Väärtused ja eetilised valikud uute tehnoloogiate arendamisel

Margit Sutrop

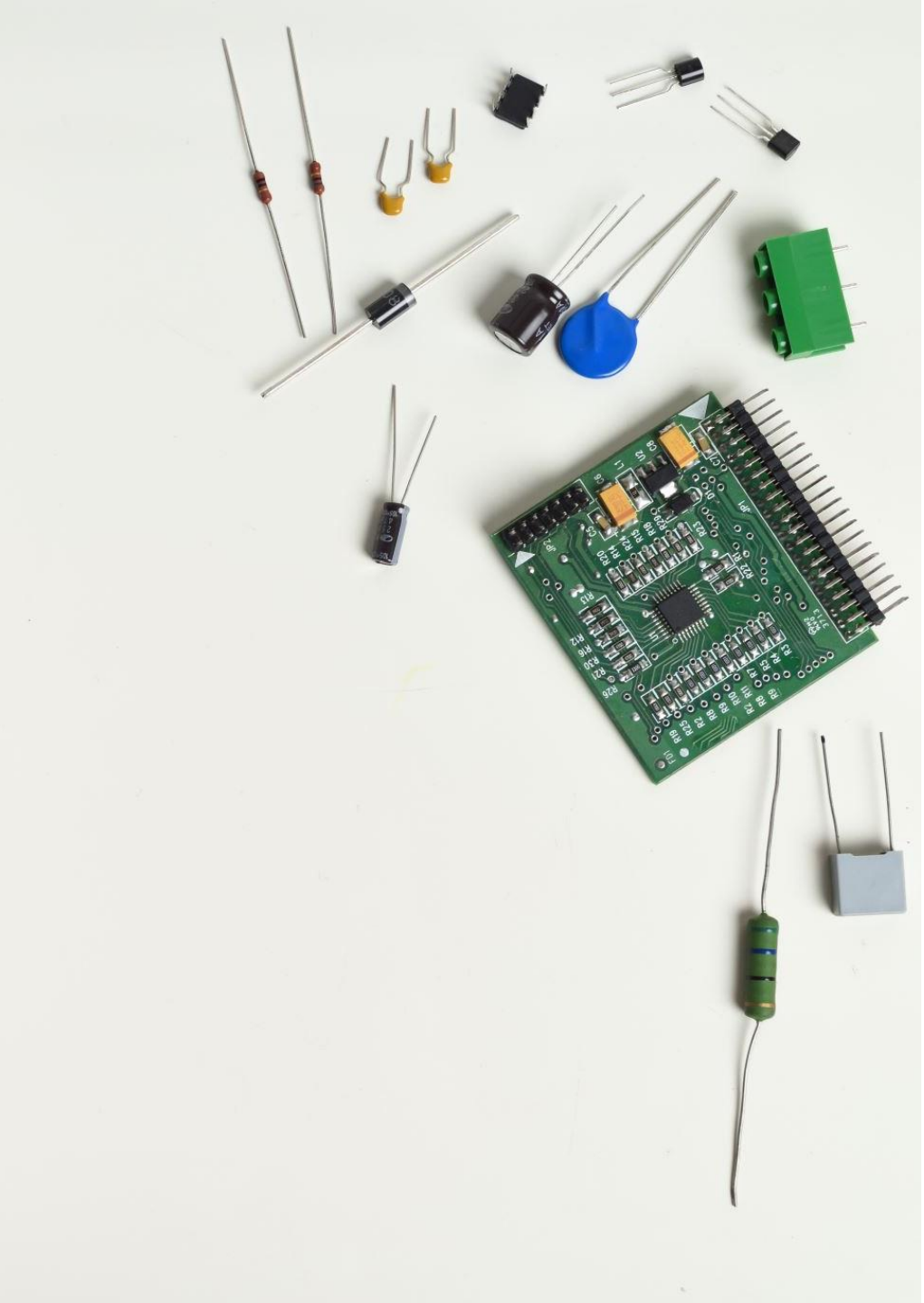
Riigikogu liige, Tartu Ülikooli praktilise filosoofia professor

Teaduspoliitika konverents Riigikogus

11. oktoobril 2023

Kava

- Tehnoloogia mitu palet
- Eetika ja tehnolooogia
- Väärtused ja väärtuste kaalumine
- Kuidas anda tehisarule inimestele olulised väärtused?



