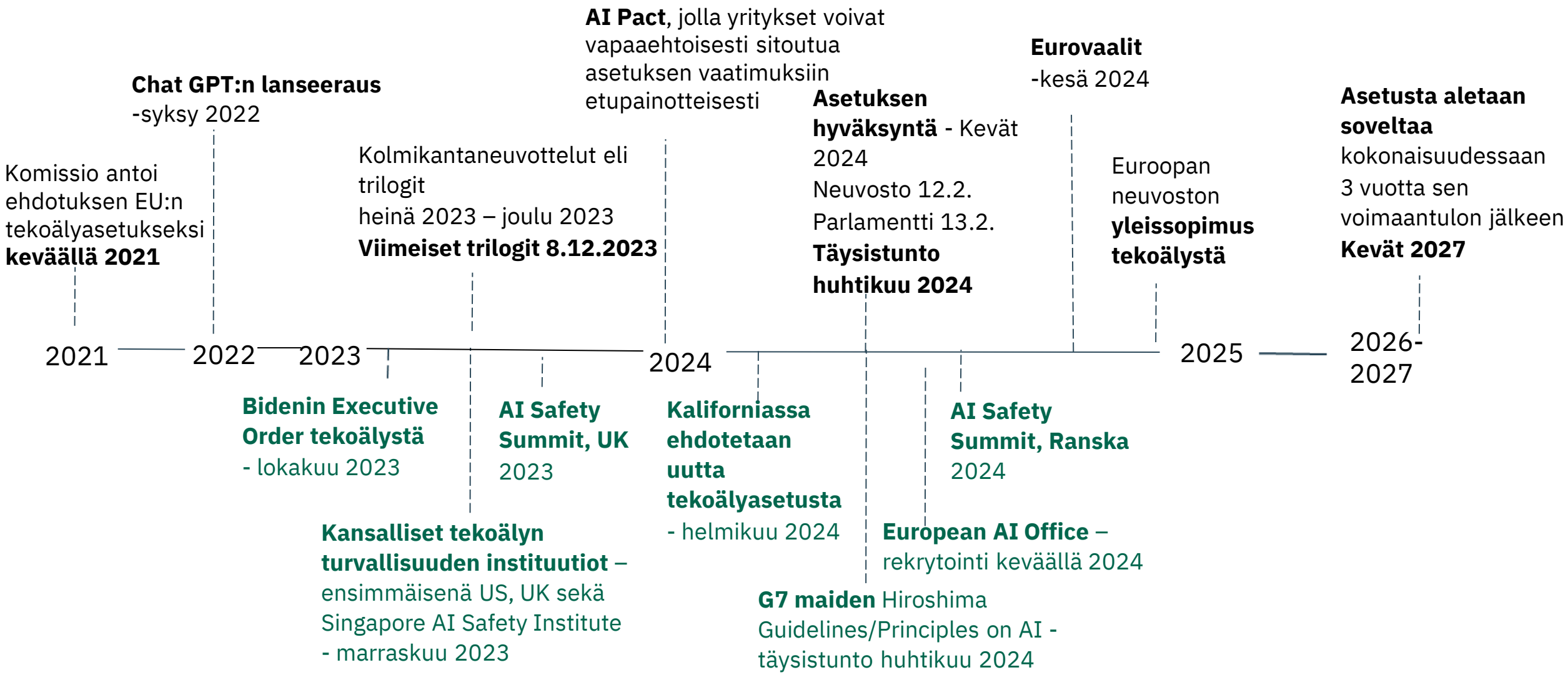
Peppiina Huhtala, asiantuntija, EK, puh. 045 678 1306
peppiina.huhtala@ek.fi

EU:n tekoälyasetus (AI Act) osa I

Elinkeinoelämän keskusliitto

Aikajana EU:n tekoälyasetus ja tekoälyn KV-sääntely

Tekoälyturvallisuus ja standardien kehitys kansainvälisessä kontekstissa



Mikä tekoälyasetus?

