TURUN YLIOPISTO

Euroopan unionin osarahoittama

Tekoälysäädös ja siihen valmistautuminen: Näin pääset alkuun

EU:n tekoälysäädöksen valmistelu aloitettiin vuonna 2021. Säädös astui voimaan viimein elokuussa 2024 pitkien poliittisten neuvottelujen tuloksena. Säädös asettaa tekoälyjärjestelmien kehittäjille ja käyttöönottajille vaatimuksia sen mukaan, mihin säädöksen riskiluokkaan järjestelmä kuuluu ja mikä organisaation rooli on suhteessa järjestelmään.

Laki on monin paikoin monimutkainen, ja odotammekin vielä EU:lta tulkintaohjeita sekä standardeja, joiden avulla säädösten vaatimuksiin voidaan tosiasiassa tarttua. Ensimmäiset asteittain voimaan tulevista vaatimuksista vaatimukset kuitenkin koskevat organisaatioita jo helmikuussa 2025, joten nyt on aika aloittaa valmistautuminen.

Tämän oppaan tarkoituksena on auttaa organisaatioita alkuun tekoälysäädöksen kanssa. Tämän oppaan avulla saat käsityksen siitä, millaisia vaatimuksia tekoälysäädös tuo ja kenelle, mistä valmistautuminen kannattaa aloittaa ja millaista työtä vaatimusten täyttäminen lopulta vaatii. Tarvitset työssä välillä myös tukea juristilta, joka tietää toimintaympäristösi erityispiirteet. Tästä pääset kuitenkin alkuun.

1. Tunnista, onko järjestelmä tekoälyä

Ensimmäinen askel on tunnistaa, onko käytössäsi järjestelmä, joka lasketaan EU:n tekoälysäädöksen mukaan tekoälyksi. Yrityksillä on nykyisellään taipumus mainostaa tekoälynä järjestelmiä myös silloin, kun ne perustuvat edistyneeseen analytiikkaan.

Tekoälysäädöksen mukaan tekoälyjärjestelmänä pidetään:

> *“konepohjaista järjestelmää, joka on suunniteltu toimimaan käyttöönoton jälkeen vaihtelevilla autonomian tasoilla ja jossa voi ilmetä mukautuvuutta käyttöönoton jälkeen ja joka päättelee vastaanottamastaan syötteestä eksplisiittisiä tai implisiittisiä tavoitteita varten, miten tuottaa tuotoksia, kuten ennusteita, sisältöä, suosituksia tai päätöksiä, jotka voivat vaikuttaa fyysisiin tai virtuaalisiin ympäristöihin”* (Artikla 3.1).

Vaikka tekoälysäädös ei sitä suoranaisesti vaadi, on tekoälyjärjestelmät hyvä kirjata esimerkiksi tekoälyrekisteriin. Näin organisaatiosi on helppo pitää kirjaa siitä, mitä tekoälytyökaluja teillä on käytössä, mihin tekoälysäädöksen kategoriaan ne menevät, ja millaisia hallintatoimenpiteitä niiden eteen on jo suunniteltu tai tehty. Rekisteriin kannattaa merkitä myös järjestelmät, jotka eivät ole tekoälyä, jos määritelmä ei ole ollut täysin selvä. Näin voitte kirjata ylös myös perusteet sille, miksi järjestelmää ei ole katsottu säädöksen mukaiseksi tekoälyksi. Tästä on hyötyä, jos työkalun kohdalla on myöhemmin epäselvyyksiä.

2. Tunnista organisaation rooli suhteessa tekoälyyn

Tekoälysäädös määrittelee erilaisia vaatimuksia organisaatioille sen mukaan, mikä heidän roolinsa on suhteessa tekoälyjärjestelmään. Oman organisaation roolin tunnistaminen on tärkeää, jotta osaat kohdentaa resurssisi vain niihin vaatimuksiin, jotka todella koskevat sinun organisaatiotasi. Keskeisimpiä rooleja ovat:

A. **Tarjoajat** (*providers*), eli ne tahot, jotka kehittävät (tai kehityttävät) tekoälyjärjestelmän ja tuovat sen markkinoille tai omaan käyttöönsä. Tarjoajan roolilta ei voi välttyä tilaamalla sitä alihankkijalta, jos sen kuitenkin tuo markkinoille omalla brändillä varustettuna. Tarjoajan roolin voi myös saada, jos toisen kehittämään järjestelmään tekee merkittäviä muutoksia, tai käyttää sitä tarkoitukseen, johon alkuperäinen tarjoaja ei ole järjestelmää kehittänyt. Tekoälysäädöksen vaatimuksista suurin osa kohdistuu tekoälyjärjestelmien tarjoajiin.

