

# Automatiseren van morele oordeelsvorming: Be careful what you wish for

Co-referaat NVBe 11 juni 2021

Marianne Boenink  
Ethics of Healthcare, IQ Healthcare  
[marianne.boenink@radboudumc.nl](mailto:marianne.boenink@radboudumc.nl)



---

# Een confronterend pre-advies...

- Ethiek van AI gaat meestal over het werk van anderen
- Dit pre-advies
- Dwingt tot reflectie op eigen expertise
  - Zowel inhoudelijk
  - Als normatief
- Nodigt uit tot reflectie op de effecten van automatisering van AI voor het vakgebied

---

# Mijn respons

- Herkenbare problemen en invoelbare roep om verbetering
- Interessant: kan AI hierbij ondersteunen?
- Focus in pre-advies vaak op
  - belang van autonomie & verantwoordelijkheid, privacy
  - praktische belemmeringen
- Maar: vergt
  - inzicht in hoe mensen tot morele oordelen komen
  - en hoe zij dat zouden moeten doen
- Zelfs bij de 'simpele' (zwakke) vormen van AI al problematisch

---

# Morele oordeelsvorming

- Pre-advies:
- Intuïtie en redeneren (cf. Haidt)
- Drie manieren om AI te trainen:
  
- Bottom-up:
  - Voeden met data waarin patronen ontdekt gaan worden > AI leert herkennen wat goed/slecht is
  - NB: AI moet oordelen ook leren onderbouwen
  
- Top-down:
  - AI wordt gevoed met regels/ethische theorie
  - Maar welke?
  
- Hybride

---

# Onderliggend idee

- Morele oordelen gaan over concrete situaties, waar normatieve regels op worden toegepast
- Maar: Toepassing van regel op situatie is niet vanzelfsprekend
  - Welke regel(s) is/zijn relevant?
  - Hoe in deze situatie te interpreteren?
- Uitdagingen hierbij zijn o.a.
  - Rijkdom en dynamiek van casuïstiek
  - Meervoud aan en specificiteit van regels
  - En de combinatie van beide (ethische abductie?)

---

# Wat betekent dat voor AI?

- Casus 1: zoeken naar data: is er sprake van ondraaglijk lijden?
  - Maar: wie bepaalt wat relevant is, en hoe? Dynamische opvattingen?
  - Wat kan wel/niet 'gedataficeerd' worden?
- Casus 2: rechtvaardige verdeling van goederen
  - Duidelijke regels/instructies om situaties te classificeren ('hoogste nood')
  - Voorbeelden van goede oordeelsvorming (gouden standaard)
- Casus 3: socratische assistent (klinisch-ethicus?)
  - Voeden met normatieve opvattingen: theorieën (welke) of casuïstische oordelen?
  - Wat geldt als een verbetering van je eigen oordeel?
    - Meer aspecten, minder bias, consistentere, meer logica?

---

# Wat betekent 'beter' hier?

- Ook als manier van redeneren anders is, kan resultaat duidelijk beter zijn
- Maar op grond van welk criterium/criteria?
  - bij bv medische AI kun je trials doen om uitkomsten van 'reguliere' en 'AI-based' toepassingen te vergelijken
  - maar hoe zit dat bij morele oordeelsvorming?
  - wat zouden de 'uitkomst parameters' zijn?

---

# Be careful what you wish for!

- Automatisering van morele oordeelsvorming zou wel eens kunnen leiden tot platgeslagen
  - vormen van morele oordeelsvorming
  - visie op ethiek





---

# Twee terugkerende vragen

- Hoe doen **mensen** dat eigenlijk, moreel oordelen?
- En hoe onderscheiden we **betere** van **slechtere** oordelen?



---

# Dank, Katleen!

- Meer begrip voor defensieve reactie artsen
- (kan ik daardoor betere ethiek van AI bedrijven?)