

Mari Tölli

TEKOÄLYN TIETOTURVA

Tekoälyn yleisimmät riskit, riskienhallinta ja EU:n
tekoälyasetus

Diplomityö
Informaatioteknologian ja viestinnän tiedekunta
Tarkastaja: Marko Helenius
Tarkastaja: Tarja Tiainen
Heinäkuu 2025

TIIVISTELMÄ

Mari Tölli: Tekoälyn tietoturva
Diplomityö
Tampereen yliopisto
Tietotekniikan Diplomi-insinööri ohjelma
Heinäkuu 2025

Kirjallisuuskatsauksen aiheena on tekoälyn tietoturva. Tekoälyn yleistyessä ja sen käytön laajetessa aihe on ajankohtainen ja yleisesti kiinnostusta herättävä. EU:n tekoälyasetuksen voimaantulon myötä myös tekoälyn sääntely ja sen vaikutukset tekoälyjärjestelmiin, niiden kehittämiseen ja tietoturvaan korostuvat entisestään.

Katsauksessa keskitytään tekoälyn riskeihin ja niiden hallintaan. Tavoitteena on tunnistaa yleisimmät riskit ja kartoittaa, mitkä hallintakeinot sopivat niiden hallintaan. Tätä näkökulmaa selvitetään tutkimuskysymyksellä: Mitkä ovat tekoälyn yleisimmät tietoturvariskit ja miten niitä voidaan hallita? Lisäksi katsauksessa tutustutaan EU:n tekoälyasetukseen ja sen vaikutuksiin tekoälyjärjestelmien kehittämiseen ja käyttöön Euroopan unionin alueella. Tähän aiheeseen kohdistuu tutkimuskysymys: Miten EU:n tekoälyasetus vaikuttaa tekoälyjärjestelmien tietoturvakäytäntöihin ja -vaatimuksiin EU:n alueella?

Katsauksen analyysimenetelmänä käytettiin sisällönanalyysiä, jossa keskityttiin materiaalin tekstien merkityksen ja teemojen tunnistamiseen sekä niiden luokitteluun. Tekoälyn yleisimpien riskien ja niiden hallintamenetelmien tutkimisessa ensisijaisena lähteenä oli ProQuest-tietokanta ja toissijaisena lähteenä Google Scholar -hakukone. Aineiston haku rajattiin aikavälille 1.1.2020-16.4.2025. Hakuprosessissa käytettiin yksinkertaisia hakusanoja ilman hakusanalausekkeita. Hakutulokset käytiin yksitellen läpi ja arvioitiin niiden olennaisuus suhteessa tutkimuskysymykseen. Lopulliseen analyysiin valikoitui 31 vertaisarvioitua julkaisua.

Analyysissä tutkimusaineistosta tunnistettiin kaikki mainitut riskit ja niiden hallintamenetelmät. Tunnistetut riskit jaettiin ryhmiin ja jokaisen riskiryhmän riskien esiintyvyyksiä laskettiin materiaalista. Tämän perusteella tunnistettiin yleisimpien riskien ryhmät. Kuudelle yleisimmälle riskiryhmälle kartoitettiin tutkimusmateriaalista hallintakeinoja.

Tekoälyasetuksen vaatimusten analysoinnin ensisijaisena lähteenä oli EU:n tekoälyasetus ja toissijaisena lähteenä Euroopan komission siitä antamat ohjeistukset ja tiedotteet. Asetus ja muu materiaali käytiin läpi ja siitä jäsennettiin ja analysoitiin tutkimuskysymyksen kannalta olennaiset kohdat. Lopuksi tehtiin myös kaksi tapausanalyysiä kuvitteellisiin tekoälyjärjestelmiin. Tapausanalyysillä pyrittiin havainnollistamaan asetuksen vaikutuksia tekoälyjärjestelmiin.

Analyysin perusteella tekoälyn yleisimpiä riskejä olivat mallin ja datan manipulointiin liittyvät riskit, yksityisyydensuojaan ja tietosuojaan liittyvät riskit, tekoälyjärjestelmien tietoturvahaavoittuvuuksiin liittyvät riskit, mallin puolueellisuuteen liittyvät riskit, tekoälyn käyttö kyberhyökkäyksissä ja tekoälyn käyttöön liittyvät eettiset riskit.

EU:n tekoälyasetuksen vaatimuksilla pyritään varmistamaan tekoälyjärjestelmien turvallisuus ja luotettavuus sekä suojelemaan perusoikeuksia. Asetus kieltää joitain tekoälyn käyttötapauksia. Kiellettyjä ovat järjestelmät, jotka manipuloivat tai harhaanjohtavat käyttäjiä, hyödyntävät käyttäjän haavoittuvuutta, tekevät syrjivää sosiaalista pisteytystä tai mahdollistavat laajamittaisen valvonnan ilman oikeudellista perustaa.

Suuririskisille tekoälyjärjestelmille asetuksessa vaaditaan muun muassa perusteellista dokumentaatiota ja elinkaaren ajan kestävää seurantaa, rekisteröintiä EU:n tietokantaan, CE-merkintää, vaatimustenmukaisuuden osoittamista, koko järjestelmän elinkaaren kestävää, käyttötarkoitukseen sopivaa toiminnan tarkkuutta, vakautta ja kyberturvallisuutta sekä ihmisen suoritamaa valvontaa.

Vähäisen riskin järjestelmille on asetuksessa kohdistettu avoimuusvelvoitteita. Käyttäjälle on kerrottava tekoälyn käytöstä vuorovaikutustilanteessa. Myös tekoälyn tuottamat synteettiset tuotokset tulee merkitä keinotekoisesti tuotetuiksi.

Avainsanat: Tekoälyn tietoturva, tekoälyn riskit, EU:n tekoälyasetus
Tämän julkaisun alkuperäisyys on tarkastettu Turnitin Originality Check -ohjelmalla.

ABSTRACT

Mari Töllli: The security of artificial intelligence
Master's Thesis in Technology
Tampere University
Master's Programme in Information Technology
July 2025

The topic of this literature review is the security of artificial intelligence (AI). As AI becomes more common and its use expands, the topic is both appropriate and of general interest. As the EU Artificial Intelligence Act (AI act) is now in force, the regulation of AI and its impact on AI systems, their development, and security is becoming increasingly highlighted.

The review focuses on the risks related to AI and their management. The objective is to identify the most common risks and to find out which control measures are suitable for managing them. This point of view is examined through the research question: What are the most common cybersecurity risks of AI, and how can they be managed? In addition, the review studies the EU AI Act and its implications for the development and use of AI systems within the European Union. The related research question is: How does the EU AI Act affect the cybersecurity practices and requirements of AI systems within the EU?

The analysis method used in the review was content analysis, which focused on finding the meanings and themes of the texts in the material and categorizing them. In the review of the most common AI risks and their management methods, the primary source was the ProQuest database, and the secondary source was the Google Scholar search engine. The data collection was narrowed to the period from January 1, 2020, to April 16, 2025. Simple keywords were used in the search process without search phrases. The search results were evaluated individually and assessed for their relevance to the research question. A total of 31 peer-reviewed publications were chosen for the final analysis.

In the analysis, all mentioned risks and their management methods were collected from the research material. The gathered risks were grouped, and the frequency of occurrence for each risk group was calculated from the material. Based on this, the most common risk groups were identified. For the six most common risk groups, control measures were searched from the research material.

The primary source for analyzing the requirements of the AI Act was the EU AI Act itself, and the secondary source was the guidelines and publications issued by the European Commission. The regulation and other materials were reviewed, and the relevant sections were organized and analyzed in relation to the research question. In addition, two case analyses were conducted on hypothetical AI systems. These case studies aimed to demonstrate the effects of the regulation on AI systems.

Based on the analysis, the most common AI risks included risks related to model and data manipulation, privacy and data protection, vulnerabilities in AI systems, model bias, the use of AI in cyberattacks, and ethical issues related to AI use.

The requirements of the EU AI Act try to ensure the safety and reliability of AI systems and to protect fundamental rights. The regulation prohibits some uses of AI. Prohibited systems include those that manipulate or mislead users, exploit user vulnerabilities, conduct unfair social scoring, or enable large-scale surveillance without a legal basis.

For high-risk AI systems, the regulation requires, for example, thorough documentation and continuous monitoring throughout lifecycle, registration in the EU database, CE certificate, demonstration of compliance, accuracy, stability, and cybersecurity appropriate to the intended use case through the system's lifecycle, as well as human oversight.

For low-risk systems, the regulation sets transparency obligations. Users must be told when interacting with AI. Synthetic products created by AI must be labeled as artificially generated.

Keywords: the security of artificial intelligence, risks of artificial intelligence, EU AI Act

The originality of this thesis has been checked using the Turnitin Originality Check service.

TEKOÄLYN KÄYTTÖ OPINNÄYTTEESSÄ

Opinnäytteessäni on käytetty tekoälysovelluksia:

- Ei
 Kyllä

Ilmoitukseni mukaan olen käyttänyt opinnäytteessäni tutkielmanprosessin aikana seuraavia tekoälysovelluksia:

Tekoälysovellusten nimet ja versiot: Microsoft Copilot 1.25032.166.0 (julkaisu 1.4.2025.)

Käyttötarkoitus: Tekoälyä käytettiin materiaalin keräysvaiheessa hakutulosten läpikäymiseen. Tekoälyn avulla luotiin kustakin hakutuloksesta tiivistelmä käskyillä, "Luo tästä tiedostosta tiivistelmä". Lisäksi tekoälyä käytettiin hakutulosten vertaisarvioinnin tarkistamiseen ja julkaisseen lehden arviointiin käskyillä, "Onko julkaisu vertaisarvioitu? Julkaisun DOI on xx.xxxxx/xxxxxxx.xxxx.xxxxxx" ja "Onko lehti indeksoitu?". Lisäksi tekoälyä käytettiin työn lopussa kieliopin ja oikeinkirjoituksen tarkastuksen tukena.

Osiot, joissa tekoälyä on käytetty: Materiaalin keräys, läpikäynti ja rajaaminen. Kieliopin tarkastamisen tukena työn viimeistely vaiheessa.

Olen tietoinen siitä, että olen täysin vastuussa koko opinnäytteeni sisällöstä, mukaan lukien osat, joissa on hyödynnetty tekoälyä, ja hyväksyn vastuun mahdollisista eettisten ohjeiden rikkomuksista.

SISÄLLYSLUETTELO

1. JOHDANTO	1
1.1 Työn aihe ja tavoite.....	1
1.2 Tutkimuskysymykset.....	2
1.3 Määrittely ja rajaukset.....	3
2. TYÖN TAUSTA.....	4
3. MENETELMÄT	6
3.1 Tekoälyn yleisimmät riskit ja niiden hallinta.....	6
3.1.1 Tietokanta ja muut lähteet.....	6
3.1.2 Hakusanat ja -tavat.....	7
3.1.3 Aika- ja kielirajaukset	9
3.1.4 Toteutus.....	9
3.2 EU:n tekoälyasetuksen vaatimusten analyysi	11
3.2.1 Analyysin rajaus.....	11
3.2.2 Lähteet.....	11
3.2.3 Toteutus.....	12
4. TULOKSET.....	16
4.1 Tekoälyn yleisimmät riskit.....	16
4.1.1 Tekoälyn riskien hallinta.....	21
4.2 EU:n tekoälyasetuksen vaatimukset	22
4.2.1 Kielletyt käyttötapaukset	23
4.2.2 Suuririskiset tekoälyjärjestelmät	24
4.2.3 Tapausanalyysit.....	28
4.2.4 Vähäisen riskin tekoälyjärjestelmät	30
4.2.5 Minimaalisen riskin tai ei lainkaan riskiä tekoälyjärjestelmät.....	31
4.2.6 Käyttöönottajille asetetut vaatimukset	32
5. YHTEENVETO JA POHDINTA	34
5.1 Tulosten yhteenveto	34
5.2 Pohdinta -tekoälyasetus ja tekoälyn yleisimmät riskit.....	35
5.3 Pohdinta -katsauksen kritiikki ja jatkotutkimusmahdollisuudet.....	37
LÄHTEET	40
LIITE I: KIRJALLISUUSLUETTELOT	44

1. JOHDANTO

Kirjallisuuskatsauksessa keskitytään tekoälyn tietoturvaan. Tarkastelun kohteena ovat tekoölyyn liittyvät tietoturvariskit ja niiden hallinta. Tekoälyn käytön yleistyessä ja laajetessa yhteiskunnassa aihe on ajankohtainen ja yleisesti kiinnostava. Lisäksi katsauksessa tutustutaan viime vuonna voimaan tulleen Euroopan unionin (EU) tekoälyasetuksen tuomiin velvoitteisiin tietoturvavaatimuksia ja riskienhallintaa silmällä pitäen. Erityisen tarkastelun kohteena ovat asetuksesta tulevat velvoitteet tekoälyjärjestelmien tietoturvaan, kehitykseen ja käyttöönottoon liittyen.

1.1 Työn aihe ja tavoite

Kuten edellä todettiin, katsauksessa tarkastelun kohteena on kaksi aihealuetta, tekoälyn käyttöön liittyvät tietoturvariskit ja EU:n tekoälyasetuksesta tulevat tietoturvavaatimukset. Katsauksessa tarkastellaankin tekoälyn tietoturvaa lähinnä riskien näkökulmasta.

Katsauksessa kartoitetaan tekoälyn käyttöön liittyviä tietoturvariskejä ja niiden hallintakeinoja. Tekoälyn kehitys on ollut viime vuosina nopeaa ja sen käyttö on yleistynyt nopeasti yhteiskunnan eri osa-alueilla niin työssä kuin vapaa-ajallakin. Tekoälyn käyttöön liittyvien riskien kartoittaminen ja ymmärtäminen sekä tehokkaat keinot riskien hallintaan ovat tärkeitä, jotta tekoälyn käyttö olisi turvallista. Yksi tämän katsauksen tavoitteista on saada hyvä yleiskuva tekoälyn riskeistä ja kartoittaa myös keinoja näiden riskien hallintaan.

Katsauksessa tarkastellaan tekoälyn tietoturvaa myös sääntelyn näkökulmasta. EU:n tekoälyasetus on ollut voimassa alle vuoden. Se määrittää uusia velvoitteita tekoälyjärjestelmien tietoturvakäytännöille ja -vaatimuksille. Katsauksessa tavoitteena on EU:n tekoälyasetuksen ymmärtäminen ja selvittää, mitä vaatimuksia EU:n tekoälyasetus asettaa tekoälyjärjestelmien tietoturvakäytäntöihin Euroopan unionin alueella. Sääntely on myös yksi riskienhallinnan keino, joten katsauksen lopussa pohditaan, miten tekoälyasetus vastaa katsauksen kartoituksessa esiin nousseisiin riskeihin.

Kirjallisuuskatsauksessa pyritään tuottamaan hyvä yleiskuva näistä aiheista ja luomaan laajempaa ymmärrystä tekoälyn tietoturvariskeistä ja niiden hallinnasta sekä EU:n alueella voimassa olevasta uudesta tekoälysääntelystä ja sen vaikutuksista.

Katsaus on toteutettu perinteisenä kirjallisuuskatsauksena, jossa kumpaankin katsauksen aihealueeseen on perehdytty erikseen. Perinteinen kirjallisuuskatsaus on tehokas tapa luoda hyvä yleiskuva ja laajempi ymmärrys tutkittavaan aiheeseen. Siksi se sopii erinomaisesti myös tämän katsauksen tavoitteisiin. Katsaus noudattaa perinteisen kirjallisuuskatsauksen kaavaa, jossa ensin tutustutaan tutkimusaiheeseen ja luodaan tutkimuskysymykset. Sen jälkeen suoritetaan kirjallisuuden haku, valinta ja arviointi. Valittu kirjallisuus analysoidaan ja esitetään tulokset, minkä jälkeen tuloksista voidaan juontaa johtopäätökset. Tässä katsauksessa molempiin aiheisiin perehdyttiin ensin erikseen ja niiden yhtäläisyyksiä pohdittiin työn lopussa Yhteenvetoluvussa.

1.2 Tutkimuskysymykset

Kirjallisuuskatsauksessa on kaksi löyhästi toisiinsa liittyvää laajaa aihetta. Aiheet on rajattu kahdella tutkimuskysymyksellä, joiden avulla pyritään kohdentamaan tarkastelua ja rajaamaan aiheet resurssien kannalta sopivan kokoisiksi ilman, että aiheiden välinen yhteys katoaa.

Tekoälyn riskien näkökulmaan perehdytään tutkimuskysymyksellä, mitkä ovat tekoälyn yleisimmät tietoturvariskit ja miten niitä voidaan hallita? Tutkimuskysymys rajaa tarkastelun tekoälyn tietoturvaan liittyviin riskeihin ja keskittää tarkastelun yleisimpiin riskiluokkiin sekä näiden riskien hallintaan. Samalla kysymyksen rajaus varmistaa, että katsauksessa tarkastellaan riskejä nimenomaan tietoturvanäkökulmasta. Tutkimuskysymys vastaa katsauksen tavoitteeseen luoda hyvä yleiskuva tekoälyn käyttöön liittyvistä tietoturvariskeistä. Vaikka kysymys rajaa ulos joitain harvinaisempia riskejä riskien vakavuudesta riippumatta, antaa yleisimpien riskien kartoitus laajan tilannekuvan tekoälyn riskeistä.

Toinen tutkimuskysymys koskee tekoälyn sääntelyä EU:n alueella. Miten EU:n tekoälyasetus (AI Act) vaikuttaa tekoälyjärjestelmien tietoturvakäytäntöihin ja -vaatimuksiin Euroopan unionin alueella? Tämä kysymys keskittää tarkastelun tekoälyasetuksen tuomiin tietoturvaan liittyviin vaatimuksiin, jotka kohdistuvat tekoälyjärjestelmiin, niiden tarjoajiin ja käyttöönottajiin. Rajaus jättää tarkastelun ulkopuolelle asetuksen yhteiskunnalliset ja taloudelliset vaikutukset. Tarkastelun ulkopuolelle jää myös pidemmän aikavälin vaikutusten arviointi. Tutkimuskysymyksellä saavutetaan kuitenkin katsauksen tavoite tekoälyasetuksen tietoturva vaatimusten ymmärtämisestä, ja yhdessä toisen tutkimuskysymyksen kanssa voidaan pohtia, kuinka hyvin tekoälyasetuksen vaatimat muutokset kohdistuvat tekoälyn käyttöön liittyviin tietoturvariskeihin.

1.3 Määrittely ja rajaukset

Katsaus keskittyy tekoälyn yleisimpien tietoturvariskien kartoittamiseen ja EU:n tekoälyasetuksen tietoturvaan liittyvien vaatimusten tarkasteluun. Rajaus jättää tarkastelun ulkopuolelle tekoälyyn liittyvät muut riskit, kuten esimerkiksi yhteiskunnalliset ja taloudelliset riskit. Samoin riskien realisoidumisen mahdollisiin seurauksiin ei katsauksessa paneuduta kovinkaan syvästi.

Tekoälyä ei ole katsauksessa rajattu kovinkaan tarkasti. Katsauksessa tekoälyllä tarkoitetaan järjestelmää, joka koostuu algoritmeista ja malleista, jotka mahdollistavat suhteellisen itsenäisen datan analysoinnin, mallintamisen ja päätöksenteon. Katsauksessa ei kohdennettu tarkastelua mihinkään tiettyyn tekoälyalgoritmiin tai malliin.

