Tekoälyn aiheuttama murros: vaikutukset turvallisuuteen

VALTAKUNNALLINEN TURVALLISUUSTAPAHTUMA

SEINÄJOKI 25.5.2023

JAANA HALLAMAA

Tekoäly ja turvallisuus

Jaana Hallamaa

- Sosiaalietiikan professori, Helsingin yliopisto
- Eettinen tekoäly julkisen hallinnon ohjauksessa (ETAIROS) Strategisen tutkimuksen neuvosto

Jäsennys

- Mitä tekoäly on
- Kehitysnäkymiä ja uhkakuvia
- Tekoäly turvallisuuden tukena - ja sen heikentäjä
- Mikä neuvoksi?

Mitä tekoäly on

Ihmisen älykkäiden toimintojen koneellista jäljittelyä (Alan Turing)

Perustuu algoritmien varassa toteutettuun laskentaan.

Toistaiseksi ei yhteisesti hyväksyttyä määritelmää

Viime vuosien kehitys

Datan louhinta

Neuroverkkorakenne

Kielimallit (LLM, large language models)

Omasta datasta oppiminen luo itseään kehittäviä järjestelmiä.

Tekoäly – uusin vaihe symbolien käytön historiassa

Symbolien käyttö uhmaa ajan ja paikan rajoja.

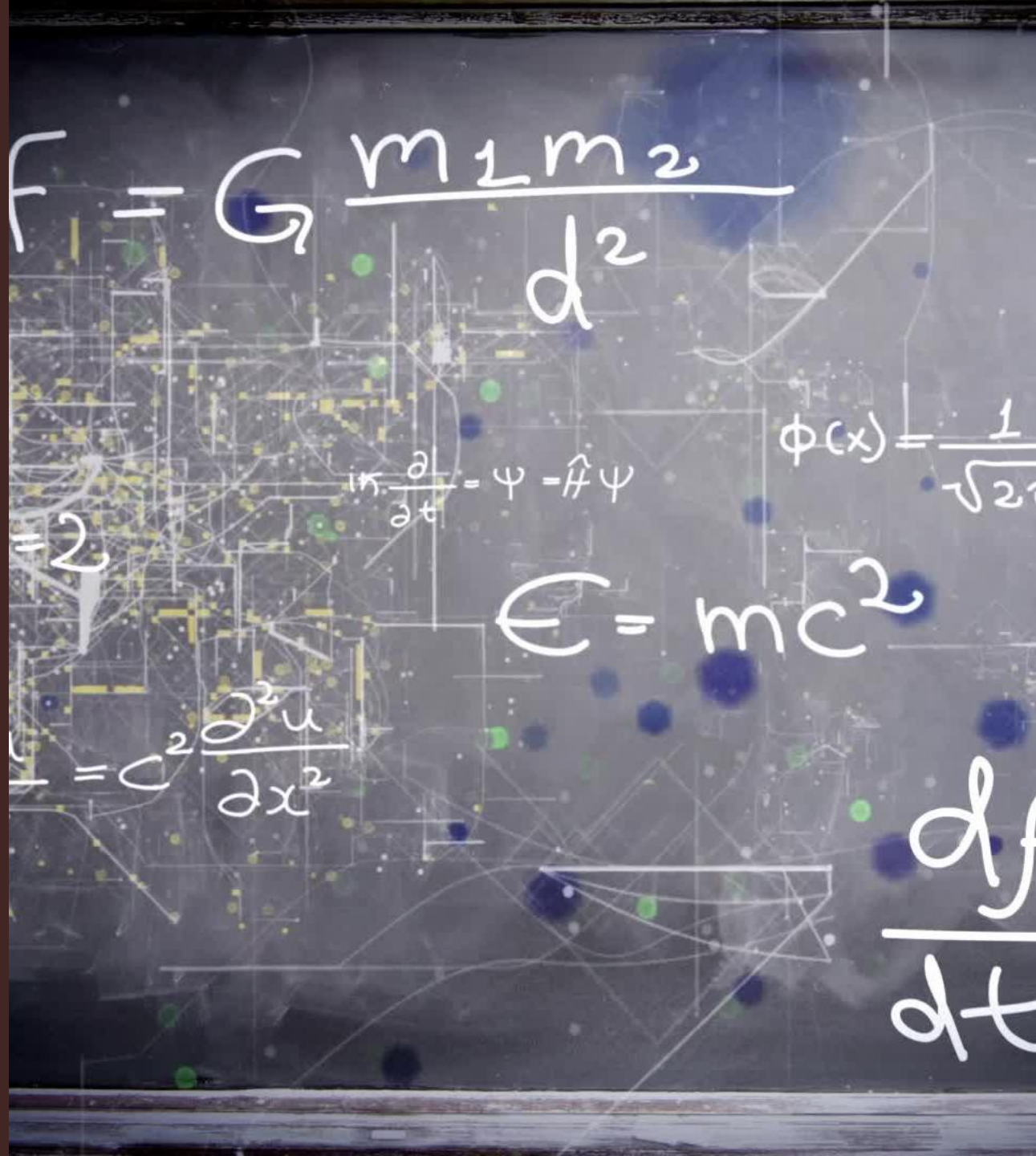
Dataa voidaan kerätä automaattisesti.