Tehnoloogia mitu palet

- Tehnoloogia kui vabastaja
- Tehnoloogia kui oht
- Tehnoloogia kui kahe teraga mõõk

Ian Barbour “Ethics in the Age of Technology” (1993)

Eetika ja tehnoloogia

Väärtuste kaitse läbi eetiliste raamistike

- Teaduseetika (RE, *research ethics*)
- Vastutustundlik teadustöö ja innovatsioon (RRI, *responsible research and innovation*)

Väärtuste integreerimine tehnoloogiasse

- Turvalisus läbi disaini (*security by design*)
- Privaatsus läbi disaini (*privacy by design*)
- Eetika läbi disaini (*ethics by design*)

Turvalisus läbi disaini

Turvalisus läbi disaini (*security by design*) on lähenemine tarkvara arendamisele, mille eesmärgiks on teha süsteemid nii rünnakukindlaks kui võimalik.

Microsoft võttis kasutusele vastavad metoodikad, standardid.

Põhielemendid arendusprotsessi

- pidev testimine,
- mitmetasemeline autentimine,
- sessiooni protokollid

Turva parimad praktikad.

Privaatsus läbi disaini

- **Privaatsus läbi disaini** (*privacy by design)* on lähenemine süsteemitehnoloogiale, mille algselt töötas välja Ann Cavoukian, Ontario (Kanada) teabe- ja privaatsusvolinik.
- Põhimõte vormistati Kanada ja Hollandi privaatsust suurendavate tehnoloogiate ühisaruandes 1995.
- Kavandatud privaatsuse raamistik avaldati 2009 ja võeti vastu 2010 rahvusvahelise privaatsuse volinike ja andmekaitse inspektsioonide poolt.
- Põhimõte viidi sisse EL andmekaitse üldmäärusesse (*General Data Protection Directive* 2018)

Usaldusväärne tehisintellekt

Euroopa Komisjoni kõrgetasemeline tehisintellekti ekspertrühm: Eetikasuunised usaldusväärse tehisintellekti arendamiseks (2019 aprill).

- **Usaldusväärse tehisintellekti 3 aspekti:**

1. see peaks olema seaduslik ja vastama kõigile kohaldatavatele õigusnormidele,
2. see peaks olema eetiline ja tagatud peaks olema eetikapõhimõtete ja -väärtuste järgimine,
3. see peaks olema nii tehniliselt kui ka sotsiaalselt töökindel.

Eetika läbi disaini

- **Eetika läbi disaini (ethics by design)** põhimõte tähendab oluliste väärtuste silmas pidamist tehisaru kavandades, arendades ja kasutades.
- **Eetika** kaitseb individuaalseid õigusi nagu vabadus ja privaatsus, võrdsus, õiglus, hoiab ära inimeste kahjustamisest ja edendab heaolu, kindlustades jätkusuutlikku keskkonda.
- **Eetilised printsiibid**: 1. Austus autonoomia, inimväärikuse ja vabaduse vastu; 2. Privaatsus ja andmekaitse; 3. Õiglus; 4. Individuaalne, sotsiaalne ja keskkonna heaolu; 5. Läbipaistvus; 6. Vastutus ja järelvalve.

”Ethics by Design and Ethics of Use Approaches for Aritificial Intelligence” (European Commission, 2021)

Väärtused

Väärtused on abstraktsed mõisted, mis vajavad tõlgendamist

Väärtused võivad põrkuda

Väärtusi saab seada hierarhiasse, võttes arvesse konteksti. Väärtuste kaalumine

Valik tehnoloogiaid ja rakendusi

- Geeniandmebaasid
- E-tervis
- Biomeetria
- Tehisintellekt

Geneetiline andmebaas kui avalik hüve

- Geneetilised andmebaasid esitasid väljakutse teaduseetika kesksele printsiibile - informeeritud nõusolekule, mis oli alates Teisest maailmasõjast teaduseetika keskne printsiip.
- Avatud nõusolek, lai nõusolek uurimistöö tegemiseks.
- Kommunitaristlik pööre bioeetikas: kollektiivsed väärtused (solidaarsus, vastastikusus, tervis versus inidividuaalsed väärtused nagu autonoomia, vabadus ja privaatsus).
- Tegelikult ei pea olema vastandlikud. Aristoteles näidanud, kuidas autonoomia ja ühishüve on ühitatavad (Sutrop 2011).