B. **Käyttöönottajat** (*deployers*), eli ne tahot, jotka käyttävät valvonnassaan olevaa tekoälyjärjestelmää ammattitoiminnassa. Osa tekoälysäädöksen vaatimuksista koskee myös käyttöönottajia.

Lisäksi säädös määrittelee muita toimitusketjuun kuuluvia rooleja (esim. jakelija, maahantuoja, valtuutettu edustaja). Jos epäilet, että sinulla on tekoälyn toimitusketjussa jokin muu kuin tarjoajan tai käyttöönottajan rooli, löydät tarkat määritelmät tekoälysäädöksen artiklasta 3.

Mieti siis, onko työkalu sellainen, että se kulkee oman organisaatiosi nimellä tai brändillä, tai teetkö siihen merkittäviä muutoksia – joko itse työkaluun tai sen käyttötarkoitukseen. Jos, niin olet hyvin todennäköisesti järjestelmän tarjoaja. Jos taas käytät tai tuot työntekijöidesi käyttöön selkeästi valmiin tuotteen sen alkuperäisessä tarkoituksessa, olet hyvin todennäköisesti käyttöönottaja.

Kummassakin tapauksessa seuraava askel on varmistaa, että täytät rooliisi kuuluvat tekoälysäädöksen vaatimukset. Ne riippuvat järjestelmän riskiluokasta, jonka tunnistamiseen menemme seuraavaksi.

3. Tunnista tekoälysovelluksen riskiluokka

Tekoälysäädös jakaa tekoälyjärjestelmät riskiluokkiin sen perusteella, millaisia riskejä ne aiheuttavat ihmisten terveydelle, turvallisuudelle ja perusoikeuksille.

1. **Ei hyväksyttävissä oleva riski**. Jos tekoäly aiheuttaa ihmisille uhkaa, jota ei voida pitää hyväksyttävänä, sen käyttö ja kehittäminen on kielletty. Tällaisia järjestelmiä ovat esimerkiksi
    - yksilöiden tai heikossa asemassa olevien ihmisten ryhmiin kuuluvien manipuloiminen
    - sosiaalinen pisteytys, joka jakaa ihmisiä ryhmiin käytöksen, sosioekonomisen aseman tai henkilökohtaisten ominaisuuksien perusteella
    - ihmisten biometrinen tunnistaminen ja luokittelu
    - reaaliaikaista ja etäisyydeltä tapahtuvaa biometristä tunnistusta tekevät järjestelmät (esim. kasvontunnistus)

    Tarkka lista kielletyistä tekoälyjärjestelmistä löytyy artiklasta 5.

2. **Korkean riskin järjestelmät.** Tähän kategoriaan lasketaan sellaiset, jotka vaikuttavat negatiivisesti ihmisten turvallisuuteen ja perusoikeuksiin. Tällaisia ovat esimerkiksi
    - EU:n tuoteturvallisuusdirektiivin alle kuuluvat tuotteet (lelut, ilmailu, autot, lääkinnälliset laitteet, hissit)
    - Kriittisen infrastruktuurin hallinta ja käsittely
    - Yleissivistävä ja ammatillinen koulutus
    - Työllistäminen, henkilöstöhallinto ja mahdollisuus itsenäiseen ammatinharjoittamiseen
    - Keskeisten yksityisten palvelujen, julkisten palvelujen ja etuusien saatavuus ja käyttö
    - lainvalvonta
    - muuttoliikkeen hallinta, turvapaikka-asiat ja rajavalvonta
    - oikeudenhoito ja demokraattiset prosessit

    Tarkka kuvaus korkeariskisistä järjestelmistä löytyy luvun III jaksosta 1 ja tekoälysäädöksen liitteestä III, johon listatut järjestelmät ovat kaikki korkeariskisiä.

3. **Järjestelmät, joihin kohdistuu avoimuusvaatimuksia.** Näiden järjestelmien riskit on katsottu maltillisiksi, mutta kuitenkin sellaisiksi, että niiden läpinäkyvyyteen kohdistuu erityisiä vaatimuksia. Tällaisia järjestelmiä ovat generoivat

tekoälyjärjestelmät, jotka imitoivat ihmisen tuottamaa sisältöä, kuten tekstiä, kuvia, videoita ja ääntä. Näiden tarkempi kuvaus löytyy säädöksen artiklasta 50.