Mistä tahansa voidaan saada ja tehdä dataa.

Dataa voidaan jalostaa informaatioksi ja tiedoksi.

Dataa keräävä ja jalostava algoritmi + tekninen laite = omaa toimintaansa ohjaava järjestelmä

Tekoälyn avulla toimintaa voidaan paitsi toteuttaa ja kontrolloida myös ennakoida, ohjata ja kehittää.



Mitä tekoäly muuttaa?

Tehostaa, nopeuttaa ja helpottaa lähes mitä tahansa toimintaa,

Hämärtää rajaa laitteiden, ohjelmistojen ja käyttäjien välillä,

Muuttaa käyttäjänsä ja hyödyntäjänsä raaka-aineekseen,

Vaikuttaa käyttäjiensä ja hyödyntäjiensä ajatteluun, toimintaan ja elämään ja muuttaa niitä.

Edellyttää uudenlaisia taitoja.

Tekee aiemmin tavanomaisista taidoista tarpeettomia.

Lupauksia!

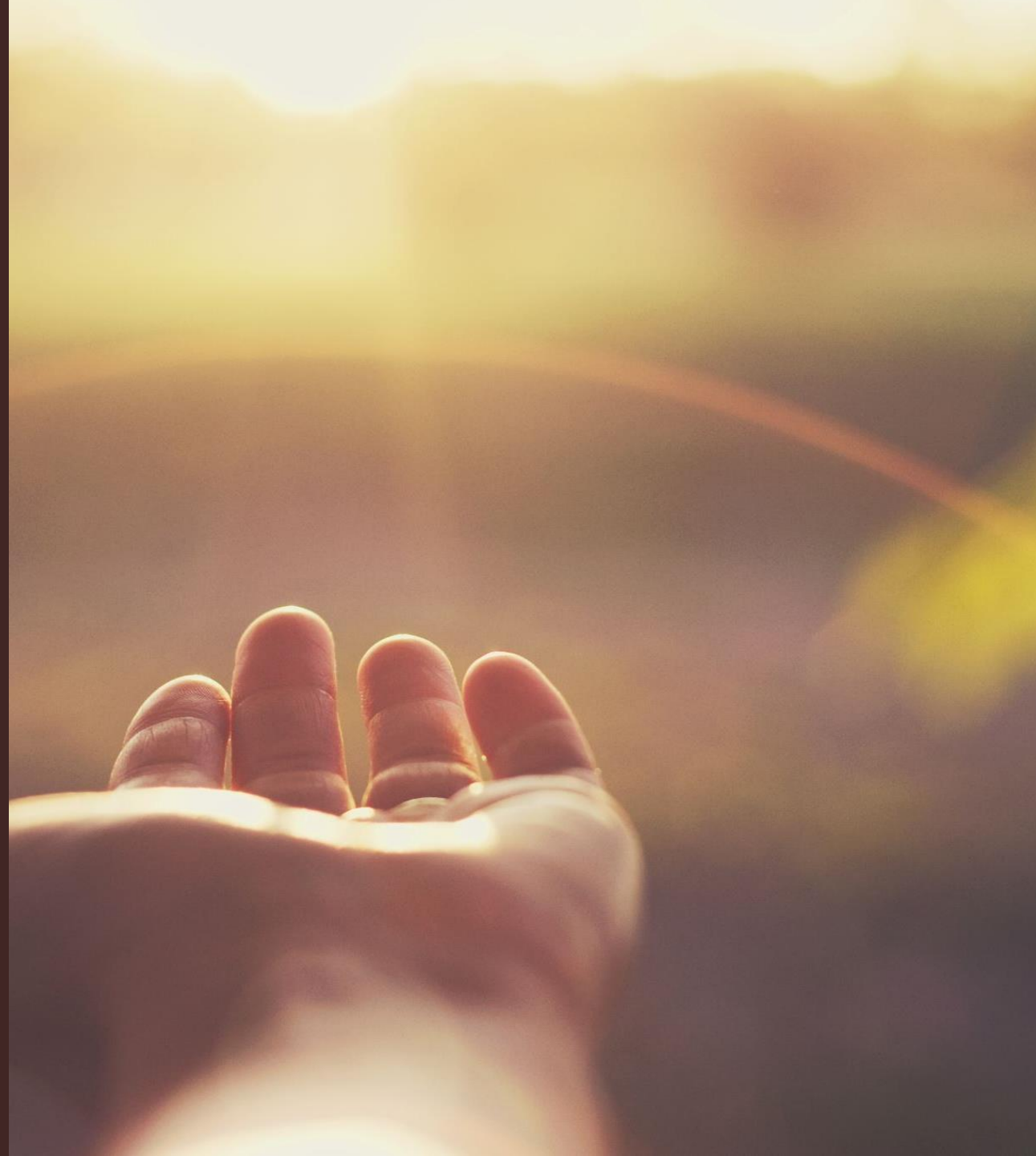
Elinkeinoelämä, viranomaistoiminta ja kansalaisten arki paranevat.

