

Voivatko tekoäly ja avoimet sisällöt leikkiä kiltisti yhdessä?

**Näkökulmia tekoälyn, CC-lisenssien ja avointen oppimateriaalien yhteispelin
haasteisiin ja mahdollisuuksiin**

Tarmo Toikkanen

tarmo.toikkanen@sitra.fi

Johtava asiantuntija, Data ja tekoäly -ohjelma, Sitra
Suomen maajohtaja, Creative Commons

2.4.2025

Näin tekoäly syö aineistoa

1

Perustamallin koulutus

Hyödynnetään yleisiä sisältöarkistoja ja -kokoelmia. Myös omia aineistoja (Meta, X, Google).

2

Järjestelmän jatkokoulutus

Suuri osa on käsityötä ja ihmistyötä, jota tehdään halpatyömaissa. Mahdollisesti omien aineistojen hyödyntäminen (Microsoft, Google)

3

Verkon sisältöjen haku käytön aikana

Tilanteen mukaan voidaan tehdä hakuja ja huomioida tulokset toiminnassa.

4

Käyttäjän aineistojen hyödyntäminen

Kaikki käyttäjän syöttämä aineisto (tekstit, kuvat, dokumentit, videot, ääni) voidaan käyttää toiminnan optimointiin.

Tekijänoikeuden sääntelemää toimintaa

- Julkinen esittäminen
- Levittäminen tai jakelu julkisesti
 - Kopiointi yrityksen jaetulle alustalle
- Datan louhinta
 - EU:n DSM: tutkimuskäyttö sallittu aina, kaupallinen käyttö sallittu, mutta voidaan estää (robots.txt)

Creative Commons -lisenssin vaikutus

- Lupa julkisesti esittää sekä julkisesti levittää ja jaella
- Mahdollisesti myös lupa julkisesti levittää muunnelmia
- Edellytyksenä tyypillisesti tekijöiden maininta (BY-ehto)

Tekijänoikeudesta vapaata toimintaa

- Esittäminen yksityisesti
- Yksityinen kopiointi
 - Kopiointi omalle työkoneelle
- Tietosisällön opiskelu ja tutkailu
- Asian esittäminen omin sanoin

robots.txt

- robots.txt on Euroopassa laillisesti tehokas keino kieltää tiedonlouhinta kaupalliseen käyttöön.
- Tekoälyn koulutus lienee tiedonlouhinta, vaikka se muodostui kiinnostavaksi vasta DSM-direktiivin jälkeen

📄 ← ↻ 🏠 🔒 <https://yle.fi/robots.txt>

```
User-agent: *
Disallow: /ylefiroot-errors
Disallow: /cgi-bin
Disallow: /klaffi
Disallow: /java
Disallow: /player
Disallow: /errorsivut/
Disallow: /global/
Disallow: /haku/
Disallow: /eurovision/
Disallow: /embed/
Allow: /global/api/load.php
Allow: /cgi-bin/tekstitv/ttv.cgi/
```

```
# karttasovellus APla 9.7.2015
Disallow: /karttasovellus/
```

```
#ai crawling
```

```
User-agent: AI2Bot
Disallow: /
```

```
User-agent: Ai2Bot-Dolma
Disallow: /
```

```
User-agent: Amazonbot
Disallow: /
```

“We recognize that there is a *perceived* tension between openness and creator choice. Namely, if we give creators choice over how to manage their works in the face of generative AI, we may run the risk of shrinking the commons. To potentially overcome, or at least better understand the effect of generative AI on the commons, we believe that finding a way for creators to indicate “no, unless...” would be positive for the commons.”

Anna Tumadóttir, [Questions for Consideration on AI & the Commons](#)