4. **Yleiskäyttöiset tekoälymallit**. Tekoälysäädös määrittelee lisäksi vaatimuksia yleiskäyttöisille tekoälymalleille. Nämä koskevat mallien tarjoajia. Vaatimukset löytyvät säädöksen luvusta IV.

5. **Muut**. Jos järjestelmä ei kuulu mihinkään tekoälysäädöksen kategorioista, siihen ei kohdistu tekoälysäädöksen osalta vaatimuksia. Huomaathan, että muut lainsäädännön vaatimukset (esim. GDPR) koskee kuitenkin tekoälyä siinä missä muitakin niiden piiriin kuuluvia tietojärjestelmiä. Muut tekoälyjärjestelmät eivät siis suinkaan ole täysin sääntelemättömiä, vaikka niille ei juuri tekoälysäädöksessä ole asetettu vaatimuksia.

4. Tunnista sinua koskevat vaatimukset

Nyt kun tiedät oman organisaatiosi roolin ja järjestelmäsi riskiluokan, voit katsoa sinua koskevat vaatimukset. Alla oleva taulukko tiivistää säädöksen keskeiset vaatimukset tarjoajille ja käyttöönottajille eri riskiluokkien mukaisesti. Huom. tarkista sinua koskevat vaatimukset tekoälysäädöksestä oman juristisi kanssa. Tämän taulukon tarkoituksena on antaa vaatimuksista yleiskatsaus, joten se ei sisällä kaikkia yksityiskohtia.

Monet vaatimuksista ovat normaaleja järjestelmäkehityksen hyviä käytänteitä, joihin on kuitenkin tekoälyn kanssa kiinnitettävä erityistä huomiota.

Riskiluokka	**Tarjoaja**	**Käyttöönottaja**
Kielletty (Luku II)	Ei saa valmistaa	Ei saa ottaa käyttöön
Korkeariskinen **(Luku III, jakso 3)**	• Riskienhallinta-järjestelmä, joka kattaa koko tekoälyn elinkaaren • Datanhallinta-järjestelmä, joka huolehtii datan olennaisuudesta ja edustavuudesta sekä virheiden ehkäisystä • Laadunhallintajärjestelmä lainmukaisuuden varmistamiseksi • Riittävä tekninen dokumentaatio • Automaattisesti tuotetut lokitiedot • Käyttöohjeet käyttöönottajille	• Huolehdittava ihmisten valvonnasta (*human ovesight*) • Huolehdittava syötettävän datan edustavuudesta (vinoumien välttäminen) • Noudatettava tarjoajan antamia käyttöohjeita • Automaattiset lokitiedot • Ilmoittaminen korkeariskisen tekoälyn käytöstä käytön kohteille, jos käyttöönottaja on työnantaja ja käytön kohteet ovat työntekijöitä

	• Korjaavat toimenpiteet ja ilmoitusvelvollisuus, kun puutteita havaitaan • Vaatimustenmukaisuuden arviointimenettely ennen markkinoille saattamista tai käyttöönottoa • CE-merkintä	• Julkisen sektorin toimijoiden on rekisteröitävä järjestelmä tekoälyrekisteriin • Perusoikeusvaikutusten arviointi, jos käyttöönottaja on julkinen organisaatio tai tarjoaa julkisia palveluita
Läpinäkyvyysvaatimusten alainen (Luku IV)	• Käyttäjille ilmoitettava, että he ovat tekemisissä tekoälyn kanssa (tämä huomioitava tekoälyjärjestelmän designissa) • Tekoälyn tuotokset ja muokkaukset merkittävä koneellisesti luettavassa muodossa, jotta ne voidaan tunnistaa koneen tuottamiksi	• Tunteentunnistus-järjestelmän tai biometrisen luokitusjärjestelmän käyttöönottajien on ilmoitettava järjestelmän toiminnasta niille luonnollisille henkilöille, jotka altistuvat järjestelmälle • Kuva-, ääni- tai videosisältöä tuottavien tai manipuloivien tekoälyjärjestelmien sisältö on merkittävä tekoälyllä tuotetuksi/muokatuksi • Tekoälyteksti on merkittävä tekoälyn tuottamaksi, jos tekstin tarkoitus on tiedottaa yleisölle yleistä etua koskevista asioista.
Yleiskäyttöiset tekoälymallit (Artikla 53–55)	• Riittävä ja ajantasainen tekninen dokumentaatio, joka on tarjolla mallien soveltajille: annettava tekoälyjärjestelmien tarjoajille hyvä käsitys yleiskäyttöisen tekoälymallin valmiuksista ja rajoituksista ja autettava niitä noudattamaan tekoälyasetuksen velvoitteita • laadittava ja asetettava julkisesti saataville tiivistelmä yleiskäyttöisen tekoälymallin koulutuksessa käytetystä sisällöstä tekoälytoimiston toimittaman mallin mukaisesti Lisäksi systeemisen riskin omaavien yleiskäyttöisten mallien tarjoajien on • suoritettava standardienmukainen riskiarviointi • arvioitava ja lievennettävä systeemisiä riskejä • seurattava, dokumentoitava viranomaisille vakavat vaaratilanteet ja niihin tehdyt toimenpiteet • varmistettava kyberturvallisuus mallille ja sen fyysiselle infrastruktuurille	Ei vaatimuksia