Toiminnoista tulee varmoja, nopeita, täsmällisiä ja turvallisia.

Palvelut voidaan tuoda kaikkien saataville.

Osallisuutta voidaan vahvistaa.

Voidaan luoda uusia osallistumiskeinoja ja -tapoja.

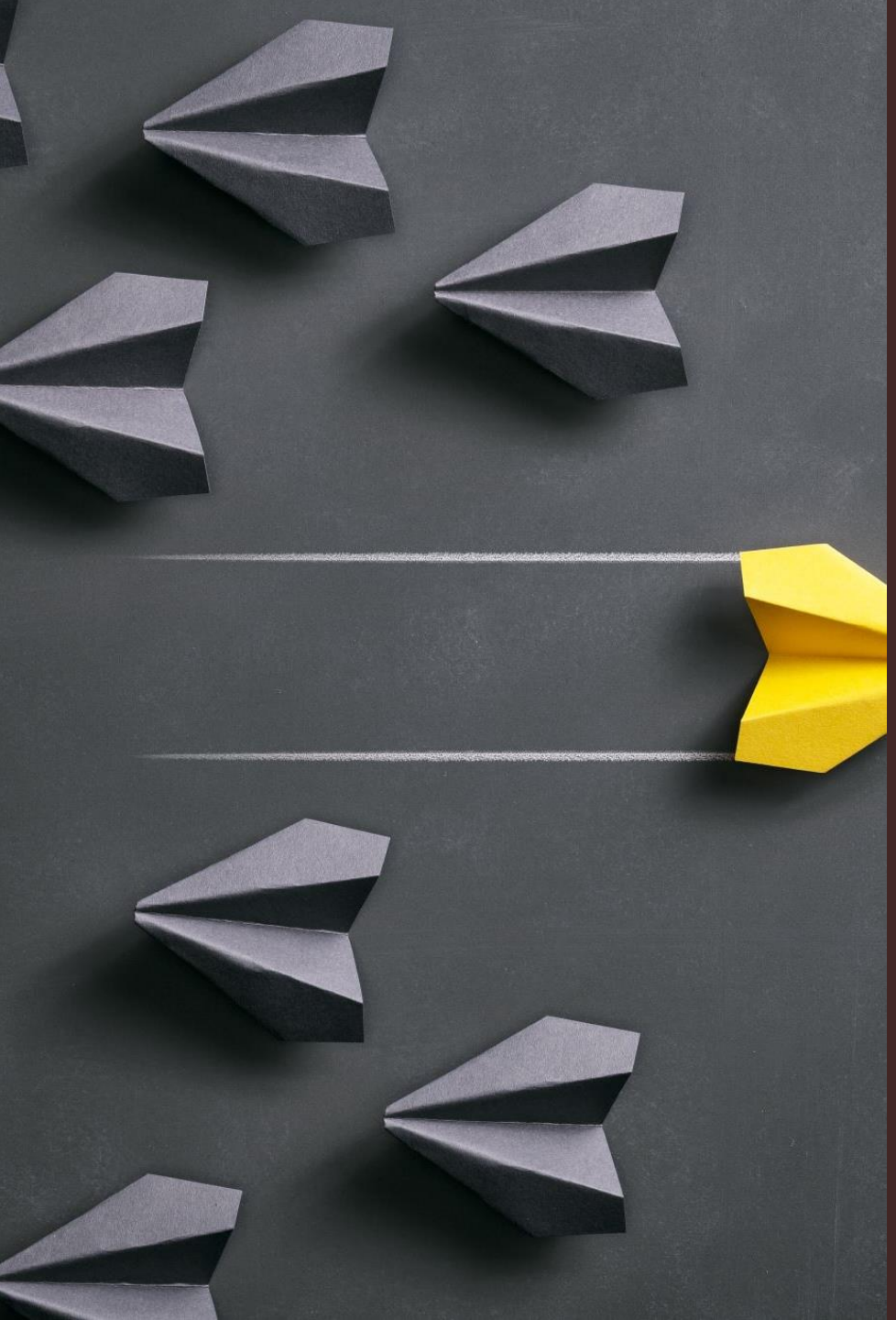


Kehityksen varjopuolia

Epäonnistuminen on toiminnan varjo. (Erik Hollnagel)

Nopeutta, tehokkuutta ja automaattisuutta voidaan käyttää hyväksi missä tahansa toiminnassa.

Toiminnan tehostuminen, valtavien datamassojen kerääminen, säilöminen ja uudet käsittelytavat => ennen näkemättömät väärinkäytön mahdollisuudet



Kokemuksia

Prosessien nopeutuminen ei aina tehosta toimintaa.

Prosessit muuttuvat läpinäkymättömiksi.

Palveluiden automatisoiminen voi tehdä asioiden hoitamisesta lähes mahdotonta.

Sopivan laitteen puuttuminen sulkee palveluiden ja yhteyksien ulkopuolelle.

Jokaisen on jatkuvasti opeteltava uusia taitoja.

On luotu itsepalveluyhteiskunta.

Tekoäly turvallisuustekijänä

- Turvallisuuden aspektit moninaistuvat.
- Mahdollisuus lisätä ja heikentää turvallisuutta vahvistuvat yhtaikaisesti.
  - Ei ole olemassa hyötyjä ilman haittoja.
- Turvallisuus on muuttunut systeemiseksi suureeksi.
  - Asiat kietoutuvat pirullisella tavalla toisiinsa häijyiksi ongelmiksi (*wicked problem).*