E-tervise andmebaas

- Inimeste terviseandmed edastatakse automaatselt, inimestele jääb võimalus teatud liiki andmed kinni panna (opt-out).
- Vastutus tervise eest jääb inimesele.
- Andmete kasutamine individuaalseteks (parem diagnostika, ravi) ja sotsiaalsteteks eesmärkideks (tervishoiu planeerimine, teadusuuringud, kvaliteedikontroll)
- Teaduseetika kesksest printsiibist – informeeritud nõusolekust taganemine, **autonoomia vähem tähtis kui tervis**. Privaatsus võib olla isegi paremini kaitstus. Sõltub sellest, kui hästi terviseandmed on kaitstud väärkasutuse vastu. (Sutrop 2011)

Biomeetriliste andmete kogumine

Biomeetria kaks kasutust: turvalisuse kaitseks

1. Vajadus eristada üht inimest teisest läbi verifitseerimise või identifitseerimise (sõrmejäljed, silma iiris)
2. Ennustada kellegi käitumist või kavatsusi (kõnnak, kehatemperatuur, lõhn, ECG)

Eetiliselt probleemne:

- Kuna teise põlvkonna biomeetrilisi tunnuseid saab tabada kaugelt, inimene ei pruugi olla teadlik sellest, et biomeetrilisi tunnuseid kogutakse ja analüüsitakse. Oht autonoomiale
- Profileerimine eetiliselt probleemne. Oht nimväärikusele vastu
- Stigmatiseerimise ja diskrimineerimise oht. Oht võrdsusele
- Oht, et andmeid kasutakse teiseks eesmärgiks, kui neid kogutakse (*function creep*). Oht privaatsusele.

(Sutrop ja Laas-Mikko 2012)

Kuidas joondada tehisintellekti inimeste väärtustega?

„**Väärtuste joondamist**” (*value alignment*) mõistetakse kui intellektiga agendi omadust, mis lubab tal järgida eesmärke ja tegevusi, mis teenivad inimeste heaolu ja on kooskõlas tema eesmärkide ja eelistustega.

Tehisintellekti joondamine väärtustega seisab vastamisi kahe väljakutsega:

- **Tehniline väljakutse** seisneb selles, kuidas õpetada tehisintellektile väärtusi?
- **Normatiivne väljakutse** seisneb küsimuses: milliseid väärtusi või kelle väärtusi tuleks tehisintellektile anda? (Sutrop 2020)