Säädökseen valmistautuminen vaatii siis toimia. Miten niitä kannattaa priorisoida?

Tekoälysäädös tulee voimaan vaiheittain, joten kaikkea ei tarvitse tehdä kerralla. Tekeminen kannattaa aloittaa tunnistamalla, missä ovat oman organisaation vahvuusalueet ja missä kehittämistä tarvitaan vielä nykyistä enemmän. Tätä varten voit käyttää esimerkiksi tästä oppaasta löytyvää maturiteettimallia.

Keskeisimmät vaatimukset astuvat voimaan tässä seuraavalla aikataululla:

1) **Helmikuu 2025**: Kiellettyjen järjestelmien kielto ja tekoälylukutaidon vaatimus
2) **Elokuu 2025**: Yleiskäyttöisten tekoälymallien tarjoajiin kohdistuvat vaatimukset
3) **Elokuu 2026**: Generatiivisen tekoälyn läpinäkyvyysvaatimukset tarjoajille ja käyttöönottajille
4) **Elokuu 2027**: Korkean riskin järjestelmien vaatimukset tarjoajille ja käyttöönottajille.

Kaiken on siis oltava valmista ja implementoitu viimeistään elokuussa 2027. Aikaa siis on, mutta sitä myös tarvitaan. Korkeariskisten järjestelmien tarjoaminen ja käyttöönotto vaativat toimenpiteitä, joita varten prosessit ja vastuut on oltava **määritelty** sekä **implementoitu** määräpäivään mennessä. Pelkkä suunnitelma ja ajatus ei siis riitä, vaan toimenpiteiden pitää olla aktiivisessa käytössä. Uusien prosessien jalkauttaminen vaatii usein aikaa ja huolellista suunnittelua, joten jos näköpiirissä on edes mahdollisuus korkeariskisille järjestelmille (esim. tekoälyn käyttö rekrytoinnissa tai työntekijöiden johtamisessa), valmistautuminen kannattaa aloittaa mahdollisimman pian.

Ensimmäisenä astuvat voimaan säädöksen yleiset vaatimukset, kuten vaatimus riittävästä tekoälylukutaidosta.

Tekoälylukutaito

> *“Tekoälyjärjestelmien tarjoajien ja käyttöönottajien on parhaansa mukaan toteutettava toimenpiteitä, joilla ne varmistavat henkilöstönsä ja muiden niiden puolesta tekoälyjärjestelmien toiminnasta ja käytöstä vastaavien henkilöiden riittävän tekoälylukutaidon, minkä yhteydessä otetaan huomioon heidän tekninen tietämyksensä, kokemuksensa, koulutuksensa ja tekoälyjärjestelmien käyttöyhteys sekä henkilöt tai henkilöryhmät, joihin tekoälyjärjestelmiä on määrä käyttää.” (Artikla 4)*

Jokaisen organisaation, joka kehittää tai ottaa käyttöön tekoälyä, on siis varmistettava henkilöstölleen riittävät tekoälytaidot. Määritelmä riittävälle tasolle ei ole nykyisellään

kuitenkaan kovinkaan tarkka, eikä tulkintaohjeita ole vielä saatavilla. Tärkeintä onkin tällä hetkellä miettiä: **millaisella koulutuksella voisimmen varmistaa, että tekoälyä työssään käyttävillä tai sen parissa työskentelevillä on riittävät edellytykset noudattaa tekoälysäädöksen vaatimuksia**?