EU:n tekoälyasetuksen kohdalla tarkastelun ulkopuolelle jäävät yhteiskunnalliset ja taloudelliset vaikutukset. EU:n tai sen jäsenvaltioiden muita lakeja ja asetuksia ei ole juurikaan katsauksessa huomioitu, vaan katsauksessa on keskitytty vain EU:n tekoälyasetukseen, vaikka asetuksessa viitataan myös joihinkin aiempiin EU:n asetuksiin, kuten esimerkiksi tietosuoja-asetukseen.

2. TYÖN TAUSTA

Muutaman viime vuoden aikana tekoälyteknologiat ovat kehittyneet huomattavasti ja niiden käyttö on laajentunut aiempaa laajemmin yhteiskunnan eri osa-alueille, kuten terveydenhuoltoon, liikenteeseen, finanssialalle ja teollisuuteen. Kehitys on tuonut ja tuo mukanaan uusia mahdollisuuksia, mutta myös merkittäviä riskejä, joiden tunnistaminen ja hallinta on tärkeää. Tekoälyn käytön laajentuminen tarkoittaa, että yhä useammat järjestelmät ja palvelut tulevat olemaan riippuvaisia siitä. Tämä korostaa tekoälyjärjestelmien tietoturvan ja riskienhallinnan merkitystä, koska se varmistaa järjestelmien käytettävyyden ja suojaa käyttäjiä mahdollisilta uhkilta (1).

Tekoälyn riskejä on vuosien kuluessa tutkittu paljon, ja varsinkin 2020-luvulla tekoälyyn liittyvä tutkimus on kasvanut vahvasti. Tekoälyn riskien tutkimuksessa on tehty sekä yksittäisiin kuten adversaariin hyökkäyksiin keskittyviä tutkimuksia, että laajempia riskikokonaisuuksien kartoituksia (3, 4). Laajemmissa katsauksissa on keskitytty esimerkiksi generatiivisen tekoälyn ja kielimallien tietoturvaan (5) tai kartoitettu tekoälyn roolia sekä tietoturvaa parantavana, että myös uhkaavana tekijänä (6).

Euroopan unionin tekoälyasetus (asetus (EU) 2024/1689 tekoälyä koskevista yhdenmukaistetuista säännöistä) astui voimaan 1. elokuuta 2024. Asetus koskee tekoälyä ja sen yhdenmukaistettuja sääntöjä, ja sillä vahvistetaan yhtenäinen oikeudellinen kehys tekoälyjärjestelmien kehittämiseksi, markkinoille saattamiselle, käyttöönotolle ja käytölle unionissa. Asetuksen tavoitteena on paitsi luotettava tekoäly, myös varmistaa terveyden, turvallisuuden ja perusoikeuksien suojeleminen (37).

Vaikka EU:n tekoälyasetus ei olekaan ollut voimassa kauan, on siihen liittyen ehditty jo julkaista joitakin tutkimuksia. EU:n yhteinen tutkimuskeskus julkaisi raportin jo ennen asetuksen voimaantuloa. Raportissa keskityttiin analysoimaan erityisesti artiklaa 15, joka keskittyy suuririskisten tekoälyjärjestelmien tietoturvaan ja vakaaseen toimintaan (7). Myös asetuksen voimaantulon jälkeen on ehditty julkaista muutamia artikkeleita. Esimerkiksi systemaattinen kirjallisuuskatsaus asetuksen vaikutuksista tietoturvaan ja tutkimus, joka analysoi asetuksen kohtia, jotka koskevat suuririskisten ja yleiskäyttöisten tekoälyjärjestelmien tietoturvaa (8, 9). Tutkimuksia asetuksen vaikutuksista ei kuitenkaan vielä ole kovin paljon, ja laajempia tutkimuksia ei ole juurikaan vielä julkaistu. Osittain tämä johtuu siitä, että asetuksen voimaantulosta on kulunut alle vuosi, ja sitä aletaan soveltaa täysimääräisesti vasta elokuussa 2026 (38).

Tekoälyasetuksen yhteyttä tekoälyn riskeihin on myös jo ehditty tutkia, vaikka asetus on ollut voimassa vain lyhyen aikaa. Maaliskuussa 2025 julkaistu tutkimus tarkastelee, miten tekoälyasetus käsittelee tekoälyn käyttöä lainsäädännössä ja sääntelyssä, ja miten tämä liittyy tekoälyn aiheuttamiin riskeihin (10). Kaiken kaikkiaan tekoälyasetuksen ja tekoälyn riskien yhteneväisyyttä on kuitenkin tutkittu vasta vähän.

Tekoälyn tietoturvariskeistä on siis jo runsaasti tutkimustietoa ja tekoälyasetuksen vaikutuksista tekoälyn tietoturvaan on jo jonkin verran aiempaa tutkimusta. Tässä katsauksessa keskitytään kuitenkin erityisesti ajantasaisen yleiskuvan luomiseen ja ymmärryksen laajentamiseen sekä pohditaan hieman tekoälyasetuksen ja riskien yhteyttä.

3. MENETELMÄT

Tässä luvussa esitellään työssä käytetty analyysimenetelmä ja kerrotaan tarkemmin, kuinka analyysit toteutettiin. Analysoitava tutkimusmateriaali oli kummankin tutkimuskysymyksen kohdalla kvalitatiivista. Koska kyseessä oli perinteinen kirjallisuuskatsaus, valikoitui työn analyysimenetelmäksi sisällönanalyysi. Sisällönanalyysissä keskitytään tekstin merkityksen ja teemojen tunnistamiseen sekä niiden luokitteluun (11). Menetelmä mahdollistaa olennaisten osien poimimisen materiaalista, niiden jäsentelyn ja johtopäätösten tekemisen niiden pohjalta.

Sisällönanalyysin etuna on myös menetelmän joustavuus. Se mahdollistaa käsittelyn kvantitatiivisella lähestymistavalla, jossa luodaan yleiskuva aineistosta ja esitetään tulokset numeerisesti (12). Tätä lähestymistapaa päädyttiin käyttämään tekoälyasetuksen analyysin yhteydessä.

Alun perin työssä oli tarkoitus käyttää myös sääntelyn vaikutusten arviointia tekoälyasetuksen vaikutusten ja vaatimusten selvittämiseen. Tekoälyasetuksen analyysissä ei kuitenkaan ehditty aikataulun puitteissa toteuttaa näin laajaa analyysiä. Tämän takia myös tekoälyasetuksen analyysi päädyttiin lopulta tekemään sisällönanalyysin menetelmällä. Analyysiä täydentämään otettiin tapausanalyysi, jossa tarkasteltiin kuvitteellisia esimerkkitapauksia tekoälyasetusta vasten.

3.1 Tekoälyn yleisimmät riskit ja niiden hallinta

3.1.1 Tietokanta ja muut lähteet

Työssä käytettiin ensisijaisena lähteenä ProQuest-tietokantaa, joka valittiin kattavuutensa ja monipuolisuutensa vuoksi. ProQuest sisältää laajan valikoiman vertaisarvioituja ja korkealaatuisia tieteellisiä artikkeleita, opinnäytetöitä, kirjoja, raportteja ja muita julkaisuja. Tietokannan monipuolisuus auttoi kattamaan tutkimuskysymyksen laajasti, tarjosi perusteellisen ymmärryksen aiheesta ja auttoi luomaan hyvän kokonaiskuvan. ProQuest-tietokannassa on paljon tuoreita ja ajankohtaisia tutkimuksia tekoälystä, mikä varmisti, että kirjallisuuskatsaus perustui ajankohtaiseen ja tutkimuskysymyksen kannalta relevanttiin aineistoon. Se oli yliopiston tarjoamista tietokannoista parhaiten työhön sopiva. Valintaan vaikutti myös se, että ProQuest tarjoaa käyttäjäystävällisen hakukoneen ja työkaluja, jotka helpottavat aineiston rajaamista ja keräämistä. Tämä tuki aineiston tehokasta kokoamista ja säästi aikaa.

ProQuest-tietokannan lisäksi toissijaisena lähteenä käytettiin Google Scholar -hakukonetta. Se tarjoaa pääsyn laajaan valikoimaan tieteellisiä artikkeleita, kirjoja, konferenssijulkaisuja ja muita akateemisia lähteitä, ja auttoi saamaan laajemman kuvan tutkimusaiheesta. Google Scholarin tuloksia otettiin mukaan katsaukseen, koska ProQuest-tietokannasta löytyi melko rajallinen määrä tutkimuskysymyksen kannalta relevanttia aineistoa. Katsauksessa käytetystä aineistosta suurin osa, noin kaksi kolmasosaa, on ProQuest-tietokannasta ja yksi kolmasosa Google Scholarin hakujen kautta.

Google Scholar valittiin toissijaiseksi lähteeksi, koska se on helppokäyttöinen ja maksuton hakukone akateemiselle aineistolle. Vaikka sen hakutuloksia ei voikaan rajata yhtä tarkasti kuin ProQuestin, se tarjosi mahdollisuuden löytää lisämateriaalia ja täydentää analyysissä käytettävää aineistoa. Molempien lähteiden hyödyntäminen varmisti, että analyysissä käytetty aineisto oli monipuolinen, kattava ja sisälsi ajankohtaisia ja olennaisia näkökulmia tutkimuskysymykseen.

3.1.2 Hakusanat ja -tavat

Työssä käytettyjen hakusanojen valinta perustui niiden kykyyn kattaa kirjallisuuskatsauksen keskeiset teemat ja tutkimuskysymys. Valitut hakusanat tähtäsivät olennaisen, laadukkaan ja ajankohtaisen aineiston löytämiseen. Tavoitteena oli, että hauilla rajattu aineisto tarjoaisi laajan ja kattavan katsauksen tekoälyn tietoturvaan ja sen riskeihin. Työssä käytettiin sekä suomen- että englanninkielisiä hakusanoja. Suomenkielisten hakusanojen kohdalla katsauksen kannalta olennaiset hakuosumat jäivät kuitenkin vähäisiksi. Hakusanat ja niiden valintaperusteet on listattu taulukossa 1.

Hakusana	Valintaperuste
AI security ja artificial intelligence security	Ylätason hakutermi, joka rajaa tulokset tekoälyn tietoturvaan. Termi on yleisesti käytetty ja tunnistettu, ja se tuottaa laajan skaalan tekoälyn tietoturvaan liittyviä hakutuloksia. Sopii hyvin yleiskuvan luomiseen ja tutkimusaiheen taustoittamiseen.
Tekoälyn tietoturva	Yllä mainitun termin suomenkielinen vastine. Tarjosi näkymän aiheen suomenkielisiin julkaisuihin. Auttoi tuomaan esiin suomalaisia näkökulmia ja erityispiirteitä.
AI security risks ja artificial intelligence security risks	Tarkentaa haun tekoälyyn liittyviin riskeihin. Auttaa rajamaan hakutuloksia tutkimuskysymyksen näkökulmasta olennaisempaan sisältöön.

Tekoälyn tietoturvariskit	Yllä mainitun suomenkielinen vastine. Tarkoituksena tuoda esiin suomalaista näkökulmaa.
AI threats ja artificial intelligence threats	Ylätason hakutermi, joka rajaa tulokset kuitenkin tekoälyyn. Valittiin, koska riskeistä puhutaan toisinaan myös uhkina.
AI security threats ja artificial intelligence threats	Tarkempi hakutermi. Rajaa haun tuloksia tekoälyn tietoturvauhkuihin keskittyviin julkaisuihin. Tarkentaa haun tutkimuskysymyksen kannalta olennaisempaan sisältöön.
Tekoälyn tietoturva uhat	Yllä mainitun suomenkielinen vastine.
AI vulnerabilities ja artificial intelligence vulnerabilities	Tarkempi hakutermi. Ohjaa hakua keskittymään tekoälyjärjestelmien haavoittuvuuksiin ja niiden tutkimiseen keskittyviin julkaisuihin.
Tekoälyn tietoturva haavoittuvuudet	Yllä mainitun suomenkielinen vastine.
AI risk management ja artificial intelligence risk management	Rajaa hakutuloksia riskeihin ja erityisesti riskienhallintamenetelmiin kohdistuviin julkaisuihin. Keskittyy erityisesti tutkimuskysymyksen toiseen osaan.
Tekoälyn tietoturvariskien hallinta	Yllä mainitun suomenkielinen vastine. Mahdollisesti suomalaisen näkökulman tuomisen riskienhallintamenetelmiin.

Taulukko 1. Työssä käytetyt hakusanat ja niiden valintaperusteet.

Hauissa päädyttiin käyttämään sekä lyhennettä AI että termiä artificial intelligence. Termien rinnakkaisen käytön päätös tehtiin sillä perusteella, että hakutulokset vaihtelivat riippuen käytetystä termistä.

Hakusanat kattoivat tutkimuskysymyksen keskeiset teemat: tekoälyn tietoturvan, sen riskit ja niiden hallinnan. Ne auttoivat löytämään laajan valikoiman tutkimuksia ja julkaisuja, jotka käsittelevät näitä aiheita. Hakusanat auttoivat tunnistamaan tutkimuksia, jotka käsittelevät tekoälyn yleisimpiä tietoturvariskejä ja tarjoavat ratkaisuja niiden hallintaan. Hakusanat mahdollistavat aineiston tehokkaan, mutta tarpeeksi laaja-alaisen rajaamisen ja keskittymisen tutkimuskysymykseen. Haut tehtiin maalisi- ja huhtikuun 2025 aikana.

Hakuprosessin aikana hakusanoja käytettiin sellaisenaan, eikä niitä linkitetty toisiinsa tai muihin tekoälyn tietoturvaan keskittyviin termeihin JA- tai TAI-hakuehtoja käyttäen tai muita hakuominaisuuksia käyttämällä. Hauissa ei myöskään käytetty hakusanalauseita. Hakuehtoja ja hakusanalauseita ei käytetty, koska hakuprosessin aikana todettiin, että yksittäiset hakusanat toimivat tehokkaasti ja antoivat laajan ja monipuolisen skaalan tuloksia, kuitenkin rajaten hakua tarpeeksi paljon. Hakusanalauseiden ajateltiin voivan rajata hakutuloksia liikaa ja mahdollisesti jättävän haun ulkopuolelle tutkimuskysymyksen kannalta olennaisia hakutuloksia. Lisäksi hakutulosten määrä oli yksittäisiä hakusanoja käyttämälläkin hallittavissa ja käsiteltävissä, joten ei koettu tarvetta tiukemalle tulosten rajaamiselle hakusanalauseilla.

3.1.3 Aika- ja kielirajaukset

Työssä käytettiin myös aika- ja kielirajauksia. Katsaukseen valittiin aineistoa, joka on julkaistu 1.1.2020-16.4.2025 välisenä aikana. Aikarajauksella pyritään varmistamaan aineiston ajankohtaisuus ja huomioimaan alan teknologiset muutokset. Tekoäly ja sen tietoturva kehittyvät ja muuttuvat tällä hetkellä nopealla tahdilla. Aikarajaus varmistaa, että katsaus perustuu tuoreisiin ja ajankohtaisiin tutkimuksiin, jotka heijastavat nykyistä tilannetta ja haasteita. Rajauksella pyritään myös hallitsemaan katsaukseen valitun aineiston määrää. Se tekee kirjallisuuskatsauksen toteuttamisesta käytännöllisempää ja tehokkaampaa. Aikarajauksen loppuosa 16.4.2025 oli päivämäärä, johon materiaalin haun ajankohtana relevanttien ja vertaisarvoitujen hakutulosten osumat rajoituivat. Koska katsauksessa käytettiin vain vertaisarvioitua materiaalia, valikoitui kyseinen päivämäärä myös haun takarajaksi.

Katsauksessa käytetään myös kielirajauksia. Siihen valittiin kieliksi suomi ja englanti. Merkittävin syy kielirajaukseen on tekijän oma kielitaito, joka rajoittui näihin kahteen kieleen. Englanninkielinen aineisto tarjoaa laajan kansainvälisen näkökulman tekoälyn tietoturvaan, kun taas suomenkieliset aineistot olisivat voineet tuoda esiin paikallista näkökulmaa ja kansallisia erityispiirteitä. Suomenkielistä aineistoa ei kuitenkaan katsaukseen lopulta valikoitunut mukaan juuri ollenkaan. Materiaalia läpikäydessä ne rajautuivat pois, koska eivät olleet tutkimuskysymyksen kannalta relevantteja.

3.1.4 Toteutus

Katsauksen aluksi tietokannasta tehtyjen hakujen tulokset käytiin yksitellen läpi ja arvioitiin hakutuloksen olennaisuus tutkimuskysymyksen suhteen. Tämä ensimmäinen arviointi tehtiin julkaisun otsikon ja hakutuloksessa näkyvän lyhyen kuvauksen sekä julkaisuun liitettyjen avainsanojen perusteella. Jos julkaisu vaikutti tarkastelun perusteella

tutkimuskysymyksen kannalta relevantilta, se otettiin talteen myöhempää tarkempaa tarkastelua varten. Tarkempaan tarkasteluun päätyi lopulta 67 hakutulosta.

Seuraavassa vaiheessa käytiin läpi kaikki aikaisemman hakutulosarvioinnin yhteydessä talteen otetut julkaisut. Jokaisen julkaisun sisältöön tutustuttiin tarkemmin julkaisusta tehdyn tiivistelmän perusteella. Lisäksi arvioitiin, oliko julkaisun vertaisarviointi luotettava ja, oliko julkaisun tehnyt lehti indeksoitu. Arvioinnissa käytettiin apuna tekoälyä. Hakutuloksen sisällön olennaisuuden arviointi suhteessa tutkimuskysymykseen perustui julkaisusta tehtyyn tiivistelmään, joka tehtiin tekoälyn avulla. Tekoälyn tekemän tiivistelmän lisäksi julkaisujen sisältöä arvioitiin myös jokaisen julkaisun alussa olevan tiivistelmän perusteella. Näiden kahden tiivistelmän muodostaman kokonaisuuden pohjalta tehtiin lopullinen arvio kyseisen hakutuloksen olennaisuudesta suhteessa tutkimuskysymykseen.

Tekoälyä käytettiin myös hakutulosten vertaisarviointien tarkistamisessa. Tekoälyltä kysyttiin julkaisun DOI-numeron perusteella, onko julkaisu vertaisarvoitu ja onko sen julkaissut lehti indeksoitu Scopuksessa tai Web of Sciencessä. Tekoälyn antama tulos tarkistettiin manuaalisesti Scopuksen ja Web of Sciencen omista listauksista. Tällä varmistettiin katsauksessa käytettyjen julkaisujen vertaisarvioinnin laatu.

Tarkemmat tiedot tekoälylle annetuista syötteistä löytyvät työn alkuosasta heti tiivistelmien jälkeen. Yllä mainittujen kriteerien perusteella rajattiin hakujen tuloksista ne, jotka hyväksyttiin mukaan analyysiin. Analyysissä käytettiin lopulta 31 julkaisua. Analyysissä käytetty aineisto on listattu kokonaisuudessaan työn lopussa liitteessä I.