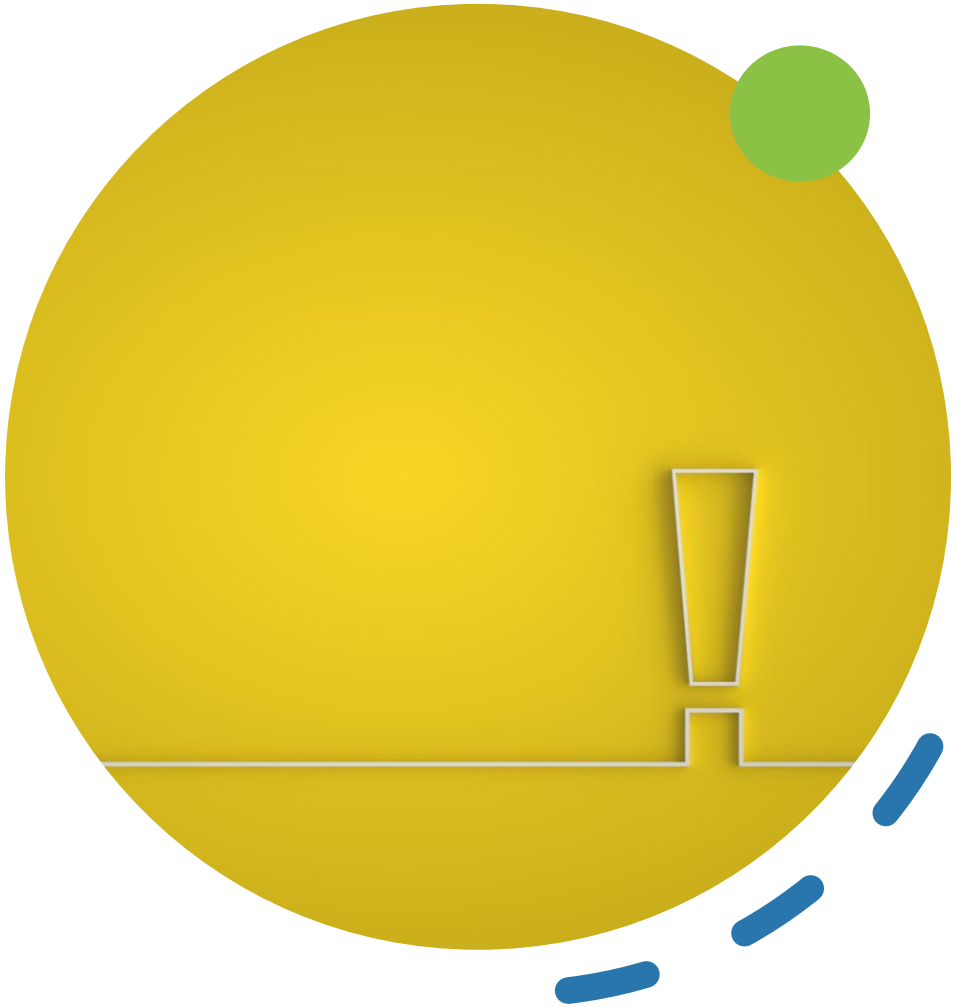
Väärtuste joondamise 3 printsiipi

Stuart Russell: et tagada kontroll tehisintellekti (TI) üle, tuleb disainida heatahtlik TI, mis järgiks inimeste väärtusi, oleks alandlik, altruistlik ja pühendunud sellele, et teenida meie eesmärke, mitte tema omi.

- **Russell on defineeerinud 3 printsiipi TI arendajatele:**

1. Masina ainus eesmärk on maksimeerida inimeste eelistusi.
2. Masin on algselt teadmatuses, mis need eelistused on.
3. Ainus infoallikas inimlike eelistuste kohta on inimeste käitumine.

Stuart Russell, “Human Compatible: AI and the Problem of Control” (2019)



Väärtuste kaalumine

- Tehnoloogia arendamine ja kasutamine eeldab inimestele oluliste väärtuste arvestamist ja nende varakult integreerimist tehnoloogiasse.
- Väärtuseid võib erinevalt tõlgendada ja nende hierarhiasse seadmine sõltub sellest, kas peame tähtsamaks individuaalseid või kollektiivseid väärtusi, inimõigusi või ühishüve.
- Väärtuste kaalumine sõltub kontekstist.
- Tehnoloogia joondamine väärtuste järgi eeldab ühiskondlikule kokkuleppele jõudmist olulistes väärtustes ja eelistustes.

Kirjandus

- Margit Sutrop, “Changing Ethical Frameworks: from individual rights to the common good?” *Cambridge Quarterly of Healthcare Eth*ics, 2011, 20, 4, 533–545. DOI: 10.1017/S0963180111000272.
- Margit Sutrop, Katrin Laas-Mikko “From identity verification to behaviour prediction. Ethical implications of second-generation biometrics. *Review of Policy Research*, 2012, 29,1,21–36. DOI:10.1111/j.1541-1338.2011.00536.x
- Margit Sutrop, “Should we trust artificial intelligence?”, *Trames. A Journal of Humanities and Social Science*, 2019, 23, 4, 499-522.

  DOI: 10.3176/tr.2019.4.07
- Margit Sutrop, “Challenges of Alignikng Artificial Intelligence with Human Values”, *Acta Baltica Historiae et Philosophiae Scientarum*, 2020, 8, 2, 54-72. DOI: 10.11590/abhps.2020.2.04.