- Tuttujen ja tunnettujen turvallisuus- ja vaaratekijöiden pohtiminen ei riitä.
  - Mitä ovat meille tuntemattomat tuntemattomat? (Joharin ikkuna, Joseph Luft ja Harrington Ingham)

Pelkoja ja aavistuksia

- Tekoälyn maailmanvalta?
  - *Vaikuttajien varoitukset tekoälyn vallankaappauksesta*
- Tekoäly voimistaa (myös) ihmisen huonoimpia puolia
  - *Eriarvoisuus voimistuu sekä paikallisesti että globaalisti.*
  - *Voitontavoittelu ohjaa kehitystä, ei yleinen etu.*
  - *Demokratiat kuihtuvat.*
  - *Vastakkainasettelut voimistuvat.*
  - *Koneet korvaavat ihmiskontaktit.*
  - *Tekoäly tekee ihmisistä tarpeettomia.*

Moraaliturvallisuus

Tekoäly on ihmisen luoma apuväline, jonka kehittämisestä ja käytöstä me ihmiset olemme vastuussa.

Eettisesti kestävä tekoäly on keskeinen yhteiskunnan turvallisuustekijä.

Kysy, epäile, tiukkaa!

Kuinka tavoitteet valitaan?

Onko päämääränä ratkaista aito ongelma vai ottaa käyttöön järjestelmä, pysyä mukana kilpailussa tai toteuttaa muotiajatusta?

Kuinka järjestelmä palvelee perustehtävää?

Tukeeko sovellus perustehtävää vai valjastaako se organisaation palvelukseensa?

Kuka hyötyy ja kuka maksaa kustannukset?

Mitä ihmiselle tapahtuu?

Välittömät ja pitkän aikavälin vaikutukset

Mitä jää näkemättä?

Mitä emme halua tuntea ja tunnustaa?