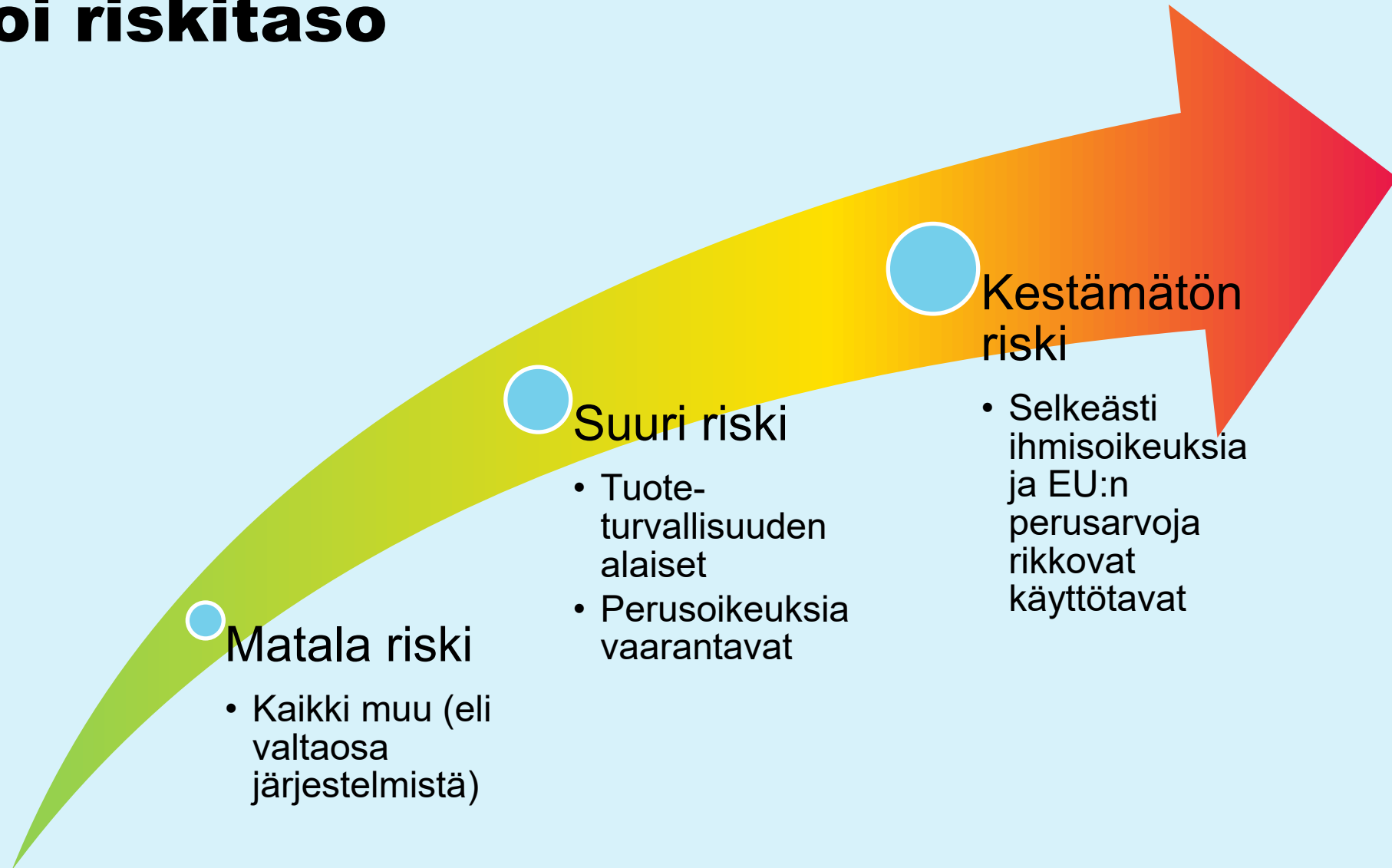
AI + ©: Selkeää

- Tekoälyn tuotoksilla ei ole tekijänoikeuden suojaa
- Jos käyttäjä syöttää aineistoa tekoälypalveluun, on hän vastuussa seurauksista.

AI + ©: Epäselvää

- Onko perustamallin kouluttaminen ”fair use”
 - Onko tulkinta erilaista aineistolajeittain? (tietoteksti, proosa, musiikki, visuaalinen taide, tietokannat...)

Arvioi riskitaso



Matala riski

- Kaikki muu (eli valtaosa järjestelmistä)