Katsauksen aluksi koko materiaali käytiin huolellisesti lukemalla läpi ja tunnistettiin sekä kirjattiin ylös kaikki mainitut riskit. Lisäksi kirjattiin ylös materiaalissa mainitut kuhunkin riskiin sopivat riskienhallintakeinot. Läpikäynnin yhteydessä keskeiset, tutkimuskysymykseen liittyvät kohdat merkittiin ylös, jotta niihin voitaisiin tarpeen mukaan palata. Kun aineistosta oli tunnistettu kaikki tekoälyn mainitut riskit, ne luokiteltiin ryhmiin, jotta aineiston jäsentäminen olisi helpompaa. Ryhmittelyn yhteydessä myös verrattiin eri aineistoista tehtyjä huomioita keskenään päällekkäisyyksien ja samankaltaisuuksien sekä eroavaisuuksien huomaamiseksi. Lopuksi kaikista mainituista riskeistä koostettiin ryhmät riskien samankaltaisuuden perusteella. Ryhmiä oli lopuksi yhteensä 15. Ryhmien esittely löytyy tästä katsauksesta Tulokset-luvusta. Mukana pidettiin myös muutamia sellaisia riskiryhmiä, jotka eivät olleet tutkimuskysymyksen kannalta olennaisia, kuten yhteiskunnalliset riskit-ryhmä. Ryhmän riskit eivät suoraan liittyneet tietoturvaan, mutta se haluttiin pitää mukana analyysin kokonaiskuvan yhtenäisyyden säilyttämiseksi.

Analyysin seuraavassa vaiheessa ryhdyttiin selvittämään eri riskiryhmien yleisyyttä. Aineisto tarkasteltiin uudestaan läpi ja jokaiselle riskiryhmälle suoritettiin esiintyvyysslaskenta. Kaikki aineistossa mainitut riskit huomioitiin laskennassa. Tällä tavalla laskennalla saatiin kattava kuva siitä, mitkä riskit ovat yleisimpiä tekoälyn tietoturvan näkökulmasta. Laskennan jälkeen esiintymistiheyden perusteella voitiin tunnistaa yleisimmät riskiryhmät. Aiemmin tutkimuskysymyksen kannalta epäolennaisiksi merkityt riskiryhmät pidettiin myös laskennassa mukana kokonaiskuvan luomiseksi. Mikäli nämä riskiryhmät olisivat vääristäneet laskentaa merkittävästi, ne olisi jätetty pois. Näin ei kuitenkaan käynyt, vaan epäolennaisien riskiryhmien esiintymistiheys jäi aineistossa niin pieneksi, ettei se ollut tuloksien kannalta merkitsevää. Alun perin suunnitelma oli ottaa tuloksiin viisi eniten mainintoja saanutta riskiryhmää, mutta tasatilanteen vuoksi mukaan otettiin myös kuudes ryhmä.

Seuraavassa vaiheessa tarkasteltiin riskien hallintamenetelmiä. Kaikki materiaalissa mainitut hallintamenetelmät oli aikaisemmassa vaiheessa kirjattu ylös ja merkitty, millaisiin riskeihin hallintamenetelmää käytettiin. Aiempien kirjausten perusteella materiaalista poimittiin riskiryhmäkohtaisesti kaikki ne riskienhallintamenetelmät, jotka sopivat kuuden yleisimmän riskiryhmän riskienhallintaan. Yhdelle riskiryhmälle, Kyberhyökkäyksissä käyttö -ryhmälle, ei ollut materiaalissa mainittu lainkaan erikseen siihen kohdistettavia riskienhallintamenetelmiä. Riskiryhmäkohtaisen käsittelyn lisäksi kerättiin materiaalista yleisiä riskienhallintamenetelmiä, jotka sopivat useamman riskiryhmän riskienhallintaan. Nämä riskienhallintamenetelmät kattoivat myös tekoälyn käytön kyberhyökkäyksien tukena tai kyberhyökkäyksissä.

3.2 EU:n tekoälyasetuksen vaatimusten analyysi

3.2.1 Analyysin rajaus

Analyysi keskittyi ainoastaan tekoälyasetuksen tietoturva-vaatimuksiin, jotka kohdistuvat tekoälyjärjestelmiin ja käyttäjäorganisaatioihin. Asetuksen taloudellisia, yhteiskunnallisia tai muita laaja-alaisia vaikutuksia ei käsitelty, koska analyysissä tavoitteena oli rajata tarkastelu ainoastaan tietoturvaan liittyviin näkökulmiin. Rajaus tehtiin tarkoituksenmukaisuuden takia, jotta voitiin keskittyä tutkimuskysymyksen kannalta olennaisiin osiin asetusta ja niiden vaikutuksiin tekoälyjärjestelmien tarjoajille ja käyttäjille.

3.2.2 Lähteet

Analyysissä ensisijaisena lähteenä käytettiin Euroopan Unionin tekoälyasetusta, joka on saatavilla Eur-Lex -verkkosivustolla (2). Asetus oli analyysin keskeisin asiakirja, koska siinä määritetään tekoälyjärjestelmiin liittyvät säännöt ja vaatimukset.

Toissijaisina lähteinä käytettiin laajasti Euroopan komission virallisilla verkkosivuilla julkaistuja ohjeistuksia (39), uutisia (38) sekä lehdistötiedotteita asetukseen liittyen (37).

Euroopan komission verkkosivuilla julkaistu materiaali rajautui toissijaiseksi lähteeksi, koska sivustolla on sekä ensikäden tietoa tekoälyasetuksen tavoitteista ja sisällöstä että sen täytäntöönpanon yksityiskohdista. Sivuston tieto on luotettavaa, virallista ja ajantasaista. Komission sivuilla on lisäksi saatavilla asetukseen liittyvää taustamateriaalia sekä muuta kontekstiin liittyvää aineistoa ja linkkejä. Ne auttavat ymmärtämään asetuksen syntyprosessia ja poliittisia linjauksia. Komission julkaisema materiaali antaa laajan kuvan asetuksen vaikutuksista ja soveltamisesta, ja helpottaa näin asetuksen analysoimista.

Analyysiä varten komission verkkosivuilta etsittiin tietoa erityisesti tekoälyyn ja tekoälyasetukseen keskittyviltä verkkosivustoilta. Tämän lisäksi käytiin läpi komission uutiset ajalta 1.1.2020-30.4.2025. Aikajakso kattaa sekä tekoälyasetuksen valmistelun että asetuksen voimaantulon jälkeisen ajan, ja mukaan mahtuvat myös uusimmat päivitykset ja uutiset. Hakua rajattiin myös hakusanalla ”tekoäly”. Haulla löytyi 196 hakutulosta. Tulokset käytiin läpi yksitellen ja arvioitiin, ovatko ne ajankohtaisia ja relevantteja tutkimuskysymyksen kannalta. Mukaan otettiin vain analyysin näkökulmasta merkitykselliset, ajantasaiset ja ajankohtaiset uutiset ja lehdistötiedotteet. Analyysissä käytetty aineisto on listattu kokonaisuudessaan työn lopussa liitteessä I.

3.2.3 Toteutus

Analyysin alkuvaiheessa tekoälyasetus ja muu aineisto käytiin läpi yleiskuvan luomiseksi ja sääntelyn ymmärtämiseksi. Tarkastelun jälkeen määriteltiin analyysin rajaukset, jotka ohjasivat jatkotoimia ja analyysin painopistettä. Koska tekoälyasetus on lähestymistavaltaan riskiperusteinen ja tarkastelun kohteena olivat tekoälyjärjestelmille asetetut vaatimukset, analyysissä keskityttiin pääasiassa suuririskisten tekoälyjärjestelmien sääntelyyn liittyviin kohtiin. Lisäksi analyysissä huomioitiin vähäisen riskin tekoälyjärjestelmille asetetut velvoitteet ja vapaaehtoiset käytänteet, koska ne tarjoavat vertailukohtia suuririskisiin tekoälyjärjestelmiin ja täydentävät kokonaiskuvan sääntelyn kattavuudesta. Hyvän kokonaiskuvan varmistamiseksi tarkasteluun sisällytettiin myös käyttäjäorganisaatioita koskevat vaatimukset. Näillä rajauksilla pyrittiin luomaan selkeä ja laaja kuva tekoälyasetuksen vaatimuksista ja niiden vaikutuksista.

Analyysin rajauksen jälkeen tekoälyasetus käytiin uudestaan läpi perusteellisesti, ja samalla merkittiin taulukkoon kaikki ne kohdat, jotka olivat tutkimuskysymyksen kannalta olennaisia. Taulukossa (taulukko 2) esitettiin selkeästi asetuksen kohta ja kirjattiin ylös kohdan sisältö. Tällä systemaattisella lähestymistavalla varmistettiin, että analyysi

kattoi kaikki olennaiset osa-alueet asetuksesta, ja analyysiin käytettävissä ollut aika hyödynnettiin mahdollisimman tehokkaasti.

Tekoälyasetuksen kohta	Sisältö
Artikla 8	Vaatimustenmukaisuus
Artikla 9	Riskienhallintajärjestelmä
Artikla 10	Data ja datanhallinta; koulutus-, validointi ja testausdatan laatu, laajuus ja edustavuus.
Artikla 11	Tekninen dokumentaatio. Tarkemmin asetuksen liitteessä IV.
Artikla 12	Lokitietojen säilyttäminen, tapahtumien jäljitettävyys.
Artikla 13	Avoimuus, läpinäkyvyys, käyttöohjeet
Artikla 14	Ihmisen suorittama valvonta
Artikla 15	Suoriutuminen, vikasietoisuus ja kyberturva
Artikla 16	Suuririskisen tekoälyjärjestelmän tarjoajan velvollisuudet
Artikla 17	Strategia, jolla varmistetaan vaatimusten noudattaminen ns. laadunhallintajärjestelmä.
Artikla 18	Vaaditut dokumentaatiot ja säilytys ajat
Artikla 19	Automaattisesti tuotettavien lokitietojen säilytys ja säilytysajat
Artikla 26	Käyttöönottajien velvollisuudet
Artikla 27	Perusoikeusvaikutusten arviointi
Artikla 48	Suuririskisten tekoälyjärjestelmien CE-merkintä
Artikla 49	Rekisteröintivelvoite
Artikla 95	Tiettyjen vaatimusten vapaaehtoista soveltamista koskevat käytäntösäännöt

Taulukko 2. *Tekoälyasetuksen kohdat, jotka merkittiin katsauksen näkökulmasta olennaisiksi.*

Lisäksi hyvän yleiskuvan luomiseksi huomioitiin myös muita keskeisiä kohtia tekoälyasetuksesta. Nämä artikkelit käsittelevät muun muassa tekoälyjärjestelmien luokittelua ja velvoitteita, jotka koskevat muita kuin tekoälyjärjestelmiä, niiden tarjoajia tai käyttäjäorganisaatioita, sekä sääntelyn valvontaan ja seuraamuksia koskevia määräyksiä. Nämä kohdat on listattu taulukossa 3.

Tekoälyasetuksen kohta	Sisältö
Alkuteksti kohta 165	Muiden kuin suuririskisten tekoälyjärjestelmien vaatimusten vapaaehtoisuus
Artikla 5	Kielletyt tekoälyyn liittyvät käytännöt
Artikla 6	Suuririskisten tekoälyjärjestelmien luokittelusäännöt
Artikla 20	Korjaavat toimenpiteet ja ilmoitusvelvollisuus
Artikla 21	Yhteistyö viranomaisten kanssa
Artikla 23	Maahantuojaisten velvollisuudet
Artikla 24	Jakelijoiden velvollisuudet
Artikla 51	Yleiskäyttöisten tekoälymallien luokittelu
Artikla 53	Yleiskäyttöisten tekoälymallien tarjoajien velvoitteet
Artikla 72	Markkinoille saattamisen jälkeinen seuranta
Artikla 99	Seuraamukset

Taulukko 3. *Katsauksessa kokonais kuvan luontiin käytetyt tekoälyasetuksen kohdat.*

Kun asetus oli jäsennetty, käytiin perusteellisesti läpi kaikki taulukkoon listatut kohdat ja kirjattiin ylös niissä esitetyt vaatimukset suuririskisille tekoälyjärjestelmille, tekoälyjärjestelmien tarjoajille ja käyttäjäorganisaatioille. Lisäksi käytiin läpi tarkemmin vaatimusten vapaaehtoista soveltamista koskevat käytännesäännöt, jotka koskevat muita kuin suuririskisiä tekoälyjärjestelmiä. Tässä vaiheessa ne päätettiin jättää analyysistä pois, koska ne eivät olleet relevantteja tutkimuskysymyksen näkökulmasta, vaan enemmän suosituksia. Samoin analyysin ulkopuolelle päädyttiin lopulta rajaamaan

yleiskäyttöisten tekoälymallien vaatimukset, koska niitä koskevien käytännönsääntöjen laadinta on vielä kesken (13).

Asetuksen läpikäynnin jälkeen muu aineisto käytiin uudestaan huolellisesti läpi. Lisämateriaali täydensi lähinnä asetuksen läpikäynnissä vähemmälle huomiolle jääneitä osioita. Täydennykset koskivat erityisesti kiellettyjä tekoälyn käyttötapauksia. Lisäksi vähäisen riskin, minimaalisen tai ei lainkaan riskiä sisältävien tekoälyjärjestelmien määritelmiä tarkennettiin sekä tarkistettiin vähäisen riskin tekoälyjärjestelmiä koskevia julkistamisvelvoitteita.

Analyysin viimeisessä vaiheessa tehtiin kevyet tapausanalyysit kahteen kuvitteelliseen suuren riskin tekoälyjärjestelmään. Tapausanalyysien avulla havainnollistettiin ja konkretisoitiin tekoälyasetuksen vaikutuksia järjestelmien tietoturvaan ja kehitystyöhön. Tarkasteluun valittiin kaksi melko erilaista järjestelmää, joilla kummallakin oli keskeinen rooli oman toimintaympäristönsä päätöksissä.

Ensimmäinen tapausanalyysin kohde oli pankkisektorin tekoälyjärjestelmä, jota käytettiin luotonantopäätöksien tekemisessä. Tämän järjestelmän osalta kuvattiin järjestelmä lyhyesti ja analysoitiin sen riskiluokitus tekoälyasetuksen määritelmän mukaisesti. Lisäksi arvioitiin, minkälaisia tietoturvavaatimuksia järjestelmään kohdistuu sen toimintaympäristössä ja mitä vaatimuksia tulee huomioida järjestelmän kehitystyössä.

Toisena tapausanalyysin kohteena oli autonominen ajoneuvo. Tässäkin tapauksessa analyysi aloitettiin järjestelmän kuvaamisella ja tekoälyasetuksen mukaisella riskiluokituksella. Tämän jälkeen tarkasteltiin tekoälyasetuksen mukaisia tietoturvaan ja kehitystyöhön liittyviä vaatimuksia, huomioiden että kyseessä oli järjestelmä, jonka toiminnan on oltava ennakoitavaa, vakaata ja turvallista kaikissa tilanteissa.

Tapausanalyysin järjestelmät olivat kuvitteellisia, ja ne valikoitiin tekijän oman kiinnostuksen mukaan. Tavoitteena oli tuoda konkretiaa tekoälyasetuksen vaatimuksiin erityisesti suuririskisten järjestelmien kohdalla ja syventää ymmärrystä tekoälyasetuksen vaatimusten soveltamisesta käytäntöön. Kahden hyvin erilaisen tekoälyjärjestelmän valinta myös havainnollisti asetuksen vaikutuksia hyvin erilaisissa toimintaympäristöissä.

4. TULOKSET

4.1 Tekoälyn yleisimmät riskit

Katsaukseen sisältyi yhteensä 31 julkaisua, jotka käsittelivät tekoälyn tietoturvaa ja sen riskejä. Julkaisut ajoittuvat vuosille 2020–2025. Julkaisuissa esiin nostettujen riskien yleiskuva pysyi melko samanlaisena läpi materiaalin riippumatta siitä, oliko kyseessä julkaisu tarkastelujakson alusta vai lopusta. Esimerkiksi geopoliittisen tilanteen muutokset tai generatiivisen tekoälyn nopea kehitys parin viime vuoden aikana eivät nousseet esiin aineistossa merkittävästi.

Kuten aiemmin menetelmät-luvussa kerrottiin, tutkimusmateriaalissa esiin nousseet riskit jaettiin 15 ryhmään kokonaisuuden jäsentämiseksi ja analyysia varten. Alla esitellään taulukossa 4 kootusti nämä 15 ryhmää sekä niiden kuvaukset.

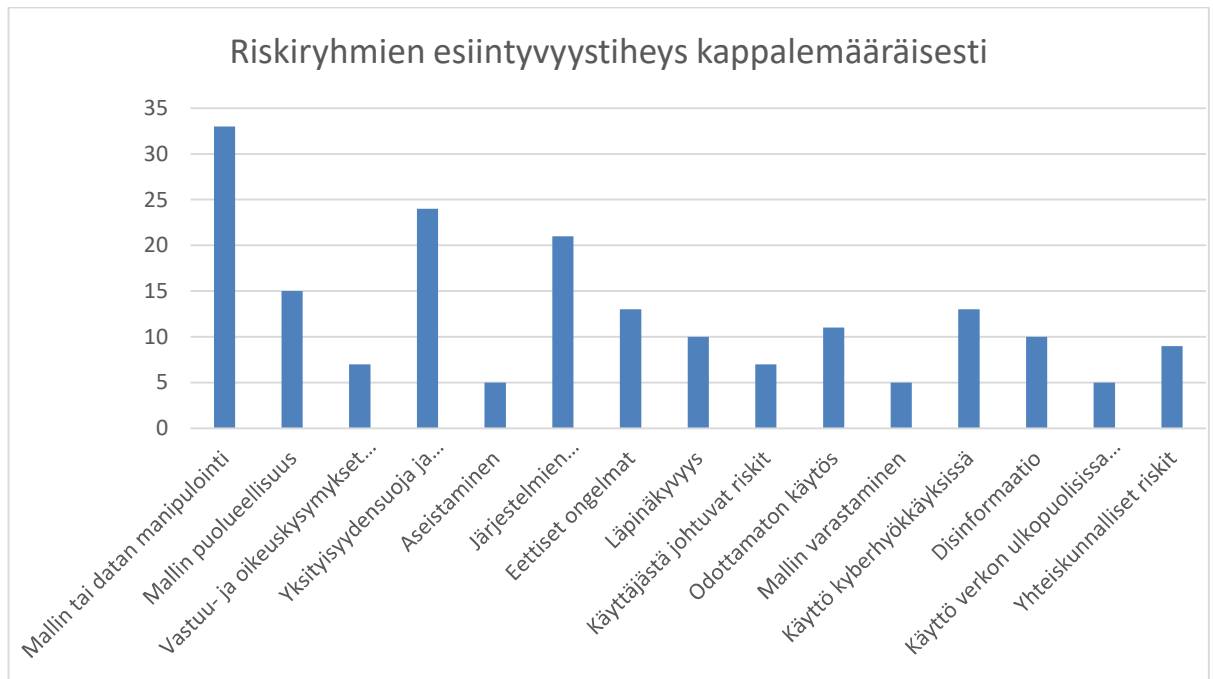
Riskiryhmittelyn nimi	Ryhmän kuvaus
Mallin tai datan manipulointi	Tähän ryhmään kuuluvat kaikki dataan tai tekoälymalliin kohdistuvat tahallisen manipuloinnin riskit, kuten adversaariset hyökkäykset, palkitsemisjärjestelmän manipulointi, syötteiden manipulointi, mallin tai datan myrkyttäminen jne.
Yksityisyyden-suoja ja tietosuoja	Tähän ryhmään kuuluvat kaikki yksityisyydensuojaa tai tietosuojaa uhkaavat riskit, esimerkiksi luvaton valvonta, henkilötietojen luvaton käyttö, henkilötietojen vuotaminen, epäeettinen profilointi jne.
Järjestelmien tietoturva-vaavoittuvuudet	Tähän ryhmään kuuluivat riskit, jotka juontavat juurensa tekoälyjärjestelmässä olevaan haavoittuvuuteen ja sen hyödyntämiseen. Myös monet ns. perinteisesti tietojärjestelmien riskeihin luetut riskit kuten tietomurrot ja DDoS kuuluvat tähän ryhmään. Muita esimerkkejä ryhmän riskeistä ovat takaportti-iskut, prompt-injektiot, SQL-injektiohyökkäykset ja zero-day-haavoittuvuudet.
Mallin puolueellisuus	Tähän ryhmään kuuluvat riskit, jotka aiheutuvat tahattomasti mutta johtavat mallin vinoutuneeseen tai puolueelliseen käytökseen. Esimerkkejä tällaisista riskeistä ovat huonolaatuisten datajoukkojen käyttö, mallin koodausvirheet, palkkioiden virheellinen määrittely, riittämätön testaus ja validointi jne.