Tekoälyjärjestelmien turvallisuussäädös

- Tavoitteena on suojata terveyttä, turvallisuutta ja perusoikeuksia tekoälyjärjestelmien käytöstä aiheutuvilta riskeiltä ja haitoilta.
- Euroopasta ihmiskeskeisen tekoälyn kehittämiselle ja soveltamiselle suotuisa ja teknologisesti edistyksellinen, mutta samaan aikaan ihmisoikeuksia kunnioittava sekä turvallinen toimintaympäristö.
- Asettaa tekoälyjärjestelmille EU:n laajuiset säännöt, joita sovelletaan sellaisinaan kaikissa jäsenmaissa.
- Riskiperusteinen lähestymistapa.

Asetuksen soveltamisalaan eivät kuulu

→ **Tutkimus- ja kehitysvaiheessa olevat tekoälyjärjestelmät ja -mallit, joita ei ole vielä asetettu markkinoille tai otettu käyttöön.**

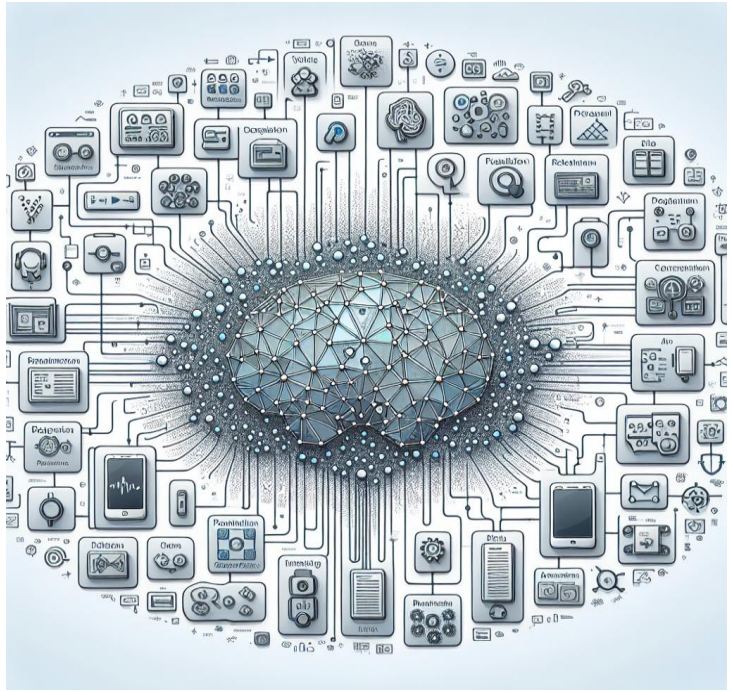
→ **Tekoälyjärjestelmät, jotka on tarkoitettu yksinomaan sotilaalliseen, puolustukselliseen tai kansallisen turvallisuuden käyttöön.**

→ **Avoimen lähdekoodin tekoälyjärjestelmät, jos kyseessä ei kielletyissä tai korkean riskin tapauksissa käytettävä AI-järjestelmä tai korkean riskin GPAI-malli.**

