 8.9.2024.

Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D. & Jacquet, A. 2024. Responsible AI Pattern Catalogue: A Collection of Best Practices for AI Governance and Engineering. ACM Computing Surveys. Volume 56, Issue 7 Article No.: 173, 1–35. Luettavissa: https://doi.org/10.1145/3626234. Luettu: 13.10.2024.

LUT Academic Library 2024. Tekoäly tiedonhankinnan apuna. LibGuides. Luettavissa: https://libguides.lut.fi/tekoaly/chatpgt_kielimallit. Luettu: 13.9.2024.

Microsoft Microsoft Security. Mitä tietoturva on? Luettavissa: https://www.microsoft.com/fi-fi/security/business/security-101/what-is-data-security. Luettu: 8.9.2024.

Netskope 2024. Cloud and Threat Report: AI Apps in the Enterprise. Luettavissa: https://www.netskope.com/netskope-threat-labs/cloud-threat-report/july-2024-ai-apps-in-the-enterprise. Luettu: 1.10.2024.

OpenAI 2024. Enterprise privacy at OpenAI. Luettavissa: https://openai.com/enterprise-privacy/. Luettu: 27.9.2024.

Paramesha, M., Rane, N. L. & Rane, J. 2024. Artificial Intelligence, Machine Learning, and Deep Learning for Cybersecurity Solutions: A Review of Emerging Technologies and Applications. Partners Universal Multidisciplinary Research Journal (PUMRJ), 01(02), 84–109. Luettavissa: https://doi.org/10.5281/zenodo.12827076. Luettu: 10.10.2024.

Ponemon institute 2024. State of AI in cybersecurity, Report 2024. Luettavissa: https://mixmode.ai/state-of-ai-in-cybersecurity-2024-download/ (vaatii kirjautumisen). Luettu: 5.10.2024.

Saghiri, A., Vahidipour, S., Jabbarpour, M., Sookhak, M., Forestiero, A. 2022. Survey of Artificial Intelligence Challenges: Analyzing the Definitions, Relationships, and Evolutions. Applied Sciences 2022, 12(8), 4054. Luettavissa: https://doi.org/10.3390/app12084054. Luettu: 12.9.2024.

SFS ry. Suomen standardisoimisliitto 2023. Johdanto. Yleistä. Luettavissa: https://sales.sfs.fi/fi/index/tuotteet/SFS/CENISO/ID2/2/1288464.html.stx. Luettu 8.9.2024.

Samoili, S., López Cobo, M., Delipetrev, B., Martínez-Plumed, F., Gómez, E., and De Prato, G. 2021. AI Watch. Defining Artificial Intelligence 2.0. Towards an operational definition and taxonomy for the AI landscape, EUR 30873 EN, Publications Office of the European Union, Luxembourg, 2021 Luettavissa: https://data.europa.eu/doi/10.2760/019901. Luettu: 9.9.2024.

UK National Cyber Security Centre (NCSC), the US Cybersecurity and Infrastructure Security Agency (CISA) ja kansainväliset kumppanit 2023. Guidelines for secure AI system development. Luettavissa: https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development. Luettu: 6.10.2024.

Vilkka, H. 2023. Kirjallisuuskatsaus metodina, opinnäytetyön osana ja tekstilajina. Art House. Helsinki.

Williams, C.M., Chaturvedi, R. & Chakravarthy, K. 2020. Cybersecurity Risks in a Pandemic. Journal of Medical Internet Research, Vol 22, No 9 September. Luettavissa: https://www.jmir.org/2020/9/e23692/. Luettu: 14.9.2024.

Yhdysvaltain valtiovarainministeriö 2024. Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Services Sector. Luettavissa: https://home.treasury.gov/news/press-releases/jy2212 (PDF). Luettu: 20.9.2024.

Zscaler 2024. Zscaler ThreatLabz 2024 AI Security Report. Luettavissa: https://info.zscaler.com/resources-industry-reports-threatlabz-ai-security-2024 (vaatii kirjautumisen). Luettu: 15.9.2024.