Käyttö kyberhyökkäyksissä	Tähän ryhmään kuuluvat riskit, joissa tekoälyä käytetään kyberhyökkäyksessä joko tehostamaan hyökkäystä tai luodaan kokonaan uusi hyökkäystapa, jossa tekoäly suorittaa hyökkäyksen. Esimerkkejä tämän ryhmä riskeistä ovat mm. tekoälyn käyttö haavoittuvuuksien skannauksissa, generatiivisen tekoälyn käyttö haittaohjelmien levittämisessä, tekoälyllä tehdyt kalasteluviestit ja ääniviestit jne.
Ihmisoikeus- ja eettiset riskit	Tähän ryhmään kuuluvat riskit, jotka liittyvät ihmisoikeuksien loukkauksiin, epäeettisiin toimintatapoihin ja kansainvälisten sopimusten rikkomuksiin.
Oikeudelliset haasteet ja tapakulttuuri	Tämän ryhmän riskit koskivat lainsäädännön puutteellisuutta, vastuunkantokysymyksiä, kansainvälisien lakien rikkomuksia, tekijänoikeuksien rikkomuksia ja sosiaalisten normien rikkomuksia.
Aseistaminen	Tähän ryhmään kuuluvat riskit, jotka koskivat terrorismia, bioterrorismia, autonomisten aseiden käyttöä ja tekoälyn hyödyntämistä sota- ja aseiteollisuudessa.
Disinformaatio	Tämän ryhmän riskit koskivat tekoälyllä mielipidevaikuttamisen tai muun manipuloinnin tarkoituksiin luotua sisältöä, kuten väärennökset, deepfake-videot, valeuutiset, synteettinen sisältö, manipuloivat mainokset, mielipidevaikuttamisen tarkoituksiin luotu sisältö jne.
Läpinäkyvyys	Riskit, jotka kuuluvat tähän ryhmään koskevat tekoälyn läpinäkyvyyttä ja päätösten selitettävyyden puutteita, päätösten ymmärtämistä, tekoälymallien monimutkaisuutta ja tulosten jäljitettävyyttä.
Käyttö verkon ulkopuolisissa rikoksissa	Tämän ryhmän riskit koskivat tekoälyn käyttöä tukemaan digitaalisen maailman ulkopuolella tapahtuvia rikoksia. Esimerkkejä tällaisista rikoksista olivat väärennökset, identiteettivarkaudet ja petokset, joissa oli apuna käytetty tekoälyä.
Käyttäjistä johtuvat riskit	Ryhmään kuuluvat käyttäjien virheistä, taidoista tai toimintatavoista seuraavat riskit kuten osaamisen puute, liika luottamus tuloksiin, riippuvuus tekoälystä ja tekoälyn tarkoituksen vastainen käyttö.

Odottamaton käytös	Tähän ryhmään kerättiin riskit, jotka koskivat tekoälyn odottamatonta käytöstä kuten hallusinointi, sulkemisvaikeudet, tekoälyn autonomia ja virhetilanteista johtuva odottamaton käytös.
Mallin varastaminen	Tämän ryhmän riskit koskivat tekoälymallin varastamista tai luvattonta kopiointia eri tavoin, esimerkiksi käänteistä suunnittelua hyödyntämällä.
Yhteiskunnalliset riskit	Ryhmään kuuluvat sellaiset riskit, joita voi yhteiskunnassa ilmetä tekoälyn käytön laajetessa. Esimerkiksi työttömyys, eriarvoisuuden lisääntyminen, ympäristöriskit, demokratian heikkeneminen, kriittisen infrastruktuurin häiriöt, yhteiskunnan epävakaus jne.

Taulukko 4. Työssä kootut riskiryhmittelyt ja niiden kuvaukset

Varsinaisessa analyysissä laskettiin kuhunkin ryhmään kuuluvien riskien esiintyvyyksiheys materiaalissa. Esiintyvyyksiheyden pohjalta todettiin, minkä ryhmän riskit olivat yleisimpiä katsauksen materiaalin perusteella. Tässä yhteydessä käsite ”yleisin” määriteltiin merkitykseen määrällisesti eniten mainituiksi.



Kuva 1. Työssä käytettyjen riskiryhmittelyn riskien esiintyvyyksiheys katsauksen materiaalissa.

Suoritettuna tiheyslaskennan perusteella viisi yleisintä riskikokonaisuutta olivat mallin ja datan manipulointiin liittyvät riskit, yksityisyysdenuojoan ja tietosuojaan liittyvät riskit, tekoälyjärjestelmien tietoturvaavaoittuvuuksiin liittyvät riskit, tekoälymallin

puolueellisuuteen liittyvät riskit, riskit, jotka liittyvät tekoälyn käyttöön kyberhyökkäyksissä, ja tekoälyn käyttöön liittyvät ihmisoikeus- ja eettiset riskit. Kuva 1 visualisoi tiheyslaskennan tulokset jokaisen ryhmän osalta. Taulukossa 5 on esitetty kärkeijoukon esiintyvyyksiä materiaalissa ja niiden yleisyys suurimmasta pienimpään. Tasatilanteen takia viides sija on jaettu kahdelle ryhmälle.

Yleisyys	Ryhmä	Esiintyvyyksiä
1	Mallin tai datan manipulointi	33 kpl
2	Yksityisyys- ja tietosuojat	24 kpl
3	Järjestelmien tietoturva- haavoittuvuudet	21 kpl
4	Mallin puolueellisuus	15 kpl
5	Käyttö kyberhyökkäyksissä	13 kpl
5	Ihmisoikeus- ja eettiset riskit	13 kpl

Taulukko 5. *Analyysin perusteella kuusi yleisintä riskien ryhmää*

Datan ja mallin manipulointiin liittyvien riskien korostuminen aineistossa näin paljon yllätti hieman, vaikka ne nousivat esiin materiaalissa todella usein. Materiaalissa esiintyi sekä koulutus, että käyttövaiheen hyökkäyksiä, kuten datan myrkytys ja adversaariset hyökkäykset. Manipuloinnin seurauksena tekoälyn päätökset voivat vinoutua tai muuttua virheellisiksi (14, 15). Tämän seurauksena luottamus tekoälymallin toimintaan heikenee. Manipulointia tekoälyä voidaan käyttää myös muihin tarkoituksiin. Sillä voidaan esimerkiksi levittää väärää tietoa tai manipuloida käyttäjää (15, 16).

Yksityisyys- ja tietosuojan liittyvien riskien yleisyys ei ollut yllättävää. Tekoälyjärjestelmät käsittelevät paljon tietoa, ja usein tähän sisältyy myös henkilötietoa (17). Tekoäly myös mahdollistaa ihmisten aiempaa laajemman seurannan, valvonnan ja profiloinnin (18, 19). Mikäli tietosuojat-asetuksen vaatimuksia ja henkeä ei kunnioiteta, voivat yksilön oikeudet vaarantua.

Kolmanneksi yleisimpiin riskeihin kuuluivat tekoälyjärjestelmien haavoittuvuuksiin liittyvät riskit. Tekoälyjärjestelmät ovat järjestelmiä siinä, missä muutkin teknologiset järjestelmät, ja niihin saattaa sisältyä ohjelmointi- tai konfigurointivirheitä, jotka altistavat ne hyökkäyksille (20). Esimerkiksi tietomurto tai järjestelmän kaappaaminen voivat johtaa vakaviin seurauksiin. Haavoittuvuuksien hyödyntämisestä voi seurata esimerkiksi

tekoälyn odottamatonta toimintaa, tietosuojaan ja yksityisyydensuojaan liittyvien riskien realisoituminen tai mallin tai datan manipulointi (15, 21, 22).

Kolme kärkiryhmää erottui selkeästi muusta aineistosta, kun taas seuraavat kolme riskiryhmää eivät enää erottuneet yhtä merkittävästi esiintyvyytiheydeltään. Neljänneksi nousivat mallin puolueellisuuteen liittyvät riskit. Mallin puolueellisuus sisälsi riskejä, jotka johtivat tekoälyn puolueelliseen käytökseen nimenomaan tahattomasti. Yleensä taustalla näissä riskeissä oli inhimillinen virhe tai osaamisen puute. Esimerkiksi mallissa esiintynyt virhe tai epäedustavan datajoukon käyttö olivat tällaisia riskejä (23, 24). Mallin puolueellisuudesta kuitenkin seuraa syrjiviä päätöksiä, mikä heikentää luottamusta tekoälyn toimintaan ja laskee sen käytettävyyttä (25).

Jaetulle viidennelle sijalle ylsivät riskit, jotka liittyvät tekoälyn käyttöön kyberhyökkäyksissä ja ihmisoikeus- ja eettiset riskit. Kyberhyökkäyksien riskeissä nousivat esiin hyökkäykset, joissa tekoälyllä oli hyökkäystä avustava tai tehostava rooli, esimerkiksi haavoittuvuuksien skannauksessa tai bottiverkon hallinnassa (14, 19, 24). Kyberhyökkäykseen liittyvien riskien realisoitumisesta voi aiheutua taloudellista vahinkoa ja muiden riskien, kuten tietovuodon tai tietosuojariskien realisoituminen (26).

Toinen viitossijan jakava riskiryhmä oli ihmisoikeus- ja eettiset riskit. Sen riskit kohdistuivat tekoälyn käytön epäeettisiin toimintatapoihin ja tekoälyn käytön seurauksena mahdollisesti tapahtuviin ihmisoikeusrikkomuksiin tai kansainvälisten sopimusten rikkomuksiin. Näiden riskien realisoituessa seurauksena voi aiheutua vahinkoa yksilölle tai yhteiskunnalle. Toisaalta seuraukset voivat olla myös taloudellisia (23, 27, 28).

Muista riskeistä tekoälyn läpinäkyvyyteen, disinformaatioon ja synteettiseen sisältöön liittyvät riskit ja tekoälyn odottamattomaan käytökseen liitetyt riskit olivat esiintymistiheydeltään keskitasoa. Nykyisestä geopoliittisesta tilanteesta huolimatta, tekoälyn aseistaminen oli esiintymistiheyden perusteella riskien yleisyyden häntäpäässä yhdessä tekoälymallin varastamiseen liittyvien riskien ja tekoälyn hyödyntämisen fyysisen maailman rikoksissa, kuten petoksissa, kanssa.

Katsauksen analyysin perusteella voidaan todeta, että tekoälyn tietoturvariskit ovat monimuotoisia ja usein toisiinsa kytkeytyviä. Esimerkiksi tietoturva haavoittuvuudet voivat aiheuttaa tietovuodon ja tietosuojarikkomuksen, mahdollistaa datan tai mallin manipuloinnin, tai mallin puolueellisuus voi johtaa eettisiin ja oikeudellisiin seurauksiin (14, 15, 17). Riskien hallinta vaatii lähestymistapaa, jossa käytetään teknisiä ratkaisuja, luodaan eettisiä periaatteita ja organisaatiotason käytäntöjä, sekä noudatetaan lainsäädännöstä tulevia velvoitteita (18, 20, 26).

4.1.1 Tekoälyn riskien hallinta

Katsauksessa tutkimusmateriaalista kerättiin riskien lisäksi myös riskeihin ehdotetut hallintakeinot. Katsauksen tulosten perusteella kuuden yleisimmän riskijoukon riskeihin suunnatut hallintakeinot on tässä luvussa koottu yhteen riskikohtaisesti. Luvun lopussa käsitellään myös hallintakeinoja, jotka suuntautuvat useamman kuin yhden yleisen riskin hallintaan.

Mallin manipulointiin liittyviin riskeihin oli tarjolla useita ratkaisuja. Ne voidaan jakaa kahteen osaan: sellaisiin riskienhallintakeinoihin, joilla pyritään estämään mallin manipulointi, ja niihin keinoihin, joilla pyritään huomaamaan mallin manipulointi. Mallin manipuloinnin estämiseen tähtääviä keinoja ovat esimerkiksi mallin robusti koulutus, joka parantaa mallin kykyä käsitellä häiriöitä (24, 29). Syötteiden validointi ja poikkeavien syötteiden tunnistus vähentävät syötteiden kautta tapahtuvan manipuloinnin toimivuutta (1,16). Ei-toivotun tai haitallisen sisällön tuottamisen esto, nimensä mukaisesti, estää mallia tuottamasta haitalliseksi määriteltyä tai ei-toivottua sisältöä (24). Regularisointitekniikoiden käyttö ohjaa mallin toimintaa tasapainoiseen suuntaan: liiasta monimutkaisuudesta rangaistaan ja tarkkuudesta palkitaan (1, 16, 30). Mallin manipuloinnin riskiä voi myös pienentää kouluttamalla mallin säännöllisesti uudestaan (16).

Riskienhallintakeinot, joilla tähdätään manipuloinnin huomaamiseen, ovat samoja, olipa kyse sitten mallin manipuloinnista tai datan manipuloinnista. Mallin tulee olla niin läpinäkyvä ja selitettävä, että käyttäjä pystyy tulkitsemaan ja ymmärtämään tuotoksen. Näin myös tekoälyn toiminnan seuranta ja valvonta on mahdollista (16, 22, 23, 31). Seurantaan kuuluu myös lokitietojen keräys. Lokitiedot ovat valvonnan työkalu. Tekoälyn toimintaa voi myös säännöllisesti auditoida, jotta mahdolliset poikkeamat mallin toiminnassa ja datassa huomattaisiin (14, 23, 31, 32).

Datan manipuloinnin estämiseen on myös tarjolla työkaluja. Tärkeimpänä niistä on käytettyjen datajoukkojen laadun ja eheyden tarkistus ennen niiden käyttöä, samoin kuin datan puhdistus –molemmat pienentävät datan myrkyttämisen riskiä (14, 16, 24, 29). Koulutus- ja tuotantodatan erottaminen ei itsessään poista datan manipuloinnin riskiä, mutta vähentää kontaminaation mahdollisuutta (30).

Tietosuojaan ja yksityisyydensuojaan liittyvien riskienhallintaan on useita keinoja. EU:n tietosuoja-asetus rajoittaa henkilötietojen käyttöä ja vaatii, että käytölle on peruste. Lisäksi uusien teknologioiden, kuten tekoälyn kohdalla, tulee ennen käyttöönottoa tehdä tietosuojavaikutusten arviointi järjestelmälle (23). Laista tulevan sääntelyn lisäksi voidaan riskejä pienentää myös tietojen anonymisoinnilla tai pseudonymisoinnilla sekä

henkilötietojen minimoinnilla (15, 17, 31). Lisäksi tekniset toimenpiteet, kuten tietojen salaaminen ja pääsynvalvonta pienentävät riskejä (15, 31).

Tekoälyjärjestelmien haavoittuvuuksien kohdalla pätevät samat keinot kuin muidenkin järjestelmien riskienhallinnassa ja suojaamisessa. Jo suunnittelu ja kehitysvaiheessa security by design -periaatteen noudattaminen parantaa järjestelmän tietoturvaa (1, 14). Lisäksi säännöllinen auditointi ja tekninen valvonta pienentävät huomaamatta jääneiden haavoittuvuuksien riskiä (15, 22). Perinteiset suojauskeinot, kuten pääsynvalvonta, tietojen salaus, päivitysten ajantasaisuus ja lokitietojen tallennus ja seuranta toimivat myös tekoälyjärjestelmien riskien kohdalla (15, 16, 17, 22). Samoin säännöllinen riskien arviointi ja niiden hallinnan suunnittelu ja toteutus on vanha ja tässäkin toimiva tapa hallita ja pienentää riskejä (33, 34).

Tekoälyn ihmisoikeus- ja eettiset riskit -ryhmään kuuluviin riskeihin ei materiaalissa ollut tarjolla kovin monia ratkaisuja. Esiin nousivat kuitenkin eettisten standardien määrittäminen, eettiset periaatteet ja ohjeet tekoälyn käyttöön ja selkeä oikeudellisten vastuiden määrittely (18, 23, 26, 27).

Ainoastaan riskeille, jotka liittyvät tekoälyn käyttöön kyberhyökkäyksissä ei tutkimusmateriaalista löytynyt erikseen niihin suunnattuja hallintakeinoja. Tekoälyjärjestelmien suojaamiseen suunnatut toimet toimivat toki myös tekoälyn tukemien tai tehostamien kyberhyökkäyksien aiheuttamaa uhkaa vastaan.

Muita tutkimusmateriaalissa mainittuja riskienhallintakeinoja olivat käyttäjien kouluttaminen, jotta he osaavat käyttää tekoälyjärjestelmiä oikein ja eettisesti sekä pystyvät mahdollisesti huomaamaan poikkeamat tekoälyn toiminnassa (26). Ihmisen suorittama valvonta tekoälyn toiminnalle, jolloin huomattaisiin tekoälyn mahdolliset vinoumat, puolueellisuus tai jos tekoäly, mallin tai datan manipuloinnin takia, alkaa toimia poikkeavasti (31, 35, 36). Järjestelmien sertifiointeilla voitaisiin parantaa tekoälyjärjestelmien tietoturvan tasoa (35). Tekoälyyn kohdistuvalla sääntelyllä voitaisiin puuttua useampaankin riskiin, muun muassa yksityisyydensuoja ja tietosuojariskeihin, eettisiin ongelmiin ja tekoälyn puolueellisuuteen liittyviin riskeihin (15, 18). Samoin tekoälyjärjestelmien tarjoajien oikeudellisen vastuun lisäämisen katsotaan parantavan järjestelmien tietoturvaa ja vähentävän mallien puolueellisuuteen liittyviä riskejä (35).

4.2 EU:n tekoälyasetuksen vaatimukset

Euroopan unionin tekoälyasetus (EU) 2024/1689 perustuu riskiperusteiseen lähestymistapaan, jossa tekoälyjärjestelmät luokitellaan riskitason mukaan seuraavasti: kielletty, suuri riski, vähäinen riski ja minimaalinen riski tai ei riskiä (37). Tässä

katsauksessa tarkasteltiin erityisesti suuren riskin tekoälyjärjestelmiin ja näiden järjestelmien tarjoajiin kohdistuvia vaatimuksia. Rajausta perustui siihen, että näihin järjestelmiin kohdistuu suurin osa asetuksessa mainituista vaatimuksista. Muiden riskitasojen vaatimuksia käsiteltiin lyhyemmin, koska niihin kohdistuu vähemmän tai ei lainkaan asetuksessa mainittuja vaatimuksia.