EU:n tekoälyasetus perustuu kahteen osioon

Tekoälyjärjestelmien sääntely käyttötarkoituksen riskiperusteisuuden mukaan

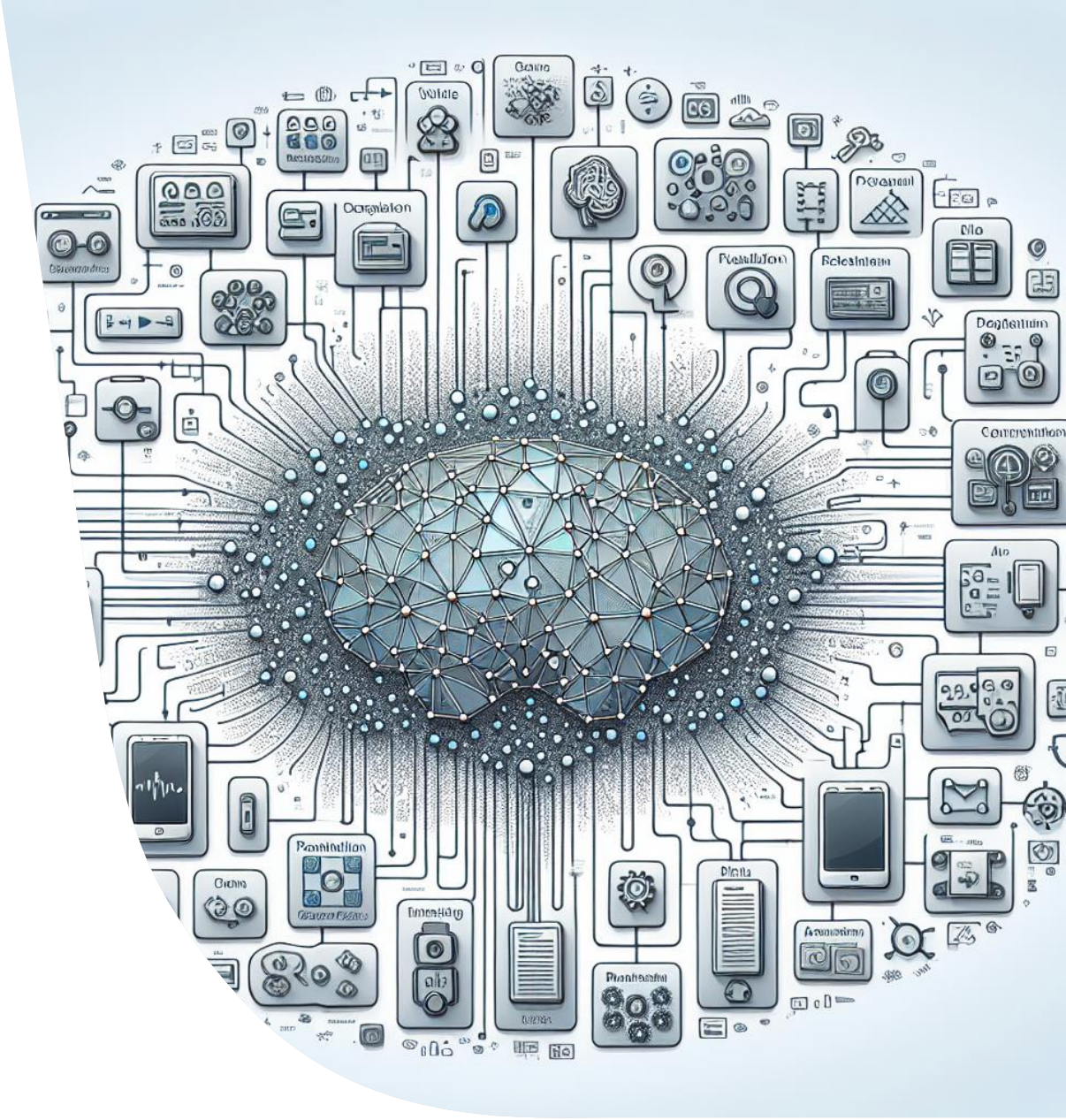
Yleiskäyttöisten tekoälymallien ja järjestelmäriskin sisältävien -mallien sääntely



Yleiskäyttöisen tekoälymallin määritelmä

“General-purpose AI model is an AI model, including when **trained with a large amount of data** using self-supervision at scale, that displays significant generality and **is capable to competently perform a wide range of distinct tasks** regardless of the way the model is released on the market and that can be integrated into a variety of downstream systems or applications.”

Yleiskäyttöinen tekoälymalli on tekoälymalli, joka on **koulutettu suurella määrällä dataa** käyttäen itseohjautuvaa oppimista mittakaavassa ja joka osoittaa merkittävää yleisyyttä ja **kykenee suorittamaan pätevästi laajan valikoiman erilaisia tehtäviä** riippumatta siitä, miten malli julkaistaan markkinoilla, ja joka **voidaan integroida erilaisiin järjestelmiin tai sovelluksiin.**



Kuva DALL-E x Peppiina

Yleiskäyttöiset tekoälymallit, jotka voivat aiheuttaa systeemisen riskin

*"**High-impact general-purpose AI models** that can cause systemic risk in the future, as well as on high-risk **AI systems**"*