Kaikkien ei siis tarvitse osata koodata tekoälyjärjestelmiä, tai edes ymmärtää työkalujen teknisiä yksityiskohtia syvällisesti. Kehittäjiltä vaaditaan parempaa teknistä osaamista ja kykyä valita sopivat, ihmisten perusoikeuksia kunnioittavat tekoälymallit, kun taas esimerkiksi generatiivisen tekoälyn käyttäjiltä vaaditaan taitoa tunnistaa järjestelmien haasteita ja rajallisuuksia, ja hahmottaa niiden toiminnan perusperiaatteet. Näillä taidoilla on siis tarkoitus varmistaa, että kullakin on heidän omaan rooliinsa nähden mahdollisuus kehittää ja käyttää tekoälyä eettisesti ja lainmukaisesti.

On myös oletettavaa, että korkeamman riskitason ratkaisuja käytettäessä vaatimukset tekoälylukutaidolle ovat korkeammat kuin matalariskisiä järjestelmiä käytettäessä.

Pohdi seuraavia kysymyksiä:

1) Millaista osaamista organisaatiossamme on tekoälyn osalta?
2) Millaisiin käyttötarkoituksiin tekoälyä meillä käytetään?
3) Olemmeko tarjonneet tekijöillemme riittävän osaamisen ja valmiuden, jotta he tietävät, mihin tarkoituksiin tekoälyä kannattaa ja ei kannata käyttää, miten arvioida kriittisesti tekoälyn tuotoksia, ja miten varmistaa, että tekoälyä käytetään lain mukaisesti?

Muista myös dokumentoida toimet, joita olette tehneet tekoälylukutaidosta huolehtimiseksi, jotta voit tarvittaessa näyttää organisaatiosi tehneen parhaansa tekoälylukutaidon varmistamiseksi ja siten tekoälysäädöksen vaatimuksen täyttämiseksi. Lisää tulkintaohjeita ja parhaita käytäntöjä tekoälylukutaidon saavuttamiseksi on odotettavissa EU:n tekoälytoimistolta vielä lähiaikoina.

Voit käyttää apuna myös tästä oppaasta löytyviä itsearviointityökaluja valmistautuessasi tekoälysäädökseen.

Cheat sheet:

1) Tunnista tekoälyjärjestelmät.
2) Tunnista oman organisaatiosi rooli.
3) Tunnista käytössä ja kehitteillä olevien tekoälyjärjestelmien riskitaso

4) Tekoälylukutaito: Hahmota oman organisaatiosi työntekijöiden roolit tekoälyn kehittämisessä ja käytössä ja osaaminen, jota roolit vaativat tekoälyn lainmukaiseksi käyttämiseksi
5) Arvioi nykyinen kypsyystaso yllä olevilla osa-alueilla ja priorisoi askeleet kohti lainmukaisuutta
6) Aloita korkeimman prioriteetin prosesseista, pala kerrallaan.

Resursseja tekoälysäädöksen soveltamisen ja tekoälyn hallinnan tueksi:

- Kirja tekoälysäädöksen varhaisen tulkinnan tueksi: Hense & Mustac (2024) *AI Act Compact. Compliance, management & use cases in corporate practice*.
- GenAI Risk Assessment tool (GAIRA): https://www.vischer.com/en/knowledge/blog/gaira-now-with-an-ai-act-check-an-additional-dpia-and-in-german/
- Sitran työkalu digisääntelyn soveltamiseen (tulossa pian!): https://www.sitra.fi/hankkeet/tyokalut-digisaantelyn-soveltamiseen-liiketoiminnassa/

Muita:

- OECD.AI catalogue of tools and metrics: https://oecd.ai/en/catalogue/overview
- AI Standards hub (UK): https://aistandardshub.org/
- Alan Turing Institute's platform for AI Governance and Ethics in Practice: https://www.turing.ac.uk/research/research-projects/ai-ethics-and-governance-practice
- AI Risk Repository (MIT): https://airisk.mit.edu/
- European Data Protection Board's AI Auditing project deliverables: https://www.edpb.europa.eu/our-work-tools/our-documents/support-pool-experts-projects/ai-auditing_en