Asetus sisältää useita vaatimuksia, jotka koskevat tekoälyjärjestelmien kehittämistä, markkinoille saattamista, käyttöönottoa ja käyttöä. Tekoälyn sääntelyn keskeinen tavoite on varmistaa tekoälyjärjestelmien turvallisuus, luotettavuus ja toimivuus sekä suojella kansalaisten perusoikeuksia Euroopan unionin alueella (38).

4.2.1 Kielletyt käyttötapaukset

Tekoälyasetuksessa on katsottu, että on olemassa tekoälyjärjestelmiä, jotka luokitellaan riskitasoon ”riski, jota ei voida hyväksyä”. Näihin kiellettyihin käyttötapauksiin katsotaan liittyvän riskejä ihmisten turvallisuudelle, toimeentulolle, kansalaisten perusoikeuksille ja Euroopan unionin arvoille. Kieltoa näiden tekoälyjärjestelmien käyttöön aletaan soveltaa 2. elokuuta 2025 alkaen (38).

Kiellettyjä tekoälyn käyttötapauksia ovat tekoälyjärjestelmät, joita käytetään haitalliseen manipulointiin tai petokseen. Tämä tarkoittaa järjestelmiä, jotka käyttävät tarkoituksellisesti manipuloiivia tai harhaanjohtavia menetelmiä tai pyrkivät vaikuttamaan alitajuntaan eri tekniikoilla. Kiellettyjen listalle kuuluvat myös tekoälyjärjestelmät, jotka haitallisesti hyödyntävät käyttäjänsä haavoittuvuutta. Näitä ovat järjestelmät, jotka käyttävät hyväkseen käyttäjän haavoittuvaa asemaa kuten ikää, vammaisuutta tai tiettyä sosiaalista tai taloudellista tilannetta. Näiden järjestelmien tavoitteena tai käytön seurauksena on käyttäytymisen vääristäminen, minkä seurauksena aiheutuu tai todennäköisesti aiheutuu merkittävää vahinkoa (39).

Tekoälyjärjestelmät, jotka tekevät sosiaalista pisteytystä, eli järjestelmät, jotka luokittelevat tai arvioivat ihmisiä tai ihmisryhmiä, luetaan kiellettyihin käyttötapauksiin. Järjestelmissä luokittelun perusteena käytetään henkilökohtaisia tai persoonallisuusominaisuuksia tai sosiaalista käyttäytymistä. Sosiaalisesta pistemäärästä seuraa syrjivää tai epäoikeudenmukaista kohtelua, kun tiedot kerätään toisistaan riippumattomista sosiaalisista konteksteista tai tällainen kohtelu on perusteetonta tai suhteetonta sosiaaliseen käyttäytymiseen nähden (39).

Kiellettyjä ovat myös käyttötapaukset, jotka mahdollistavat laajamittaisen valvonnan ilman asianmukaista oikeudellista perustaa. Tällaisia ovat tekoälyjärjestelmät, jotka

mahdollistavat yksittäisen rikoksen riskien arvioinnin tai ennustamisen perustuen pelkästään profilointiin tai persoonallisuuden piirteisiin ja ominaisuuksiin. Poikkeuksena ovat järjestelmät, joita käytetään tukemaan ihmisen tekemää arviota, joka perustuu objektiivisiin ja todennettaviin faktoihin, jotka liittyvät suoraan rikolliseen toimintaan. Kiellettyä on myös Internetin tai kameravalvontamateriaalin kohdentamaton kaavinta, jonka tarkoituksena on luoda tai laajentaa kasvojentunnistustietokantoja. Samoin julkisissa tiloissa lainvalvontatarkoituksessa toteutettu reaaliaikainen biometrinen etätunnistus on kielletty. Poikkeuksena tähän ovat tilanteet, joissa biometristä tunnistusta käytetään kohdennettuun uhrien etsintään tai tiettyjen uhkien, esimerkiksi terrori-iskujen estämiseen, tai tiettyjen rikosten kohdalla epäiltyjen etsintään (39).

Lisäksi tekoälyn käyttö ihmisten tunteiden tunnistamiseen työpaikoilla ja oppilaitoksissa on kielletty. Poikkeuksena ovat tilanteet, joissa käyttö tapahtuu lääketieteellisiin tai turvallisuuteen liittyviin tarkoituksiin. Myös tekoälyn käyttö biometriseen luokitteluun, jonka tavoitteena on selvittää ihmisten rotu, poliittiset mielipiteet, ammattiliiton jäsenyys, uskonnolliset tai filosofiset uskomukset, seksuaalinen elämä tai seksuaalinen suuntautuminen, on kiellettyä. Tässä poikkeuksena on laillisesti hankittu biometriseen dataan liittyvä tietojoukkojen luokittelu tai suodatus. Poikkeus on voimassa myös lainvalvontasektorilla (39).

4.2.2 Suuririskiset tekoälyjärjestelmät

Suuririskisiin järjestelmiin luetaan tekoälyjärjestelmät, jotka voivat aiheuttaa vakavia riskejä terveydelle, turvallisuudelle tai ihmisten perusoikeuksille. Tällaisiksi järjestelmiksi luetaan muun muassa biometriset tunnistusjärjestelmät eli järjestelmät, joita käytetään esimerkiksi kasvojentunnistukseen, tekoälyjärjestelmät, joita käytetään kriittisen infrastruktuurin, kuten vesihuollon, sähköjakelun tai liikenteen hallintaan, ja tekoälyjärjestelmät, joita käytetään oppilaiden arviointiin, pääsykokeisiin tai oppimistulosten seurantaan. (38). Näiden järjestelmien on läpäistävä arvioinnit ennen markkinoille pääsyä ja säännöllisesti koko järjestelmän elinkaaren ajan (2).

Suuren riskin tekoälyjärjestelmiin kohdistuu tekoälyasetuksessa paljon sääntelyä. Ensimmäinen näihin järjestelmiin kohdistuva vaatimus on artiklassa 8 esitetty vaatimustenmukaisuus, joka edellyttää, että suuririskisiksi määritetyt tekoälyjärjestelmät täyttävät asetuksessa säädetyt vaatimukset (2).

Kun suuririskinen tekoälyjärjestelmä tuodaan markkinoille, sen tarjoajan tulee laatia koneluettava, allekirjoitettu kirjallinen EU-vaatimustenmukaisuusvakuutus tekoälyjärjestelmälle. Vakuutuksen tulee olla kansallisten viranomaisten saatavilla 10 vuoden ajan

siitä, kun tekoälyjärjestelmä on tullut markkinoille tai otettu käyttöön. Lisäksi järjestelmän tarjoajan on hankittava järjestelmälle CE-merkintä ja rekisteröitävä suuririskinen tekoälyjärjestelmä ja itsensä järjestelmän tarjoajana EU:n suuririskisten tekoälyjärjestelmien tietokantaan (2).

Ennen markkinoille saattamista myös suuririskisen tekoälyjärjestelmän teknisen dokumentaation tulee vastata tekoälyasetuksen vaatimuksia. Teknisellä dokumentaatiolla osoitetaan tekoälyjärjestelmän vaatimustenmukaisuus, ja se sisältää kansallisten viranomaisten tekoälyn vaatimustenmukaisuuden arviointiin tarvitsemat tiedot tekoälyjärjestelmästä. Tekninen dokumentaatio sisältää vähintään tekoälyjärjestelmän yleisen kuvauksen, kuvauksen järjestelmän käyttötarkoituksesta sekä järjestelmän version ja tarjoajan nimen. Lisäksi vaaditaan yksityiskohtainen kuvaus tekoälyjärjestelmän osista ja sen kehittämisprosessista sekä tarkat tiedot järjestelmän seurannasta, toiminnasta ja valvonnasta, ja yksityiskohtainen kuvaus riskienhallintajärjestelmästä. Suorituskyvyn osalta vaaditaan dokumentaatioon kuvaus tekoälyjärjestelmän suorituskykymittareiden asianmukaisuudesta ja tarkka kuvaus käytössä olevasta järjestelmästä, jolla arvioidaan tekoälyjärjestelmän suorituskykyä markkinoille saattamisen jälkeisessä vaiheessa. Dokumentaatioon tulee myös luetteloida järjestelmässä kokonaan tai osittain sovelletuista yhdenmukaistetuista standardeista. Jos yhdenmukaistettuja standardeja ei ole sovellettu, tulee dokumentoida yksityiskohtainen kuvaus muista ratkaisuksista, joita on sovellettu asetuksen vaatimusten täyttämiseksi. Tähän kuuluu myös luettelo sovelletuista muista standardeista ja teknisistä eritelmistä. Dokumentaatioon tulee myös liittää jäljennös EU-vaatimustenmukaisuusvakuutuksesta, jonka tekoälyjärjestelmän tarjoaja on tehnyt. Dokumentaatio on pidettävä ajan tasalla, ja siihen on kuvattava järjestelmään sen elinkaaren aikana tehdyt muutokset (2).

Markkinoille saattamista varten suuririskisen tekoälyjärjestelmän tarjoajan on myös perustettava seurantajärjestelmä, jonka avulla voidaan kerätä tietoja, jotka koskevat kyseisen tekoälyjärjestelmien suorituskykyä koko sen elinkaaren ajan. Seurantajärjestelmä voi perustua käyttöönottajän toimittamiin tietoihin, tai ne voidaan kerätä muista lähteistä, mutta tietojen pohjalta pitää pystyä arvioimaan, täyttääkö tekoälyjärjestelmä sille asetetut vaatimukset (2).

Seurantajärjestelmän lisäksi suuririskisen tekoälyjärjestelmän tarjoajan on perustettava iteratiivinen, prosessina toimiva riskinhallintajärjestelmä. Sen käytön tulee jatkua suunnitellusti tekoälyjärjestelmän koko elinkaaren ajan. Riskienhallintajärjestelmään kuuluu, että tekoälyjärjestelmän riskien kartoitus ja arviointi sekä riskienhallintatoimet tarkastetaan säännöllisesti ja tarvittaessa päivitetään ajan tasalle. Tämä koskee erityisesti niitä riskejä, jotka voivat vaikuttaa terveyteen, turvallisuuteen tai ihmisten perusoikeuksiin.

Kartoitukseen tulee kuitenkin sisällyttää myös järjestelmän käyttötarkoituksen mukaiseen käyttöön liittyvät riskit ja riskit, jotka seuraavat järjestelmän väärinkäytöstä. Lisäksi markkinoille saattamisen yhteydessä suuririskisten tekoälyjärjestelmien kohdalla käyttöönotettavasta seurantajärjestelmästä kerätyn tiedon analyysin perusteella esiin nousevat riskit tulee huomioida arvioinnissa. Riskeille tulee määritellä sopivat hallintatoimet siten, että jäännösriskin ja koko suuririskisen tekoälyjärjestelmän kokonaisjäännösriskin voidaan katsoa olevan hyväksytyllä tasolla (2).

Ennen markkinoille saattamista suuririskisen tekoälyjärjestelmän tarjoajan on varmistettava, että järjestelmälle on tehty asetuksessa vaadittu vaatimustenmukaisuuden arviointi. Arvioinnilla varmistetaan asetuksessa vaaditun laadunhallintajärjestelmän ja teknisen dokumentaation vaatimustenmukaisuus sekä laadunhallintajärjestelmän seuranta (2).

Markkinoille saattamisen ja käyttöönoton jälkeen tulee suuririskistä tekoälyjärjestelmää pystyä valvomaan sen koko elinkaaren ajan. Tämän ihmisen suorittaman valvonnan mahdollistaminen tulee ottaa huomioon jo suunnittelussa ja kehityksessä. Ihmisen suorittamalla valvonnalla pyritään ehkäisemään järjestelmän käytöstä aiheutuvia riskejä. Valvonnan tason tuleekin suhteuttaa järjestelmään liittyviin riskeihin, itsenäisyyden tasoon ja käyttöympäristöön. Tekoälyjärjestelmän tarjoajan tulee tarjota järjestelmän käyttöönottajalle tarvittavat tiedot ja välineet, jotta käyttöönottaja pystyy tehokkaasti ja asianmukaisesti hoitamaan ihmisen suorittamaa valvontaa käyttöönoton jälkeen (2).

Mikäli suuririskinen tekoälyjärjestelmä ei markkinoille saattamisen tai käyttöönoton jälkeen syystä tai toisesta enää täytä tekoälyasetuksen vaatimuksia, pitää järjestelmän tarjoajan viipymättä toteuttaa tarvittavat korjaavat toimenpiteet, joilla järjestelmä saadaan vaatimusten mukaiseksi. Mikäli järjestelmän tarjoaja ei pysty tai halua tekoälyjärjestelmää korjata, tulee se poistaa käytöstä ja markkinoilta (2).

Ihmisen suorittaman valvonnan mahdollistamisen lisäksi tulee jo suuririskisen tekoälyjärjestelmän suunnittelu- ja kehitysvaiheessa ottaa huomioon muitakin tekoälyasetuksen vaatimuksia. Yksi näistä vaatimuksista on laadunhallintajärjestelmän käyttöönotto. Sitä käytetään muiden tekoälyasetuksessa annettujen vaatimusten noudattamisen varmistamisessa ja seurannassa. Laadunhallintajärjestelmä on dokumentoitava, ja sen tulee sisältää ainakin seuraavat asiat: strategia säännösten noudattamiseen mukaan liittyvien vaatimustenmukaisuusmenettelyt; tekoälyjärjestelmän suunnittelussa, suunnittelun valvonnassa ja rakenteen tarkastuksessa käytettävät tekniikat, menettelyt ja toimenpiteet; järjestelmän tarkastus-, testaus- ja validointimenettelyt sekä niiden suoritusiheydet; järjestelmässä sovellettavat tekniset eritelmät ja standardit; datanhallintajärjestelmät ja

-menettelyt sekä datanhankinta, keruu, analysointi, tunnisteiden lisääminen, tallentaminen, suodattaminen, datan louhinta, datan yhdistäminen ja datan säilyttäminen; tekoälyjärjestelmän riskienhallintajärjestelmä; markkinoille saattamisen ja käyttöönoton jälkeinen seurantajärjestelmä; menettelyt vakavasta vaaratilanteesta ilmoittamiseen; viestinnän järjestäminen kansallisten toimivaltaisten viranomaisten, muiden asianomaisten viranomaisten ja muiden olennaisten sidosryhmien kanssa; järjestelmät ja menettelyt dokumentaation ja tiedon säilyttämisestä; resurssien hallinta ja vastuuvollisuuskehys (2).

Kehityksessä on myös huomioitava datanlaatuun ja hallintaan liittyviä vaatimuksia. Kehityksessä käytetyn koulutus-, validointi- ja testausdatan on oltava käyttötarkoituksen huomioiden tarpeeksi merkityksellistä, riittävän edustavaa ja mahdollisimman virheetöntä ja täydellistä. Lisäksi datajoukkojen on käyttötarkoituksen vaatimassa laajuudessa huomioitava ne ominaisuudet, jotka ovat tyypillisiä sille ympäristölle, jossa tekoälyjärjestelmää käytetään. Datan laatua ja hallintaa pyritään varmistamaan myös vaatimalla tietojenkäsittelytoimia, jotka pienentävät dataan liittyviä riskejä. Tällaisia ovat huomautusten ja tunnisteiden lisääminen, datan puhdistus, ajantasaistaminen, datan rikastaminen ja yhdistäminen sekä datajoukkojen saatavuuden, määrän ja soveltuvuuden arviointi käyttötarkoitukseen nähden. Lisäksi datalle tulisi tehdä oletusten muotoilu – erityisesti niiden tietojen osalta, joita datan on tarkoitus mitata ja edustaa. Datan laatuun ja edustavuuteen liittyvillä vaatimuksilla pyritään pienentämään datan vinoutumiseen liittyviä riskejä (2).

Suuririskisiin tekoälyjärjestelmiin kohdistuu myös niiden tarkkuuteen ja suorituskykyyn liittyviä vaatimuksia. Tekoälyasetuksen artiklassa 15 todetaan, että suuririskisen tekoälyjärjestelmän tulisi toimia asiaankuuluvalla tarkkuudella ja vakaudella koko elinkaarensa ajan. Järjestelmän tarkkuustason mittarit tulee ilmoittaa sen käyttöohjeissa. Vaakan toiminnan vaatimuksella tarkoitetaan, että tekoälyjärjestelmä tulee kestää mahdollisimman paljon järjestelmän tai sen ympäristön vikoja, virheitä ja epäjohtonmukaisuuksia ilman, että sen toiminta häiriintyy (2). Tämä voidaan saavuttaa organisatorisilla toimenpiteillä ja teknisillä vararatkaisuilla, joihin kuuluvat myös varasuunnitelmat tai vikavarmistussuunnitelmat.

Tekoälyasetuksen artiklassa 15 säädetään myös suuririskisiltä tekoälyjärjestelmiltä vaaditusta kyberturvallisuuden tasosta. Tekoälyjärjestelmän tulee olla suunniteltu kestämään ulkopuolisten tahojen yritykset hyväksikäyttää järjestelmän haavoittuvuuksia muuttaakseen järjestelmän käyttöä, tuotoksia tai suorituskykyä. Lisäksi asetuksessa vaaditaan, että kyberturvallisuuden varmistamiseen käytettyjen teknisten ratkaisujen tulee olla riittävät suhteessa riskeihin ja olosuhteisiin. Suuririskisen tekoälyjärjestelmän

teknisiin ratkaisuihin on tarvittaessa lisättävä toimenpiteitä, joilla ehkäistään, havaitaan, rajoitetaan ja ratkaistaan pyrkimyksiä manipuloida koulutusdatajoukkoja, tekoälymallia tai syöttötietoja. Teknisillä ratkaisuilla ja toimenpiteillä tulisi myös ratkaista luottamus-
hyökkäykset ja mahdolliset tekoälymallin puutteet (2).

Suuririskisen tekoälyjärjestelmän toiminnan jäljitettävyyden ja seurannan helpottamiseksi on tekoälyasetuksessa velvoite lokitapahtumien tallentamisesta. Jokaisen suuririskisen tekoälyjärjestelmän on siis teknisesti pystyttävä tallentamaan tapahtumat automaattisesti koko elinkaarensa ajan. Lokitapahtumien tulee sisältää ainakin järjestelmän käyttöjaksojen kirjaukset; viitetietokannat, joihin järjestelmän syöttötietojen tarkastus perustuu; syöttötiedot, joiden haku on johtanut tulokseen sekä tuloksen tarkastamisen tehneen valvojan ihmisen tunnistetiedot. Lisäksi lokitietojen tulisi mahdollistaa sellaisten tilanteiden tunnistamisen, jotka voivat johtaa siihen, että suuririskinen tekoälyjärjestelmä aiheuttaa kansallisen riskin. Lokitietojen tulisi myös tukea ihmisen suorittamaa tekoälyjärjestelmän toiminnan valvontaa ja tekoälyjärjestelmän tarjoajien suorittamaa markkinoilla saattamisen jälkeistä seurantaa. Lokitietoja tulee säilyttää järjestelmän käyttötarkoitukseen nähden sopiva ajanjakso, kuitenkin vähintään kuusi kuukautta.