- Järjestelmäriskillä tarkoitetaan muun muassa seuraavia riskejä:
  - Todelliset tai kohtuudella ennakoitavissa olevat kielteiset vaikutukset, jotka liittyvät suuronnettomuuksiin, kriittisten alojen häiriöihin ja vakaviin seurauksiin kansanterveydelle ja –turvallisuudelle.
  - Todelliset tai kohtuudella ennakoitavissa olevat kielteiset vaikutukset demokraattisiin prosesseihin, julkiseen ja taloudelliseen turvallisuuteen tai laittoman, väärän tai syrjivän sisällön levittäminen.
- Tekoälymalli kuuluu tähän luokkaan, jos
  - Sen koulutuksessa on käytetty laskentatehoa vähintään 10^25 FLOPs
  - Tai jos Euroopan tekoälytoimisto katsoo mallin muihin tekijöihin perustuen käsittävän systeemisiä riskejä (esim. käyttäjämäärä, autonomisuuden aste).

Tekoälyjärjestelmän määritelmä

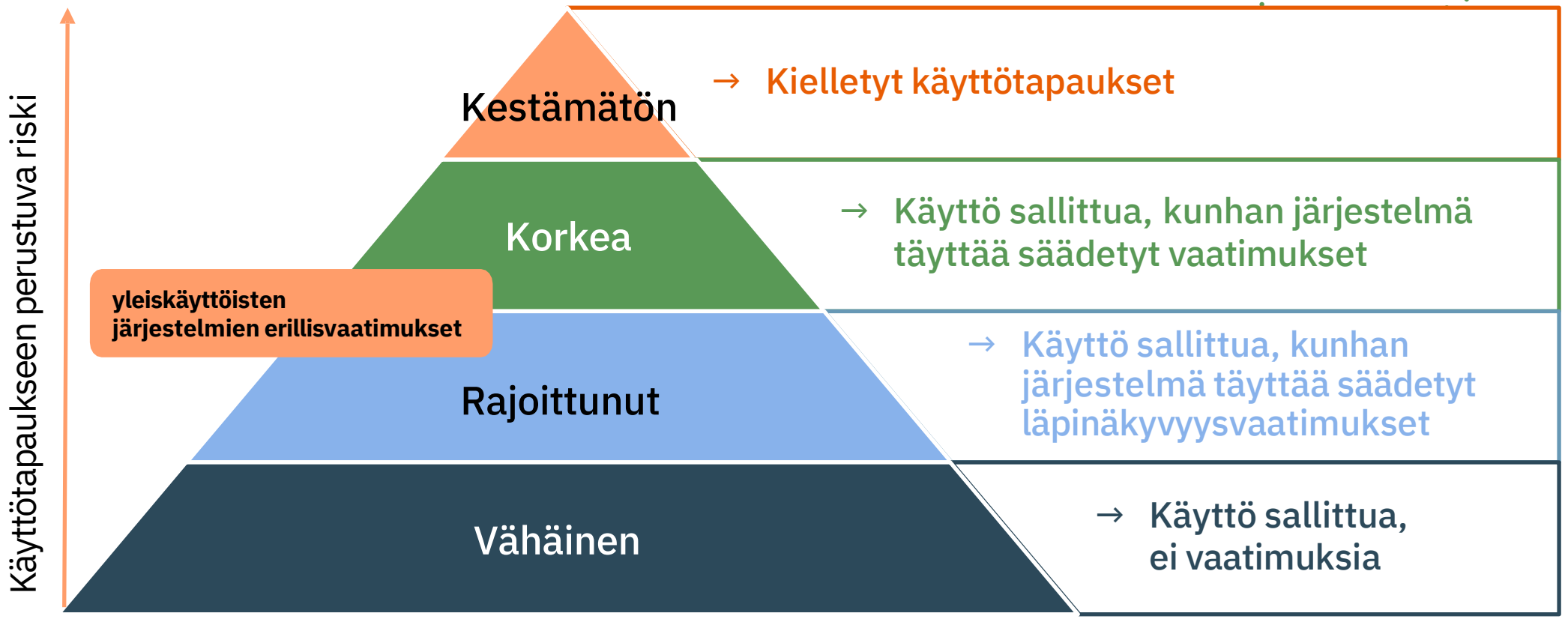
“Machine-based system that is **designed to operate with varying levels of autonomy,** and that can, for **explicit or implicit objectives,** generate outputs such as predictions, recommendations, or decisions that influence physical or virtual environments.”

