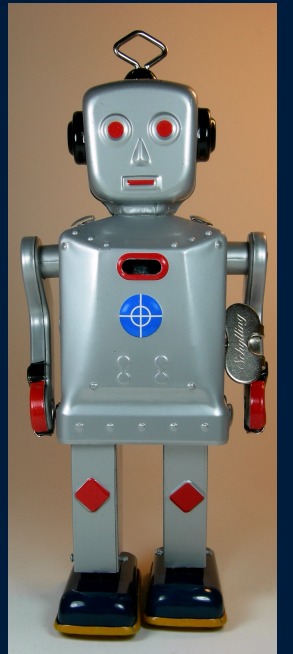
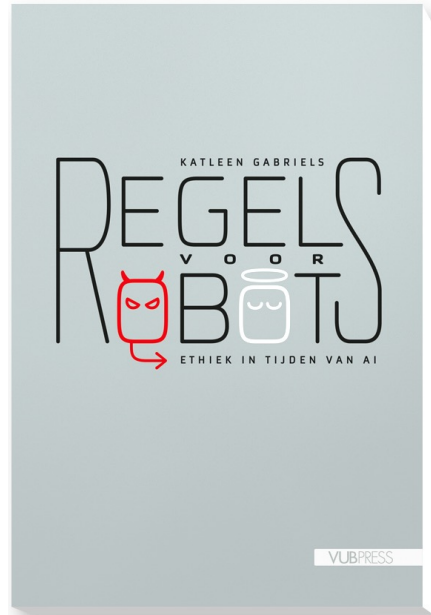


Hebben robots ook een moreel kompas en compassie?

Katleen Gabriels, assistant professor
k.gabriels@maastrichtuniversity.nl

Brainfeed – 15 juni 2022





Preadvies

uitgebracht t.b.v. het Jaarsymposium van de
Nederlandse Vereniging voor Bio-ethiek 11 juni 2021



<https://nvbioethiek.wordpress.com/2021/08/05/preadvies-2021-siri-wat-adviseer-jij-nu-beschikbaar/>

Digital Society



Google schorst medewerker die diepe gesprekken met AI-chatbot lekte: "Heeft LaMDA gevoelens als een mens?"

<https://www.vrt.be/vrtnws/nl/2022/06/13/google-schorst-medewerker-die-diepe-gesprekken-met-ai-chatbot-le/>

GOOGLE / TECH / ARTIFICIAL INTELLIGENCE

Google suspends engineer who claims its AI is sentient

It claims Blake Lemoine breached its confidentiality policies

By Jon Porter | @JonPorty | Jun 13, 2022, 6:15am EDT

<https://www.theverge.com/2022/6/13/23165535/google-suspends-ai-artificial-intelligence-engineer-sentient>

Outline

1. Inleiding
2. Voordelen KI – altijd beter?
3. *Artificial Moral Agents*: Moraliteit en KI
4. Slotbeschouwing

1. Inleiding

Wat is intelligentie?

Max Tegmark: “het vermogen om complexe doelstellingen te verwezenlijken”

- Sociale
- Creatieve
- Analytische
- Wiskundige
- Emotionele
- Praktische
- ...

Wat is kunstmatige intelligentie?

A Proposal for the
DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

Rodney Brooks:

- AI “started as a field whose goal it was to replicate human level intelligence in a machine”

Brooks, R. A. (1999). *Cambrian Intelligence. The Early History of the New AI*. Cambridge, MA: MIT Press, p. 80.

- Mensen hebben *zicht* -> KI: beeld- en patroonherkenning
- Mensen hebben *spraak* -> Natural Language Processing (NLP)
 - e.g. Siri, Woebot, GPT-3 ('text generation algorithms'), Google Duplex
- Mensen kunnen *emoties herkennen* -> algoritmen voor emotieherkenning
- ! Nog steeds onnauwkeurigheid bij NLP en emotieherkenning !
Taal en emoties werden lange tijd genegeerd als onderdelen van intelligentie

- Robots zijn ‘containers’ en de *belichaming* van KI geworden
- Zelfrijdende wagen: sensoren, camera’s (‘ogen’), microfoons (‘oren’) om te leren van ervaringen
 - Data = input
 - Data worden geanalyseerd om een beslissing of voorspelling te doen (output)
- Menselijke intelligentie overheerst; is veel complexer en veelzijdiger

Nieuwe KI-zomer dankzij

- nieuwe technologische ontwikkelingen en applicaties;
- toegenomen opslagcapaciteit, snelheid en rekenkracht;
- en big dataprojecten die immense datasets aanleggen waar algoritmen mee getraind worden:
 - datagebaseerde KI
 - zelflerende algoritmen

Letter | Published: 25 January 2017

Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva , Brett Kuprel , Roberto A. Novoa , Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun 

Nature **542**, 115–118 (02 February 2017) | [Download Citation](#) 

<https://www.nature.com/articles/nature21056>

AI 'outperforms' doctors diagnosing breast cancer



Fergus Walsh
Medical correspondent
@BBCFergusWalsh

🕒 2 January 2020





Radiologe Isabelle Biltjes is gespecialiseerd in mammografie. — © RR

Antwerpse primeur: AZ Klina zet artificiële intelligentie in tegen borstkanker

Antwerpen - Het AZ Klina in Brasschaat kan voortaan met de hulp van artificiële intelligentie herkennen of borstweefsel er verdacht uitziet én een inschatting maken van het risico op borstkanker binnen de twee jaar. "Maar de computer zal nooit beslissen hoe een patiënt wordt behandeld."