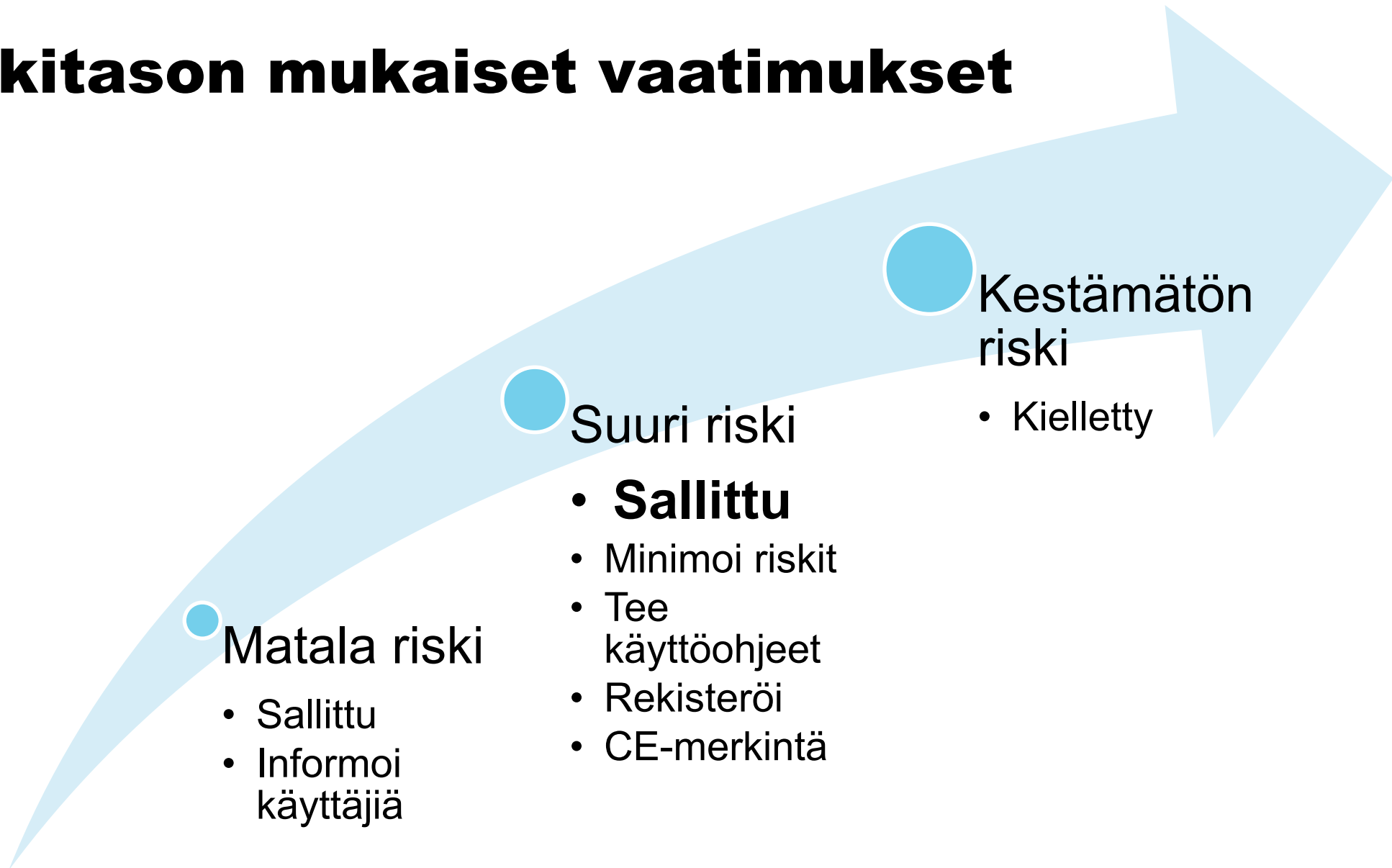
Suuri riski

- Tuote-turvallisuuden alaiset
- Perusoikeuksia vaarantavat

Kestämätön riski

- Selkeästi ihmisoikeuksia ja EU:n perusarvoja rikkovat käytötavat

Riskitason mukaiset vaatimukset



Yleisimmät väärinymmärrykset tekoälyasetuksesta - korjattuna

1. Säätää vain harvoja

Raskaasti säännellyt suuririskiset järjestelmät ovat hyvin pieni osa kaikista tekoälyn käyttötavoista.

2. Suuririskisetkin ovat sallittuja

Suuririskiset järjestelmät ovat **silti sallittuja**, ne vain pitää tehdä erityisen huolellisesti.

3. Opetus ei ole vain suuririskistä

Suuririskiset käyttötapaukset (liite III) **eivät automaattisesti** ole suuririskisiä.



Datatalouden ABC

Maksuttomia verkkokursseja EU:n digi- ja datasääntelystä.

www.sitra.fi/hankkeet/datatalouden-abc/