Koneperusteinen järjestelmä, joka on **suunniteltu toimimaan eriasteisesti itsenäisesti** ja **joka voi olla mukautuva käyttöönoton jälkeen** ja joka nimenomaisia tai epäsuoria tavoitteita varten päättelee saamastaan syötteestä, miten se **voi tuottaa tuotoksia,** kuten ennusteita, sisältöä, suosituksia tai päätöksiä**, jotka voivat vaikuttaa fyysisiin tai virtuaalisiin ympäristöihin.**

→ **Trilogeissa päädyttiin 6.12.2023 OECD:n määritelmään!**

Kuva DALL-E x Peppiina

Riskiin perustuvat säännöt tekoälyjärjestelmille



Kuvio: Teknologiateollisuus ry

Kielletyt käyttötapaukset

- Haitalliset alitajuiset tekniikat, joilla manipuloidaan ihmisten käyttäytymistä
- Tunteiden tunnistusjärjestelmät työelämässä ja koulutuksessa
- Haavoittuvien ryhmien hyväksikäyttö (esim. vammaiset ja ikääntyneet ihmiset)
- Biometriset luokittelujärjestelmät
- Ennakoivan poliisitoiminnan sovellukset
- Sosiaalinen pisteytys, joka voi johtaa henkilöiden tai ryhmien epäsuotuisaan kohteluun
- Ei kohdennettu (untargeted) kasvojen kuvien kerääminen internetistä tai valvontakameran tallenteista kasvojentunnistustietokantojen luomiseksi
- Reaaliaikaiset biometriset etätunnistusjärjestelmät julkisissa tiloissa lainvalvontatarkoituksiin, **lukuun ottamatta seuraavia poikkeuksia**

Reaaliaikaisen biometrisen tunnistamisen sallitut lainvalvonnan käyttötapaukset

- Kohdennettu etsintä sieppauksen, ihmiskaupan ja seksuaalisen hyväksikäytön uhrien löytämiseksi sekä kadonneiden henkilöiden etsimiseksi.
- Henkeä tai fyysistä turvallisuutta uhkaavan erityisen, merkittävän ja välittömän uhan tai ennakoitavissa olevan terrori-iskun uhan torjuminen.
- Seuraavista vakavista rikoksesta epäillyn henkilön paikantaminen tai tunnistaminen rikostutkinnan suorittamiseksi:
  - **Terrorismi, ihmiskauppa, seksuaalinen hyväksikäyttö, murha, lapsikaappaus, aseellinen ryöstö, osallisuus rikollisjärjestöön, tietyt ympäristörikokset.**
  - **Vaatimuksena, että rikoksesta voi jäsenvaltiossa saada vähintään neljän vuoden vankeustuomion.**

Kuva DALL-E x Peppiina

Kuva DALL-E x Peppiina

Asetuksen velvoitteet, kuka vastaa?

- Velvoitteet riippuvat siitä, onko organisaatio tekoälyn markkinoille asettaja (tarjoaja) vai käyttääkö se jonkun toisen luomaa tekoälyä (käyttöönottaja).
- **Suurin osa asetuksen vaatimuksista kohdistuu tekoälyjärjestelmän tarjoajaan.**
- Erityisen tiukat velvoitteet kohdistuvat sellaisen laajan tekoälymallin kehittäjiin, johon esimerkiksi ChatGPT:kin perustuu (GPAI).
- Tiukat velvoitteet kohdistuvat myös tarjoajiin, joiden tuottamia tekoälyjärjestelmiä käytetään esimerkiksi sairaaloissa, työpaikoilla, oppilaitoksissa tai hallinnossa, kun päätetään ihmisille tärkeistä etuuksista (korkean riskin käyttöalue).
- Tarjoajan on tällöin varmistettava, että tekoäly toimii luotettavasti ja että käyttäjät ymmärtävät, mitä dataa tekoäly käyttää ja mitä tehdä, jos tekoälyjärjestelmä ei toimi toivotulla tavalla.