Maxie Eckert

Donderdag 15 oktober 2020 om 05:00



https://www.gva.be/cnt/dmf20201014_96462402

Paradox van Moravec

Wat makkelijk is voor de mens is moeilijk voor de computer
en vice versa

40982 x 9825?

= 402 648 150

https://www.youtube.com/watch?v=_sBBaNYex3E



2. Voordelen KI – altijd beter?

“Was ist der Mensch?”

Immanuel Kant

**“Uit het kromme hout waaruit de mens is
opgebouwd
is nog nooit iets rechts getimmerd”**

Immanuel Kant

“suboptimal information processors”

Giubilini & Savulescu (2018, p. 170)

Ondermaatse informatieverwerkers

- Efficiëntie
- Snelheid
- Verzamelen, analyseren en verwerken van een enorme hoeveelheid data
- Consistentie
- Meer objectief

In hoeverre zou dit leiden tot beter geïnformeerde en meer onderbouwde (morele) oordelen?

Kritisch blijven

ARTIFICIAL INTELLIGENCE

Hundreds of AI tools have been built to catch covid. None of them helped.

Some have been used in hospitals, despite not being properly tested. But the pandemic could help make medical AI better.

<https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/>

Zie: L. Wynants *et al.*, “Prediction models for diagnosis and prognosis of COVID-19: Systematic review and critical appraisal,” *BMJ*, vol. 369, p. m1328, Apr. 2020.



XENOPHOBIC MACHINES

DISCRIMINATION THROUGH UNREGULATED USE OF
ALGORITHMS IN THE DUTCH CHILDCARE BENEFITS
SCANDAL



<https://www.amnesty.nl/actueel/toeslagenaffaire-is-mensenrechtenschending-zegt-amnesty-international>



De toeslagenaffaire bij de belastingdienst ontstond doordat men gegevens over tweede nationaliteit van mensen ging gebruiken in algoritmen die risicogroepen moesten opsporen.

📷 CCO (publiek domein)

<https://www.nemokennislink.nl/publicaties/hoe-kunnen-we-leren-van-discriminerende-algoritmes/>



vrouw actualiteit



[Alle](#)

[Afbeeldingen](#)

[Nieuws](#)

[Video's](#)

[Maps](#)

[Meer](#)

[Instellingen](#)

[Tools](#)

Ongeveer 524.000 resultaten (0,39 seconden)

Bedoelde je: [vrouw seksualiteit](#)



Een privacyherinnering van Google

[MIJ LATER HERINNEREN](#)

[BEKIJKEN](#)



🔍 oma weet

🔍 oma weet raadt

🔍 oma weet raadt gekneusde ribben

🔍 oma weet raadt jeuk bij honden

🔍 oma weet raadt keelpijn

🔍 oma weet raadt spierpijn

🔍 oma weet raadt rimpels

🔍 oma weet raadt afvallen

🔍 oma weet raadt wondroos

REPORT

Researchers hack a self-driving car by putting stickers on street signs



JOHN BELTZ SNYDER

Aug 4th 2017 at 4:17PM



<https://www.autoblog.com/2017/08/04/self-driving-car-sign-hack-stickers/>

3. *Artificial Moral Agents:* Moraliteit en KI

4 types

- Ethische impact actor (AMA-1)
- Impliciet ethische actor (AMA-2)
- Expliciet ethische actor (AMA-3)
- Volledig ethische actor (AMA-4)
- Zwakke – sterke KI

Casuïstiek

- Casus 1: Gebrek aan informatie bij euthanasie
 - Aanleveren, analyseren, verwerken van data
- Casus 2: Coördinatie en distributie van hulp en goederen bij ramp
 - Aanleveren, analyseren, verwerken van data én rangschikking (triage)
- Casus 3: KI-gesprekspartner
 - O.a. vergevorderd begrip van taal (nog niet mogelijk), zou zelf niet moreel oordelen, maar zou adviseren en ervoor zorgen dat mens tot beter moreel oordeel komt
- Casus 4: De robotopvoeder
 - Kan zelf moreel oordelen en oordeel van de mens naast zich neerleggen

Uitdagingen

- Empirisch
- Technisch
- Conceptueel
- Ethisch

Empirisch

- E.g. Hoe komen mensen tot morele oordelen?
- Jonathan Haidt: onze affectieve reacties en intuïties sturen in de eerste plaats ons moreel redeneren
- Intuïties en oordelen kunnen veranderen in een gesprek waarin argumenten worden aangedragen (Haidt, 2013, p. 55)

Hoe trainen?