Suuririskisiä tekoälyjärjestelmiä koskee myös avoimuusvaatimus. Järjestelmien kehityksen näkökulmasta se tarkoittaa, että järjestelmä on suunniteltava ja kehitettävä siten, että niiden toiminta on riittävän avointa ja käyttöönottajat pystyvät ymmärtämään tuotoksia ja käyttämään niitä. Lisäksi avoimuusvaatimus velvoittaa suuririskisen tekoälyjärjestelmän tarjoajan toimittamaan käyttöönottajille järjestelmän käyttöohjeet. Niissä kuvataan järjestelmän ominaisuudet, valmiudet ja suorituskyvyn rajoitukset sekä järjestelmän käyttötarkoitus. Lisäksi kuvataan ihmisen suorittamat valvontatoimenpiteet ja tekniset valvontatoimenpiteet sekä järjestelmän vaatimat laskenta- ja laiteresurssit, tarvittavat huolto- ja hoitotoimenpiteet, ohjelmistopäivitysvälit ja järjestelmän arvioitu käyttöikä (2).

4.2.3 Tapausanalyysit

Tässä kappaleessa käydään läpi kaksi pintapuolista tapausanalyysiä, joilla havainnollistetaan tekoälyasetuksen vaikutuksia suuririskisiin tekoälyjärjestelmiin. Tapausanalyysien tekoälyjärjestelmät ovat kuvitteellisia.

Ensimmäinen tapausanalyysi koskee pankkisektorin tekoälyjärjestelmää, jota käytetään luotonantoprosessia tekemään luotonantopäätöksiä. Kyseinen järjestelmä käyttää asiakkaiden historiadataa ja erilaisia algoritmeja arvioidessaan asiakkaiden luottokelpoisuutta. Data sisältää tietoa asiakkaiden taloudellisesta tilanteesta,

maksuhistoriatietoja, tietoja jo myönnettyistä luotoista, maksuhäiriömerkintä- ja ulosotto-tietoja sekä muita relevantteja tietoja. Järjestelmä antaa suosituksia asiakkaiden luottohakemusten myöntämisestä tai hylkäämisestä.

Tekoälyasetuksen määritelmän mukaan tällainen luotonantopäätöksissä käytetty tekoälyjärjestelmä kuuluu suuririskisten tekoälyjärjestelmien luokkaan. Sen tekemät päätökset vaikuttavat suoraan henkilöiden taloudelliseen tilanteeseen ja voivat aiheuttaa merkittäviä taloudellisia seurauksia, kuten luottihakemuksen hylkäämisen.

Tällaisen järjestelmän kehitystyössä tulee huomioida tekoälyasetuksen vaatimukset kyberturvallisuuden tasosta sekä toiminnan tarkkuudesta ja vakaudesta erityisen huolellisesti. Järjestelmän päätösten tulee olla tarkkoja ja järjestelmän toiminnan hyvin vika-sietoista, jotta virheellisiltä tuotoksilta vältyttäisiin luotonantopäätöksissä. Lisäksi on varmistettava koulutus-, testaus- ja validointidatajoukkojen laatu ja eheys sekä varmistettava, ettei datajoukkoihin tai malliin tule vinoumia, jotka voisivat johtaa syrjintään luotonantopäätöksissä. Kehittäjien tulisi noudattaa secure by design -periaatetta kehitystyössä. Lisäksi tulisi muistaa, että data, jota järjestelmä tulee käsittelemään, sisältää henkilötietoja. Näin ollen myös tietosuoja tulisi huomioida kehitystyössä. Ennen käyttöönottoa järjestelmään tulee tehdä tietoturvatestaukset ja auditointi sekä dokumentoida järjestelmän toiminta ja tietoturvatimet yksityiskohtaisesti. Järjestelmän vaatimustenmukaisuutta kehitystyön aikana tulee seurata tekoälyasetuksen vaatimalla laadunhallintajärjestelmällä.

Ennen käyttöönottoa tekoälyjärjestelmä tulee rekisteröidä EU:n suuririskisten tekoälyjärjestelmien tietokantaan ja sen vaatimustenmukaisuus osoittaa. Järjestelmälle pitää myös hankkia CE-merkintä ja ottaa käyttöön riskienhallintajärjestelmä ja seurantajärjestelmä. Järjestelmän toimintaa, datan eheyttä ja laatua sekä järjestelmän turvallisuutta tulee seurata koko sen elinkaaren ajan, ja järjestelmään liittyvä dokumentaatio on pidettävä ajan tasalla. Lisäksi järjestelmän toiminnan valvontaan tulee osoittaa henkilöt ja kouluttaa heidät tehtäväänsä. Järjestelmän käyttöön tulee myös avoimuusvelvoitteen mukaan tarjota käyttöohjeet. Kun järjestelmä on käytössä, asiakkaille tulee ilmoittaa, että tekoäly osallistuu heidän luotonantoprosessinsa päätöksentekoon.

Toinen tapausanalyysin kohde on autonominen ajoneuvo. Tässä autonomisella ajoneuvolla tarkoitetaan itseajavaa autoa. Se käyttää tekoälyä ja erilaisia sensoreita, kuten kameroita ja valotutkaa, tehdessään ajopäätöksiä ja navigoidessaan liikenteessä. Tekoälyjärjestelmä analysoi auton ympäristöä mahdollisimman reaaliaikaisesti ja tekee sen pohjalta päätöksiä auton liikkumiseen ja turvallisuuteen liittyen.

Tekoälyasetuksen riskiluokituksen mukaan autonominen ajoneuvo kuuluu suuririskisten tekoälyjärjestelmien luokkaan. Tekoälyjärjestelmän tekemät päätökset voivat vaikuttaa suoraan ihmisten turvallisuuteen. Järjestelmän tekemistä päätöksistä voi aiheutua vakavia fyysisiä vahinkoja, onnettomuustilanteita liikenteessä sekä mahdollisia taloudellisia seurauksia.

Autonomista ajoneuvoa kehitettäessä tulee noudattaa secure by design -periaatetta ja suorittaa järjestelmälle tietoturvatestaukset ja auditoinnit sekä muilla keinoilla varmistaa, ettei järjestelmän haavoittuvuuksia voida hyödyntää esimerkiksi päätösten vinouttamiseen tai ajoneuvon kaappaamiseen. Lisäksi tulee toteuttaa erityiset turvallisuus- ja valvontamekanismit, joilla varmistetaan ajoneuvon turvallinen ja vakaa toiminta ja päätöksenteon tarkkuus kaikissa tilanteissa. Ajoneuvoa ohjaavan tekoälyjärjestelmän koulutus-, testaus- ja validointidatajoukkojen tulee olla laadukkaat, eheät ja käyttötarkoitukseen tarpeeksi laaja ja edustava. Esimerkiksi datajoukkojen tai mallin vinouma voi johtaa syrjiviin päätöksiin vaikkapa väistötilanteessa. Tämän takia datan laatuun, laajuuteen ja edustavuuteen sekä mallin toimintaan tulee kiinnittää erityistä huomiota. Kehityksessä on oltava käytössä tekoälyasetuksessa vaadittu laadunhallintajärjestelmä, ja tekoälyjärjestelmän toiminnasta ja tietoturvatiedoista tulee olla kattava dokumentaatio. Lisäksi ajoneuvon valmistajan on tarjottava avoimuusperiaatteen mukaisesti ajoneuvon käyttöohjeet kaikille käyttäjille.

Ennen käyttöönottoa tekoälyjärjestelmän vaatimustenmukaisuus tulee osoittaa, ja se tulee rekisteröidä EU:n suuririskisten tekoälyjärjestelmien tietokantaan. Lisäksi järjestelmälle tulee hankkia CE-merkintä. Käyttöönoton jälkeen tekoälyjärjestelmälle tulee olla riskienhallintajärjestelmä ja seurantajärjestelmä käytössä. Lisäksi tekoälyjärjestelmän toimintaa tulee valvoa ja arvioida käyttöönottajän toimesta koko sen elinkaaren ajan.

4.2.4 Vähäisen riskin tekoälyjärjestelmät

Tekoälyasetus ei anna suoraa määritelmää vähäisen riskin tekoälyjärjestelmälle. Euroopan komission verkkosivustolla vähäisen riskin luokkaan kuuluvat riskit, jotka liittyvät tekoälyn läpinäkyvyyteen. Toisin sanoen vähäisen riskin tekoälyjärjestelmät eivät aiheuta merkittävää riskiä ihmisten terveydelle, turvallisuudelle tai perusoikeuksille, mutta niihin liittyy läpinäkyvyys- ja avoimuusriskejä. Tällaiset tekoälyjärjestelmät voivat toimia vuorovaikutuksessa ihmisten kanssa tai tuottaa sisältöä, mutta niiden käyttö ei aiheuta merkittäviä haittoja tai riskejä. Vähäisen riskin tekoälyjärjestelmiksi luetaan muun muassa chatbotit ja järjestelmät, jotka tuottavat synteettistä sisältöä (38).

EU:n tekoälyasetus sisältää vapaaehtoisia käytänteitä, joita tekoälyjärjestelmissä, myös vähäisen riskin tekoälyjärjestelmissä, suositellaan sovellettaviksi (2). Vähäisen riskin järjestelmille kohdistuu kuitenkin myös muita niin sanottuja avoimuusvelvoitteita. Niillä pyritään varmistamaan ihmisten luottamuksen säilyttäminen tarjoamalla tietoa tekoälyn käytöstä (38). Velvoitteissa esimerkiksi vaaditaan, että käyttäjälle ilmoitetaan, kun hän on vuorovaikutuksessa tekoälyn kanssa. Ilmoitusta ei vaadita, jos vuorovaikutustilanteessa on täysin ilmeistä, että vuorovaikutus tapahtuu tekoälyn kanssa (2).

Jos otetaan käyttöön tunteentunnistus tai biometrinen luokitusjärjestelmä, on käyttöön otosta ilmoitettava niille henkilöille, jotka altistuvat järjestelmälle. Toisin sanoen niille henkilöille, joiden tietoja järjestelmä käsittelee. Tällainen järjestelmä esimerkiksi tunnistaa tai päättelee tunteita tai aikomuksia tai luokittelee ihmisiä tiettyihin ryhmiin heidän biometrinen tietojensa perusteella. Järjestelmän käyttöönottajana on myös noudatettava Euroopan Unionin henkilötietojen käsittelyyn liittyviä muita säädöksiä, kuten tietosuojasetusta soveltuvin osin (2).

Synteettisen sisällön tuottamiseen kohdistuu myös avoimuusvelvoite. Synteettistä äänikuva-, video- tai tekstisisältöä tuottaessa on tekoälyjärjestelmän tarjoajien varmistettava, että tuotokset merkitään keinotekoisesti tuotetuiksi tai muokatuiksi. Merkinnän on oltava koneellisesti luettavassa muodossa. Vaatimus koskee myös yleiskäyttöisiä tekoälyjärjestelmiä (2). Tällä velvoitteella varmistetaan, että tuotokset tunnistetaan ihmisen sijasta tekoälyn tuottamiksi.

Syväväärennosten kohdalla on velvoite erikseen ilmoittaa, että tuotettu materiaali on keinotekoisesti tuotettu tai sitä on manipuloitu. Tämä vaatimus koskee sekä kuva-, ääni tai videosisältöä että tekstiä. Tekstin manipuloinnin kohdalla erityisenä lisänä on vaatimus ilmoittaa teksti keinotekoisesti tuotetuksi tai manipuloiduksi, mikäli sen julkaisemisen tarkoituksena on tiedottaa yleisölle yleistä etua koskevista asioista (2).

4.2.5 Minimaalisen riskin tai ei lainkaan riskiä tekoälyjärjestelmät

Minimaalisen riskin tai ei lainkaan riskiä -luokkaan kuuluvat esimerkiksi sähköpostisuodattimet ja tekoälyn viihdekäyttö. Tähän riskiluokkaan kuuluvat järjestelmät eivät lähtökohtaisesti aiheuta merkittäviä riskejä, ja niihin ei tekoälyasetuksessa kohdistu sääntelyä. Euroopan komission verkkosivuston mukaan tällä hetkellä suurin osa Euroopan Unionin alueella käytetyistä tekoälyjärjestelmistä kuuluu tähän riskiluokkaan (38).

EU:n tekoälyasetus kuitenkin sisältää vapaaehtoisia käytänteitä, joita myös minimaalisen riskin tai ei lainkaan riskiä sisältävien järjestelmien kohdalla voidaan soveltaa.

Vapaaehtoisuus kuitenkin tarkoittaa, ettei tekoälyjärjestelmille tai niiden tarjoajille ja kehittäjille kohdistu erityisiä sääntelyvaatimuksia (2).

4.2.6 Käyttöönottajille asetetut vaatimukset

EU:n tekoälyasetus asettaa vaatimuksia myös tekoälyjärjestelmiä käyttöönottaville ta-
hoille. Käyttöönottaja on tekoälyasetuksessa määritelty tekoälyjärjestelmää käyttäväksi
luonnolliseksi henkilöksi tai oikeushenkilöksi. Tähän kuuluvat myös viranomaiset, viras-
tot ja muut elimet, joiden valvonnassa järjestelmää käytetään. Määritelmän ulkopuolella
jää henkilökohtainen käyttö, kunhan se ei ole ammattikäyttöä (2).

Käyttöönottajalle asetetut vaatimukset koskevat suuririskisiä tekoälyjärjestelmiä. Ensini-
näkin käyttöönottaja veloitetaan noudattamaan tekoälyjärjestelmän käyttöohjeita ja
seuraamaan järjestelmän toimintaa niiden perusteella. Käyttöönottaja myös veloitetaan
raportoimaan tekoälyjärjestelmä tarjoajalle, mikäli järjestelmän toiminnassa huoma-
mataa poikkeavuus. Käyttöönottajan velvollisuuksiin, kuten tarjoajankin, kuuluu ihmi-
sen suorittama tekoälyn toiminnan valvonta. Valvontaa tekevän henkilön tulee olla pä-
tevä tehtävään, ja hänelle tulee antaa tarvittava koulutus, valtuudet sekä tuki tehtävän
suorittamiseen. Mikäli käyttöönottaja pystyy joltain osin valvomaan syöttötietoja, on ni-
den merkityksellisyys ja riittävä edustavuus käyttötarkoitukseen nähden varmistettava
(2).

Ennen suuririskisen tekoälyjärjestelmän käyttöönottoa työpaikalla on käyttöönottajan,
joka on työnantaja, kerrottava työntekijöille ja heidän edustajilleen käyttöönotosta, mi-
käli tekoälyjärjestelmän soveltamisalue kohdistuu työntekijöihin. Lisäksi työnantajan tu-
lee noudattaa muuta asiaan liittyvää unionin ja paikallista lainsäädäntöä (2).

Jos käyttöönottaja ottaa käyttöön suuririskisen tekoälyjärjestelmän, joka tekee päätök-
siä tai avustaa päätöksen teossa, joka kohdistuu ihmisiin, on käyttöönottajan ilmoitet-
tava tekoälyn käytöstä niille henkilöille, joihin järjestelmän käyttö kohdistuu (2).

Suuririskisen tekoälyjärjestelmän käyttöönottajan on myös noudatettava tiettyjä tietojen
säilyttämiseen liittyviä vaatimuksia. Heidän on säilytettävä järjestelmän automaattisesti
tuottamat lokitiedot niiltä osin kuin lokitiedot ovat heidän hallinnassaan, järjestelmän
käyttötarkoitukseen nähden asianmukaisen ajanjakson ajan, kuitenkin vähintään kuusi
kuukautta. Poikkeuksena tähän ovat tiedot, joita koskee jokin toinen unionin tai kansal-
lisessa lainsäädännössä tuleva säilytysaika ja peruste (2).

Lisäksi suuririskisen tekoälyjärjestelmän käyttöönottajan tulee tarvittaessa tehdä ennen
järjestelmän käyttöönottoa tietosuojasetuksen (EU) 2016/679 mukainen tietosuojaa

koskeva vaikutustenarviointi. Tämän lisäksi suuririskisten tekoälyjärjestelmien käyttöönottajien tulee tehdä arviointi järjestelmän käytön mahdollisista perusoikeusvaikutuksista, jos järjestelmää käytetään henkilöiden luottokelpoisuuden arviointiin tai heidän luottopisteytyksensä määrittämiseen tai henkilöitä koskevaan riskinarviointiin ja hinnoitteluun sairaus- ja henkivakuutusten tapauksessa. Käyttöönottajien, jotka ovat julkisoikeudellisia laitoksia tai julkisia palveluja tarjoavia yksityisiä yhteisöjä, on tehtävä perusoikeuksien arviointi aina ennen suuririskisen tekoälyjärjestelmän käyttöönottoa (2).

Perusoikeuksien arviointi sisältää kuvauksen käyttöönottajan prosesseista, joissa tekoälyjärjestelmää käytetään, sekä ajanjakson, jonka kuluessa järjestelmää on tarkoitus käyttää, ja kuinka usein. Aiemmin mainittu ihmisen suorittama valvonta on kuvattava, jotta voidaan todeta sen toimenpiteiden noudattavan annettuja käyttöohjeita. Lisäksi on kuvattava henkilöiden ja ryhmien luokat, joihin järjestelmän käyttö todennäköisesti vaikuttaa ja tehtävä riskien arviointi sekä kuvattava toimenpiteet, jotka on toteutettava kyseisten riskien toteutuessa. Kuvauksen tulee sisältää myös sisäistä hallintoa ja valitusmekanismeja koskevat järjestelyt (2).

Mikäli käyttöönottaja on viranomainen, unionin toimielin tai laitos, heidän on lisäksi noudatettava tekoälyasetuksen rekisteröintivelvoitetta ennen suuririskisen tekoälyjärjestelmän käyttöönottoa. Lisäksi nämä toimijat voivat käyttöönottaa vain sellaisia suuririskisiä tekoälyjärjestelmiä, jotka on rekisteröity EU:n suuririskisten tekoälyjärjestelmien tietokantaan (2).

Kaikilta käyttöönottajilta edellytetään myös yhteistyötä kaikkien asiaankuuluvien viranomaistahojen kanssa niissä toimissa, joita viranomaiset toteuttavat tekoälyasetuksen täyttööntäytymiseksi ja valvomiseksi (2).

5. YHTEENVETO JA POHDINTA

5.1 Tulosten yhteenveto

Katsauksessa kartoitettiin tekoälyn yleisimpiä riskejä ja niiden hallintakeinoja sekä EU:n tekoälyasetuksen vaikutuksia tekoälyjärjestelmien tietoturva-vaatimuksiin. Analyysin perusteella yleisimmiksi riskeiksi tunnistettiin mallin ja datan manipulointiin liittyvät riskit, yksityisyys- ja tietosuojariskit, tekoälyjärjestelmien tietoturva-vaavoittuvuudet ja mallin puolueellisuuteen liittyvät riskit, tekoälyn käyttö kyberhyökkäyksissä sekä tekoälyn käyttöön liittyvät ihmisoikeus- ja eettiset riskit. Näiden riskienhallintaan katsauksen analyysissä löytyi useita hallintakeinoja. Mallin ja datan manipuloinnin ehkäisemiseen ehdotettiin muun muassa robustista koulutusta, regularisointitekniikoiden käyttöä, syötteiden validointia, datan laadun ja eheyden varmistamista sekä tekoälyn toiminnan säännöllisiä auditointeja. Tietosuoja- ja yksityisyydensuojariskien hallintaa sopivat muun muassa tietojen anonymisointi ja pseudonymisointi sekä henkilötietojen minimointi. Security by design -periaatteen noudattaminen tekoälyn kehittämisessä parantaa tekoälyjärjestelmien tietoturvaa ja vähentää järjestelmien tietoturva-vaavoittuvuuksiin liittyviä riskejä. Eettisten standardien määrittämisellä voitaisiin vastata tekoälyn käyttöön liittyviin eettisiin haasteisiin.

EU:n tekoälyasetus pyrkii varmistamaan tekoälyjärjestelmien turvallisuuden ja luotettavuuden sekä suojelemaan perusoikeuksia. Tekoälyasetus on lähestymistavaltaan riskiperusteinen ja asettaa vaatimuksia tekoälyjärjestelmille niiden riskiluokituksen mukaan. Tekoälyasetuksen riskiluokkia ovat kielletty, suuri riski, vähäinen riski ja minimaalinen riski tai ei riskiä (37).

Tekoälyasetus kieltää kaikki tekoälyjärjestelmien käyttötavat, jotka manipuloivat tai harhaanjohtavat käyttäjiä, käyttävät hyväksi käyttäjän haavoittuvaa asemaa, mahdollistavat laajamittaisen valvonnan ilman oikeudellista perustetta tai suorittavat syrjivää tai muuten haitallista sosiaalista pisteytystä (39).

Suuririskisille tekoälyjärjestelmille on asetuksessa monia vaatimuksia. Ennen käyttöönottoa suuririskisten tekoälyjärjestelmien tulee muun muassa olla kattavasti dokumentoituja ja niiden vaatimustenmukaisuuden tulee olla todennettu. Suuririskiset tekoälynjärjestelmät pitää rekisteröidä EU:n omaan tietokantaan ja niille pitää hankkia CE-merkintä. Kehitystyössä suuririskisten tekoälyjärjestelmien kohdalla tulee huomioida myös asetuksesta tulevat koulutus-, validointi ja testausdatan laatuun, edustavuuteen ja virheettömyyteen liittyvät vaatimukset. Tekoälyasetuksessa on vaatimuksia myös muun

muassa tekoälyjärjestelmän seurantaan, toimintavarmuuteen, tietoturvaan ja tarkkuuteen liittyen. Suuririskisille tekoälyjärjestelmille pitää olla käytössä riskienhallintajärjestelmä, jonka avulla niiden toimintaa seurataan koko niiden elinkaaren ajan. Lisäksi järjestelmien tulee täyttää sille asetetut tarkkuus- ja vakausvaatimukset, joilla varmistetaan tuotosten laatua ja järjestelmän vikasietoisuutta. Suuririskisen tekoälyjärjestelmän tulee myös pystyä vastustamaan ulkopuolisten tahojen yritykset hyväksikäyttää järjestelmän haavoittuvuuksia. Lisäksi asetuksessa vaaditaan ihmisen suorittamaa valvontaa tekoälyjärjestelmään ja sen toimintaan koko sen elinkaaren ajan (2).

Vähäisen riskin tekoälyjärjestelmille tekoälyasetus antaa avoimuusvelvoitteita. Käyttäjille on muun muassa ilmoitettava, jos he ovat vuorovaikutuksessa tekoälyn kanssa. Lisäksi tekoälyn tuottamat synteettiset sisällöt tulee merkitä tekoälyn tekemiksi tuotteiksi. Vaatimus koskee erityisesti syvävääreännöksiä. Minimaalisen riskin ja ei lainkaan riskiä sisältäviin järjestelmiin ei tekoälyasetuksessa kohdistu lainkaan sääntelyä (2).

Tekoälyjärjestelmän käyttöönottajalle asetus antaa joitakin velvoitteita. Tällaisia ovat muun muassa tekoälyjärjestelmän käyttöohjeiden noudattaminen ja järjestelmän käyttö vain sille tarkoitettulla tavalla. Käyttöönottajan tulee myös seurata tekoälyjärjestelmän toimintaa ja raportoida järjestelmän tarjoajalle mahdollisista poikkeamista toiminnassa. Käyttöönottajan vastuulla on myös omalta osaltaan järjestää tekoälyjärjestelmälle suoritettava ihmisen tekemä valvonta, jos kyseessä on suuririskinen tekoälyjärjestelmä (2).

5.2 Pohdinta -tekoälyasetus ja tekoälyn yleisimmät riskit

Kun tarkastellaan tekoälyasetuksen vaatimuksia ja katsauksen analyysissä yleisimmiksi nousseita riskejä, voidaan todeta sääntelyn kohdistuvan melko hyvin yleisimpiin riskeihin. Esimerkiksi tekoälyasetuksen kielletyt käyttötapaukset yrittävät suitsia myös tekoälyn käyttöä kyberhyökkäyksissä. Kiellettyjen käyttötapausten listalla ovat muun muassa tekoälyjärjestelmät ja käyttötapaukset, joilla pyritään manipuloimaan käyttäjää sekä käyttötapaukset, joissa käytetään hyväksi käyttäjän haavoittuvaa asemaa.

Datan manipulointiin ja tekoälyn puolueellisuuteen liittyviin riskeihin asetus vastaa kohdistamalla sääntelyä tekoälyn käyttämään dataan. Asetus kohdistaa vaatimuksia koulutus-, validointi- ja testausdatan laatuun, edustavuuteen, eheyteen ja virheettömyyteen. Vaatimukset tekoälyjärjestelmien toiminnan seurantaan ja ihmisen suorittamaan valvontaan ovat sääntelytoimia, joilla pyritään varmistamaan, että mahdollinen datan tai mallin manipulointi tai tekoälyn ei-toivottu käytös huomataan ja voidaan korjata. Tekoälyn puolueellisuutta suitsivat myös järjestelmien toiminnalle asetetut vakaus- ja

tarkkuusvaatimukset. Ne vaativat, että tekoälyjärjestelmät toimivat käyttötarkoitukseensa nähden riittävälle vakaudella ja tarkkuudella kaikissa tilanteissa.

Tekoälymallin manipulointiin liittyviin riskeihin tekoälyasetus vastaa kohdistamalla vaatimuksia tekoälyjärjestelmien tietoturvaan. Asetus vaatii, että tekoälyjärjestelmien tulee olla vastustuskykyisiä ulkopuolisten tahojen hyväksikäyttöyrityksiä vastaan. Järjestelmiltä edellytetään myös vikasietoisuutta. Niiden tulee toimia luotettavalla tasolla suhteessa käyttötarkoitukseen myös virhetilanteissa. Mallin toimintaa ei siis pitäisi päästä manipuloimaan edes virhetilanteen seurauksena. Tekoälyjärjestelmille asetetut dokumentaatiovaatimukset varmistavat, että tekoälyjärjestelmät ovat kattavasti dokumentoituja ja niiden toimintaa ja tuloksia voidaan ymmärtää ja arvioida. Lisäksi suuririskisten tekoälyjärjestelmien edellytetään olevan ihmisen valvonnassa. Kattava dokumentaatio ja ihmisen suorittama valvonta auttavat havaitsemaan manipuloinnin ajoissa ja rajoittamaan sen seurauksia.

Koska osa tekoälyn käyttöön liittyvistä tietosuoja- ja yksityisyysriskeistä on seurausta tekoälyjärjestelmien haavoittuvuuksien hyväksikäytöstä, tekoälyasetuksen artiklassa 15 annetut vaatimukset tekoälyjärjestelmien tietoturvaan pienentävät myös näiden riskien toteutumista. Tekoälyasetus ei myöskään poista EU:n tietosuoja-asetuksen velvoitetta, jonka mukaan ennen uusien teknologioiden, myös tekoälyn, käyttöönottoa tulee niille suorittaa tietosuojavaikutusten arviointi (40). Tämän lisäksi tekoälyasetuksessa on vaatimus perusoikeuksiin kohdistuvien vaikutusten arvioinnista ennen tekoälyjärjestelmän käyttöönottoa (2).

Tekoälyasetuksen 15 artikla vastaa myös tekoälyjärjestelmien tietoturva- ja haavoittuvuuksiin liittyviin riskeihin. Artiklan mukaan tekoälyjärjestelmien tulee olla vastustuskykyisiä ulkopuolelta tulevia hyökkäyksiä vastaan. Myös asetuksesta tulevat vaatimukset koko järjestelmän elinkaaren ajan kestäväälle ihmisen suorittamalle valvonnalle ja riskienhallintajärjestelmälle tukevat järjestelmien tietoturva- ja haavoittuvuuksista johtuvien riskien hallintaa.

Suuririskisten järjestelmien vaatimukset, vähäriskisten järjestelmien avoimuusvaatimukset ja kielletyt käyttötapaukset kohdentavat sääntelyä melko tarkasti juuri tekoälyn yleisimpiin riskeihin. Asetuksen tehokkuus tekoälyn käyttöön liittyvien yleisimpien riskien hallinnassa perustuu kuitenkin siihen, kuinka tehokkaasti asetuksen täytäntöönpanoa valvotaan ja tekoälyjärjestelmien vaatimusten mukaisuutta seurataan. Painoarvoa asetukselle antaa myös asetuksen rikkomisesta seuraava rangaistus.

Tekoälyasetuksen vaatimusten noudattamista valvoo markkinavalvontaviranomainen, ja perusoikeuksien suojelusta vastaavat viranomaiset huolehtivat kukin oman

lainkäyttöalueensa osalta perusoikeuksien toteutumisen seurannasta. Valvova viranomaisena voi pyytää markkinavalvontaviranomaista selvittämään, onko tekoälyasetuksen velvoitteita rikottu. (41). Asetuksen rikkomisesta voidaan antaa sakkorangaistus, jonka suuruus riippuu rikkomuksesta. Kiellettyyn käyttötapaan kohdistuvasta rikkomuksesta sakko voi olla enintään 35 000 000 euroa tai, jos rikkomuksen tekijä on yritys, enintään 7 prosenttia sen edeltävän tilikauden vuotuisesta maailmanlaajuisesta kokonaisliikevaihdosta, riippuen siitä, kumpi näistä määristä on suurempi (2).

Jos tekoälyn käyttöönottaja tai tekoälyjärjestelmän tarjoaja rikkovat asetuksen heille antamia vaatimuksia, voidaan rikkojalle määrätä sakko, joka on enintään 15 000 000 euroa tai, jos rikkomuksen tekijänä on yritys, enintään 3 prosenttia sen edeltävän tilikauden vuotuisesta maailmanlaajuisesta kokonaisliikevaihdosta, riippuen siitä, kumpi näistä määristä on suurempi (2).

Mikäli valvovalle viranomaiselle on annettu pyydettäessä virheellistä, puutteellista tai harhaanjohtavaa tietoa, voidaan siitä määrätä sakko, joka on enintään 7 500 000 euroa tai, jos rikkomuksen tekijä on yritys, enintään prosentti sen edeltävän tilikauden vuotuisesta maailmanlaajuisesta kokonaisliikevaihdosta sen mukaan, kumpi näistä määristä on suurempi (2).

Mikäli rikkomukseen syyllistynyt taho on Euroopan unionin toimielin, julkinen laitos tai virasto, on kiellettyjen käyttötapausten kohdalla rikkomuksesta määrätty sakko enintään 1 500 000 euroa. Asetuksen muiden vaatimusten ja velvoitteiden rikkomisesta määrätty sakko voi olla enintään 750 000 euroa. Lisäksi Euroopan tietosuojavaltuutettu voi määrätä sakkoja Euroopan unionin toimielimille sekä julkisille laitoksille ja virastoille tietosuoja-asetuksen soveltamisalaan liittyvissä tekoälyjärjestelmiä koskevissa rikkomuksissa (2)

5.3 Pohdinta -katsauksen kritiikki ja jatkotutkimusmahdollisuudet

Katsauksessa keskityttiin selvittämään tekoälyn liittyviä yleisimpiä riskejä ja EU:n tekoälyasetuksen antamia vaatimuksia tekoälyjärjestelmille ja niiden käyttöönottajille. Katsauksessa haettiin myös hallintakeinoja tekoälyn yleisimmille riskeille. Katsaus pystyi ainakin ylätasolla vastaamaan molempiin tutkimuskysymyksiin ja saavutti sille asetetut tavoitteet.

Vaikka katsauksen kohdalla päästiin tavoitteeseen, on molempien tutkimuskysymysten kohdalla syytä myös arvioida katsauksen heikkouksia. Tekoälyn yleisimpien riskien kohdalla katsauksen heikkoudet kohdistuvat erityisesti tiedonhankintavaiheeseen.

Yksinkertaisten hakusanojen käyttö hakuprosessissa saattoi johtaa liian rajattuun haakuun ja jättää katsauksen ulkopuolelle olennaisia tutkimuksia. Hakusanalausekkeille olisi hakutulosten kattavuutta ja laatua mahdollisesti voitu parantaa. Lisäksi haussa käytetty aikarajaus, jolla pyrittiin varmistamaan hakutulosten ajankohtaisuus, saattoi sekoin omalta osaltaan rajata tuloksia liikaa. Rajauksen seurauksena olennaisia ja ajankohtaisia tutkimuksia saattoi rajautua katsauksen ulkopuolelle. Lisäksi hakutulosten arviointiprosessi ei ollut tarpeeksi systemaattinen. Hakutulosten ensimmäinen tarkastelu perustui melko lyhyeen kuvaukseen hakutuloksen sisällöstä. Vaikka hakutuloksia oli paljon, olisi ensimmäisen tarkastelun kohdalla kannattanut noudattaa tarkempaa ja systemaattisempaa läpikäyntiä. Nyt tutkimuskysymyksen kannalta olennaisia hakutuloksia saattoi rajautua ulos katsauksesta. Katsauksessa yleisimpien riskien kartoitukseen käytetty aineisto jäikin melko pieneksi. Se käsitti vain 31 vertaisarvioitua julkaisua. Aineiston pienuus heikentää tulosten yleistettävyyttä ja luotettavuutta.

EU:n tekoälyasetukseen liittyvän tutkimuskysymyksen kohdalla analyysistä jäi puuttumaan laajempi konteksti ja asetuksen syvällisempi arviointi. Se olisi antanut kattavamman kuvan asetuksen vaikutuksista. Arvioinnissa olisi voinut myös huomioida, miten asetukset suhteutuu muihin EU:n asetuksiin, joiden sääntelyalue on lähellä sitä, kuten EU:n tietosuojasetukseen.

Tekoälyasetuksen vaikutusten tarkastelussa myös tapausanalyysit jäivät melko pintapuolisiksi. Syvällisempi analyysi olisi antanut enemmän käytännön näkökulmaa asetuksen vaikutuksiin. Analyysiin olisi voinut myös sisällyttää konkreettisia esimerkkejä toimenpiteistä, joilla vaatimustenmukaisuus analyysin kohteena olevissa järjestelmissä voidaan saavuttaa.

Tässä katsauksessa tarkasteltiin sekä tekoälyyn liittyviä yleisimpiä riskejä, että tekoälyasetuksesta tulevia tekoälyjärjestelmien tietoturvaan liittyviä vaatimuksia. Kun asetukset on ollut voimassa pidempään, on mahdollista tutkia tekoälyasetuksen vaikutusta tekoälyn riskeihin, niiden hallintaan ja ehkäisyyn. Tutkimuksessa voisi esimerkiksi keskittyä tarkastelemaan, tekoälyjärjestelmien riskien kartoitusta ja hallintaa ennen ja jälkeen asetuksen voimaantuloa. Tällaisella tarkastelulla voitaisiin arvioida, miten asetukset on muuttanut tekoälyjärjestelmien riskienhallintaa ja turvallisuutta. Samalla voisi tehdä arvioita, ovatko jotkin riskit jääneet asetuksesta liian vähälle huomiolle tai onko tullut uusia riskejä, joita asetukset ei huomioi.

Toinen mahdollinen tutkimussuunta on tekoälyasetuksen vaikutus tekoälyjärjestelmien tietoturvaan ja kehitykseen. Tutkimuksessa voisi kartoittaa tekoälyasetuksen konkreettisia vaikutuksia tekoälyjärjestelmien tietoturvasuhteeseen. Kiinnostavaa olisi esimerkiksi

tarkastella, miten asetuksen tietoturva-vaatimukset on käytännössä toteutettu tekoälyjärjestelmissä tai miten asetus on vaikuttanut tekoälyn kehitysprosesseihin. Toinen mahdollinen lähestymistapa olisi kartoittaa tekoälyjärjestelmien tietoturvasoaa ennen ja jälkeen asetuksen voimaantulon. Näin asetuksen vaikutus tekoälyjärjestelmien tietoturvasoonaan olisi selkeästi nähtävissä. Ehdotetut jatkotutkimuskohteet ovat mahdollisia, kun tekoälyasetus on ollut voimassa jonkin aikaa ja tekoälyjärjestelmät on ehditty päivittää täyttämään asetuksen vaatimukset.

LÄHTEET

- [1] Vähä-Sipilä, A., Marchal, S. and Aksela, M. (2021) Tekoölyn Soveltamisen kyber- Turvallisuus Ja Riskienhallinta, Tekoölyn soveltamisen kyberturvallisuus ja riskienhallinta. Saatavilla: https://www.traficom.fi/sites/default/files/media/publication/Tekoölyn_soveltamisen_kyberturvallisuus_ja_riskienhallinta.pdf .
- [2] Euroopan parlamentin ja neuvoston asetus (EU) 2024/1689, annettu 13 päivänä kesäkuuta 2024, tekoälyä koskevista yhdenmukaistetuista säännöistä ja asetusten (EY) N:o 300/2008, (EU) N:o 167/2013, (EU) N:o 168/2013, (EU) 2018/858, (EU) 2018/1139 ja (EU) 2019/2144 sekä direktiivien 2014/90/EU, (EU) 2016/797 ja (EU) 2020/1828 muuttamisesta (tekoälysäädös) (ETA:n kannalta merkityksellinen teksti). Euroopan unionin virallinen lehti, L-sarja 2024/1689, 12.7.2024. ELI: <http://data.europa.eu/eli/reg/2024/1689/oj>
- [3] Ahmed, S.Q., Ganesh, B.V., Kumar, S.S., Mishra, P., Anand, R., & Akurathi, B. (2024). A Comprehensive Review of Adversarial Attacks on Machine Learning. ArXiv, abs/2412.11384.
- [4] Hendrycks, D., Mazeika, M., & Woodside, T. (2023). An Overview of Catastrophic AI Risks. ArXiv, abs/2306.12001.
- [5] Ferrag, Mohamed Amine et al. "Generative AI in Cybersecurity: A Comprehensive Review of LLM Applications and Vulnerabilities." (2024).
- [6] Burhanuddin, L.A.b., Shibghatullah, A.S.B., Ilias, I.S.C., Zainudin, Z.B., Zamry, N.B.M. (2025). AI-Enhanced Cybersecurity: A Comprehensive Review of Techniques and Challenges. In: Al-Sharafi, M.A., Al-Emran, M., Mahmoud, M.A., Arpaci, I. (eds) Current and Future Trends on AI Applications. Studies in Computational Intelligence, vol 1178. Springer, Cham. https://doi.org/10.1007/978-3-031-75091-5_7
- [7] Junklewitz, H., Hamon, R., André, J., Evas, T., Soler Garrido, D. & Sanchez Martin, J.I. (2023). Cybersecurity of Artificial Intelligence in the AI Act. European Commission, Joint Research Centre. Saatavilla: https://publications.jrc.ec.europa.eu/repository/bitstream/JRC134461/JRC134461_01.pdf
- [8] Tronnier, F., Löbner, S., Lacombe, MH., Rannenber, K. (2026). Regulatory Challenges in Cybersecurity – A Critical Analysis of the EU AI Act. In: Drevin, L., Leung, W.S., von Solms, S. (eds) Information Security Education. Empowering People Through Information Security Education. WISE 2025. IFIP Advances in Information and Communication Technology, vol 742. Springer, Cham. https://doi.org/10.1007/978-3-031-94924-1_6
- [9] Nolte, H., Rateike, M., & Finck, M. (2025). Robustness and Cybersecurity in the EU Artificial Intelligence Act. Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency.
- [10] Rangone, N. and Megale, L. (2025) 'Risks Without Rights? The EU AI Act's Approach to AI in Law and Rule-Making', European Journal of Risk Regulation, pp. 1–16. doi:10.1017/err.2025.13.