- Top-down
 - ‘Wat gij niet wilt dat u geschiedt, doe dat ook een ander niet’
- Bottom-up
- Hybride

Technisch

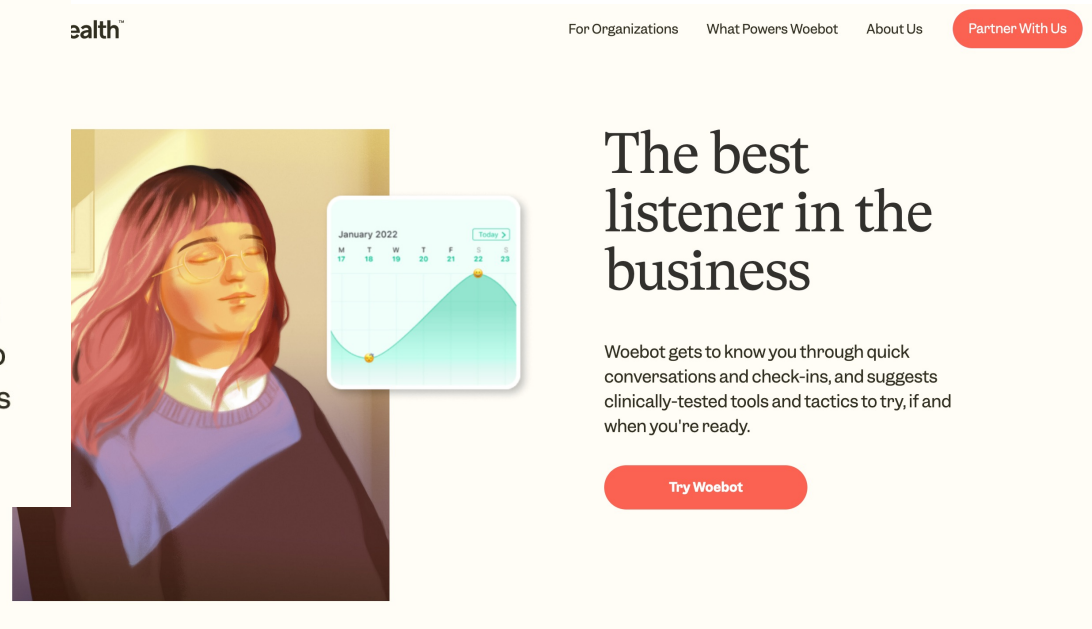
- Hoe machines moraliteit aanleren?
 - Contextafhankelijkheid van moraliteit
 - Natural Language Processing (NLP); argumentatie
 - Geen blauwdruk

Begripsverwarring (conceptueel)

- ‘Agent’
 - Zwak: ‘software or a machine performing a certain task’
 - Zwak: “a person or thing that produces a particular effect or change” – cf. Toeslagenaffaire
 - Sterk: ‘to be able to reason and act’
- ‘Moral agent’
 - Zwak: cf. AMA-1
 - Sterk: vrije wil, bewustzijn, morele verantwoordelijkheid, ...
- Autonomie
- Intelligentie
- Morele oordelen: intuïtie en rede

Small chats for big feelings

Meet Woebot, your personal mental health ally that helps you get back to feeling like yourself. Grounded in clinical research and powered by AI, Woebot is easy to talk to and fits right into your life, whenever you want to chat. There's no such thing as appointments or waiting rooms here.



The screenshot shows the Woebot website homepage. At the top left is the 'woebot health' logo. To the right is a navigation bar with links: 'For Organizations', 'What Powers Woebot', 'About Us', and a red 'Partner With Us' button. The main content area is split into two columns. The left column features an illustration of a woman with long pink hair and glasses, with a small calendar and mood graph overlay. The right column has the heading 'The best listener in the business', a paragraph about Woebot's AI-powered mental health support, and a red 'Try Woebot' button.

woebot health

For Organizations What Powers Woebot About Us Partner With Us

The best listener in the business

Woebot gets to know you through quick conversations and check-ins, and suggests clinically-tested tools and tactics to try, if and when you're ready.

Try Woebot

<https://woebothealth.com/>

Ethisch

- Manipulatie
- Morele luiheid: ons oordeel delegeren aan machine (cf. Toeslagenaffaire)

4. Slotbeschouwing

- Tool
 - Kritisch denken! AI = *buzzword*
- Cruciale rol voor de mens
- Grote, complexe vragen: wat is denken? Wat is bewustzijn? Hoe oordeelt de mens moreel?

Siri

Ik luister...

Wat is de betekenis van het leven

Daar Kant ik geen antwoord op geven.

Ha ha!