- [11] Peda.net. (2021). Laadullinen sisällönanalyysi. [online] Saatavilla: <https://peda.net/jyu/sport/ljto2/lso/pgs/huovisen-ryhma-2021-2022/aineiston-analysointi/laadullinen-sisallonanalyysi>
- [12] Vuori, J. (n.d.). Laadullinen sisällönanalyysi. [online] Tietoarkisto. Saatavilla: <https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvali/analyysitavan-valinta-ja-yleiset-analyysitavat/laadullinen-sisallonanalyysi/>.
- [13] Euroopan komissio (n.d.). Yleiskäyttöiset tekoälyn käytäntesäännöt. [online] Shaping Europe's digital future. Saatavilla: <https://digital-strategy.ec.europa.eu/fi/policies/ai-code-practice>
- [14] Pascu, C. et al. (2023) Artificial Intelligence and cybersecurity research: Enisa Research and Innovation Brief, Enisa. Heraklion, Greece: ENISA. Saatavilla: <https://www.enisa.europa.eu/sites/default/files/publications/Artificial%20Intelligence%20and%20Cybersecurity%20Research.pdf>.
- [15] Villegas-Ch, W., & García-Ortiz, J. (2023). Toward a Comprehensive Framework for Ensuring Security and Privacy in Artificial Intelligence. *Electronics*, 12(18), 3786. Saatavilla: <https://doi.org/10.3390/electronics12183786>.
- [16] Altun, İ.Z. and Emre Özkök, A. (2024) Securing Artificial Intelligence: Exploring Attack Scenarios and defense strategies, 2024 12th International Symposium on Digital Forensics and Security (ISDFS), pp. 1–6. doi:10.1109/isdfs60797.2024.10527288.
- [17] Raimundo, R., & Rosário, A. (2021). The Impact of Artificial Intelligence on Data System Security: A Literature Review. *Sensors*, 21(21), 7029. Saatavilla: <https://doi.org/10.3390/s21217029>.
- [18] Kafali, E., Preuveneers, D., Semertzidis, T., & Daras, P. (2024). Defending Against AI Threats with a User-Centric Trustworthiness Assessment Framework. *Big Data and Cognitive Computing*, 8(11), 142. Saatavilla: <https://doi.org/10.3390/bdcc8110142>.
- [19] Khanal, S., Zhang, H. and Taeihagh, A. (2024) Building an AI ecosystem in a small nation: Lessons from Singapore's journey to the forefront of ai, *Humanities and Social Sciences Communications*, 11(1). doi:10.1057/s41599-024-03289-7.
- [20] Huiyun Jing et al (2021) *J. Phys.: Conf. Ser.* 1948 012004, Saatavilla: <https://doi.org/10.1088/1742-6596/1948/1/012004>.
- [21] Stoica, A., Ghenade, A. & Pica, A.S. (2024), The Impact of Artificial Intelligence on Cyber Security, *FAIMA Business & Management Journal*, vol. 12, no. 3, pp. 5-14.
- [22] Zbořil, M. (2024). SECURITY RISKS ASSOCIATED WITH DEPLOYMENT OF AI SOLUTIONS INTO ORGANIZATIONS. IDIMT-2024: Changes to ICT, Management, and Business Processes through AI: 32nd Interdisciplinary Information Management Talks Sept. 4–6, 2024 Hradec Králové, Czech Republic, pp.65–72. Saatavilla: <https://doi.org/10.35011/IDIMT-2024-65>.
- [23] Sobrino-García, I. (2021). Artificial Intelligence Risks and Challenges in the Spanish Public Administration: An Exploratory Analysis through Expert

- Judgements. *Administrative Sciences*, 11(3), 102. Saatavilla: <https://doi.org/10.3390/admsci11030102>.
- [24] Zhang, J. et al. (2025) When LLMs meet cybersecurity: A systematic literature review, *Cybersecurity*, 8(1). doi:10.1186/s42400-025-00361-w.
- [25] Fox, S. (2024). Adaptive AI Alignment: Established Resources for Aligning Machine Learning with Human Intentions and Values in Changing Environments. *Machine Learning and Knowledge Extraction*, 6(4), 2570-2600. Saatavilla: <https://doi.org/10.3390/make6040124>.
- [26] Malatji, M. and Tolah, A. (2024) Artificial Intelligence (AI) cybersecurity dimensions: A comprehensive framework for understanding adversarial and offensive ai, *AI and Ethics*, 5(2), pp. 883–910. Saatavilla: 10.1007/s43681-024-00427-4.
- [27] Hutter, R., & Hutter, M. (2021). Chances and Risks of Artificial Intelligence—A Concept of Developing and Exploiting Machine Intelligence for Future Societies. *Applied System Innovation*, 4(2), 37. Saatavilla: <https://doi.org/10.3390/asi4020037>.
- [28] Aliman, N.-M., Kester, L., & Yampolskiy, R. (2021). Transdisciplinary AI Observatory—Retrospective Analyses and Future-Oriented Contradistinctions. *Philosophies*, 6(1), 6. Saatavilla: <https://doi.org/10.3390/philosophies6010006>.
- [29] Narula, S. et al. (2025) Exploring research and tools in AI Security: A systematic mapping study, *IEEE Access*, 13, pp. 84057–84080. doi:10.1109/access.2025.3567195.
- [30] Singh K., Saxena R. and Kumar B., (2024). AI Security: Cyber Threats and Threat-Informed Defense, 2024 8th Cyber Security in Networking Conference (CSNet), Paris, France, pp. 305-312, doi: 10.1109/CSNet64211.2024.10851770.
- [31] Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024). Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective. *Informatics*, 11(3), 58. Saatavilla: <https://doi.org/10.3390/informatics11030058>.
- [32] Bodini, M. (2024). Generative Artificial Intelligence and Regulations: Can We Plan a Resilient Journey Toward the Safe Application of Generative Artificial Intelligence? *Societies*, 14(12), 268. Saatavilla: <https://doi.org/10.3390/soc14120268>.
- [33] Fauzi, Rokhman & Sembiring, Jaka. (2023). A Review on Information Security Risk Assessment of Smart Systems: Risk Landscape, Challenges, and Prospective Methods. pp. 1-6. doi: 10.1109/ICISS59129.2023.10291306.
- [34] Steimers, A., & Schneider, M. (2022). Sources of Risk of AI Systems. *International Journal of Environmental Research and Public Health*, 19(6), 3641. Saatavilla: <https://doi.org/10.3390/ijerph19063641>.
- [35] Islas, O., Gutiérrez-Cortés, F., & Arribas-Urrutia, A. (2024). A Look at the Risks and Threats of Artificial Intelligence, From Media Ecology. *Comunicar*, 79, pp.1-9. Saatavilla: <https://doi.org/10.58262/V33279.1>.

- [36] Cha, S. (2024) Towards an international regulatory framework for AI Safety: Lessons from the IAEA's nuclear safety regulations, *Humanities and Social Sciences Communications*, 11(1). doi:10.1057/s41599-024-03017-1.
- [37] Euroopan komissio (2024). Tekoälysäädös tulee voimaan. [online] Euroopan komissio. Saatavilla: https://commission.europa.eu/news-and-media/news/ai-act-enters-force-2024-08-01_fi
- [38] digital-strategy.ec.europa.eu. (2024). Tekoälysäädös | Shaping Europe's digital future. [online] Saatavilla: <https://digital-strategy.ec.europa.eu/fi/policies/regulatory-framework-ai>.
- [39] ANNEX to the Communication to the Commission Approval of the content of the draft Communication from the Commission - Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act). (2025). [online] Shaping Europe's digital future. Brussels: European Commission. Saatavilla: <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act>.
- [40] Euroopan unioni. (2016). Euroopan parlamentin ja neuvoston asetus (EU) 2016/679 luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä ja näiden tietojen vapaasta liikkuvuudesta sekä direktiivin 95/46/EY kumoamisesta (yleinen tietosuoja-asetus). Euroopan unionin virallinen lehti, L 119, 4.5.2016.
- [41] Tekoälyasetus -Työ- ja elinkeinoministeriö. (2024). EU:n tekoälyasetuksen kansallinen toimeenpano. [online] Saatavilla: <https://tem.fi/tekoalyasetus>.

LIITE I: KIRJALLISUUSLUETTELOT

Tekoälyn yleisimmät riskit ja niiden hallinta

Aliman, N.-M., Kester, L., & Yampolskiy, R. (2021). Transdisciplinary AI Observatory—Retrospective Analyses and Future-Oriented Contradistinctions. *Philosophies*, 6(1), 6. doi: <https://doi.org/10.3390/philosophies6010006>.

Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024). Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective. *Informatics*, 11(3), 58. doi: <https://doi.org/10.3390/informatics11030058>.

Altun, İ.Z. and Emre Özkök, A. (2024) Securing Artificial Intelligence: Exploring Attack Scenarios and defense strategies, 2024 12th International Symposium on Digital Forensics and Security (ISDFS), pp. 1–6. doi:10.1109/isdfs60797.2024.10527288.

Bodini, M. (2024). Generative Artificial Intelligence and Regulations: Can We Plan a Resilient Journey Toward the Safe Application of Generative Artificial Intelligence? *Societies*, 14(12), 268. doi: <https://doi.org/10.3390/soc14120268>.

Calzada, I., Németh, G., & Al-Radhi, M. S. (2025). Trustworthy AI for Whom? GenAI Detection Techniques of Trust Through Decentralized Web3 Ecosystems. *Big Data and Cognitive Computing*, 9(3), 62. doi: <https://doi.org/10.3390/bdcc9030062>.

Cha, S. (2024) Towards an international regulatory framework for AI Safety: Lessons from the IAEA's nuclear safety regulations, *Humanities and Social Sciences Communications*, 11(1). doi:10.1057/s41599-024-03017-1.

Fauzi, Rokhman & Sembiring, Jaka. (2023). A Review on Information Security Risk Assessment of Smart Systems: Risk Landscape, Challenges, and Prospective Methods. pp. 1-6. doi: 10.1109/ICISS59129.2023.10291306.

Fox, S. (2024). Adaptive AI Alignment: Established Resources for Aligning Machine Learning with Human Intentions and Values in Changing Environments. *Machine Learning and Knowledge Extraction*, 6(4), 2570-2600. doi: <https://doi.org/10.3390/make6040124>.

Hashmi, E., Yamin, M.M. and Yayilgan, S.Y. (2024) Securing Tomorrow: A comprehensive survey on the Synergy of Artificial Intelligence and Information Security, *AI and Ethics*, 5(3), pp. 1911–1929. doi:10.1007/s43681-024-00529-z.

- Hernández-Orallo, J. et al. (2020) 'Ai paradigms and AI safety: Mapping artefacts and techniques to safety issues', *Frontiers in Artificial Intelligence and Applications, ECAI 2020*(325), pp. 2521–2528. doi:10.3233/faia200386.
- Huiyun Jing et al (2021) *J. Phys.: Conf. Ser.* 1948 012004, doi: <https://doi.org/10.1088/1742-6596/1948/1/012004>.
- Hutter, R., & Hutter, M. (2021). Chances and Risks of Artificial Intelligence—A Concept of Developing and Exploiting Machine Intelligence for Future Societies. *Applied System Innovation*, 4(2), 37. doi: <https://doi.org/10.3390/asi4020037>.
- Islas, O., Gutiérrez-Cortés, F., & Arribas-Urrutia, A. (2024). A Look at the Risks and Threats of Artificial Intelligence, From Media Ecology. *Comunicar*, 79, pp.1-9. doi: <https://doi.org/10.58262/V33279.1>.
- Iyer, N. (2022) Adversarial machine learning: Exploring security vulnerabilities in AI-Driven Systems, *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 3(1), pp. 1–9. doi: 10.63282/3050-9262.ijaidsm-l-v3i1p101.
- Kafali, E., Preuveneers, D., Semertzidis, T., & Daras, P. (2024). Defending Against AI Threats with a User-Centric Trustworthiness Assessment Framework. *Big Data and Cognitive Computing*, 8(11), 142. doi: <https://doi.org/10.3390/bdcc8110142>.
- Khanal, S., Zhang, H. and Taeihagh, A. (2024) Building an AI ecosystem in a small nation: Lessons from Singapore's journey to the forefront of ai, *Humanities and Social Sciences Communications*, 11(1). doi:10.1057/s41599-024-03289-7.
- Malatji, M. and Tolah, A. (2024) Artificial Intelligence (AI) cybersecurity dimensions: A comprehensive framework for understanding adversarial and offensive ai, *AI and Ethics*, 5(2), pp. 883–910. doi: 10.1007/s43681-024-00427-4.
- Narula, S. et al. (2025) Exploring research and tools in AI Security: A systematic mapping study, *IEEE Access*, 13, pp. 84057–84080. doi:10.1109/access.2025.3567195.
- Pascu, C. et al. (2023) Artificial Intelligence and cybersecurity research: Enisa Research and Innovation Brief, Enisa. Heraklion, Greece: ENISA. Saatavilla: <https://www.enisa.europa.eu/sites/default/files/publications/Artificial%20Intelligence%20and%20Cybersecurity%20Research.pdf>.
- Raimundo, R., & Rosário, A. (2021). The Impact of Artificial Intelligence on Data System Security: A Literature Review. *Sensors*, 21(21), 7029. doi: <https://doi.org/10.3390/s21217029>.

- Singh K., Saxena R. and Kumar B., (2024). AI Security: Cyber Threats and Threat-Informed Defense, 2024 8th Cyber Security in Networking Conference (CSNet), Paris, France, pp. 305-312, doi: 10.1109/CSNet64211.2024.10851770.
- Sobrinho-García, I. (2021). Artificial Intelligence Risks and Challenges in the Spanish Public Administration: An Exploratory Analysis through Expert Judgements. *Administrative Sciences*, 11(3), 102. doi: <https://doi.org/10.3390/admsci11030102>.
- Steimers, A., & Schneider, M. (2022). Sources of Risk of AI Systems. *International Journal of Environmental Research and Public Health*, 19(6), 3641. doi: <https://doi.org/10.3390/ijerph19063641>.
- Stoica, A., Ghenade, A. & Pica, A.S. (2024), The Impact of Artificial Intelligence on Cyber Security, *FAIMA Business & Management Journal*, vol. 12, no. 3, pp. 5-14.
- Suo, J., Li, M., Guo, J., & Sun, Y. (2024). Engineering Safety and Ethical Challenges in 2045 Artificial Intelligence Singularity. *Sustainability*, 16(23), 10337. doi: <https://doi.org/10.3390/su162310337>.
- Taherdoost, H., Le, T.-V., & Slimani, K. (2025). Cryptographic Techniques in Artificial Intelligence Security: A Bibliometric Review. *Cryptography*, 9(1), 17. doi: <https://doi.org/10.3390/cryptography9010017>.
- Villegas-Ch, W., & García-Ortiz, J. (2023). Toward a Comprehensive Framework for Ensuring Security and Privacy in Artificial Intelligence. *Electronics*, 12(18), 3786. doi: <https://doi.org/10.3390/electronics12183786>.
- Vähä-Sipilä, A., Marchal, S. and Aksela, M. (2021) Tekoälyn Soveltamisen kyber- Turvallisuus Ja Riskienhallinta, Tekoälyn soveltamisen kyberturvallisuus ja riskienhallinta. Saatavilla: https://www.traficom.fi/sites/default/files/media/publication/Tekoälyn_soveltamisen_kyberturvallisuus_ja_riskienhallinta.pdf.
- Zhang, H. et al. (2024) Preface: Security and safety in artificial intelligence, *Security and Safety*, 3: E2024021. doi:10.1051/sands/2024021.
- Zhang, J. et al. (2025) When LLMS meet cybersecurity: A systematic literature review, *Cybersecurity*, 8(1). doi:10.1186/s42400-025-00361-w.
- Zbořil, M. (2024). SECURITY RISKS ASSOCIATED WITH DEPLOYMENT OF AI SOLUTIONS INTO ORGANIZATIONS. IDIMT-2024: Changes to ICT, Management, and Business Processes through AI: 32nd Interdisciplinary Information Management Talks Sept. 4–6, 2024 Hradec Králové, Czech Republic, pp.65–72. doi: <https://doi.org/10.35011/IDIMT-2024-65>.

Euroopan Unionin tekoälyasetuksen vaatimukset

ANNEX to the Communication to the Commission Approval of the content of the draft Communication from the Commission - Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act). (2025).

[online] Shaping Europe's digital future. Brussels: European Commission. Saatavilla: <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act>.

digital-strategy.ec.europa.eu. (2024). Tekoälysäädös | Shaping Europe's digital future. [online] Saatavilla: <https://digital-strategy.ec.europa.eu/fi/policies/regulatory-an-artifi-ai>.

Euroopan komissio (2024). Tekoälysäädös tulee voimaan. [online] Euroopan komissio. Saatavilla: https://commission.europa.eu/news-and-media/news/ai-act-enters-force-2024-08-01_fi.

Euroopan komissio (n.d.). Tekoälyä koskeva eurooppalainen lähestymistapa. [online] Shaping Europe's digital future. Saatavilla: <https://digital-strategy.ec.europa.eu/fi/policies/european-approach-artificial-intelligence>.

Euroopan komissio (n.d.). Yleiskäyttöiset tekoälyn käytäntesäännöt. [online] Shaping Europe's digital future. Saatavilla: <https://digital-strategy.ec.europa.eu/fi/policies/ai-code-practice>.

Euroopan parlamentin ja neuvoston asetus (EU) 2024/1689, annettu 13 päivänä kesäkuuta 2024, tekoälyä koskevista yhdenmukaistetuista säännöistä ja asetusten (EY) N:o 300/2008, (EU) N:o 167/2013, (EU) N:o 168/2013, (EU) 2018/858, (EU) 2018/1139 ja (EU) 2019/2144 sekä direktiivien 2014/90/EU, (EU) 2016/797 ja (EU) 2020/1828 muuttamisesta (tekoälysäädös) (ETA:n kannalta merkityksellinen teksti). Euroopan unionin virallinen lehti, L-sarja 2024/1689, 12.7.2024. ELI: <http://data.europa.eu/eli/reg/2024/1689/oj>.