Regels voor robots
Ethiek in tijden van AI

Kenniscentrum Data & Maatschappij
9 december 2019

Katleen Gabriels
k.gabriels@maastrichtuniversity.nl

https://www.linkedin.com/in/katleen-gabriels-6415b938/

Maastricht University

Maastricht University

KATLEEN GABRIELS
REGELS VOOR BOTS
ETHIEK IN TIJDEN VAN AI

Elk product dat door mensen gemaakt is, is ‘made with morality’

NEWS

Pure Gym CEO: We're not a sexist company

By Jak Phillips 20 Mar 2015

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.

Photographer: Joe Mehling

Figure 1. Trenchard More, John McCarthy, Marvin Minsky, Oliver Selfridge, and Ray Solomonoff.

1956 - 2006

Maastricht University

KATLEEN GABRIELS
REGELS VOOR BOTS
ETHIEK IN TIJDEN VAN AI


Anchor @anchor
Elon Musk says artificial intelligence is humanity's greatest threat
36 minutes ago
WEIRD
Security robot drowns itself
2 hours ago

Nieuwe AI-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 AI)

Zie o.a. Russell, S. J., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach (Third Edition)*. Essex: Pearson Education Limited, p. 29.

nature
International journal of science

Letter | Published: 25 January 2017

Dermatologist-level classification of skin cancer with deep neural networks

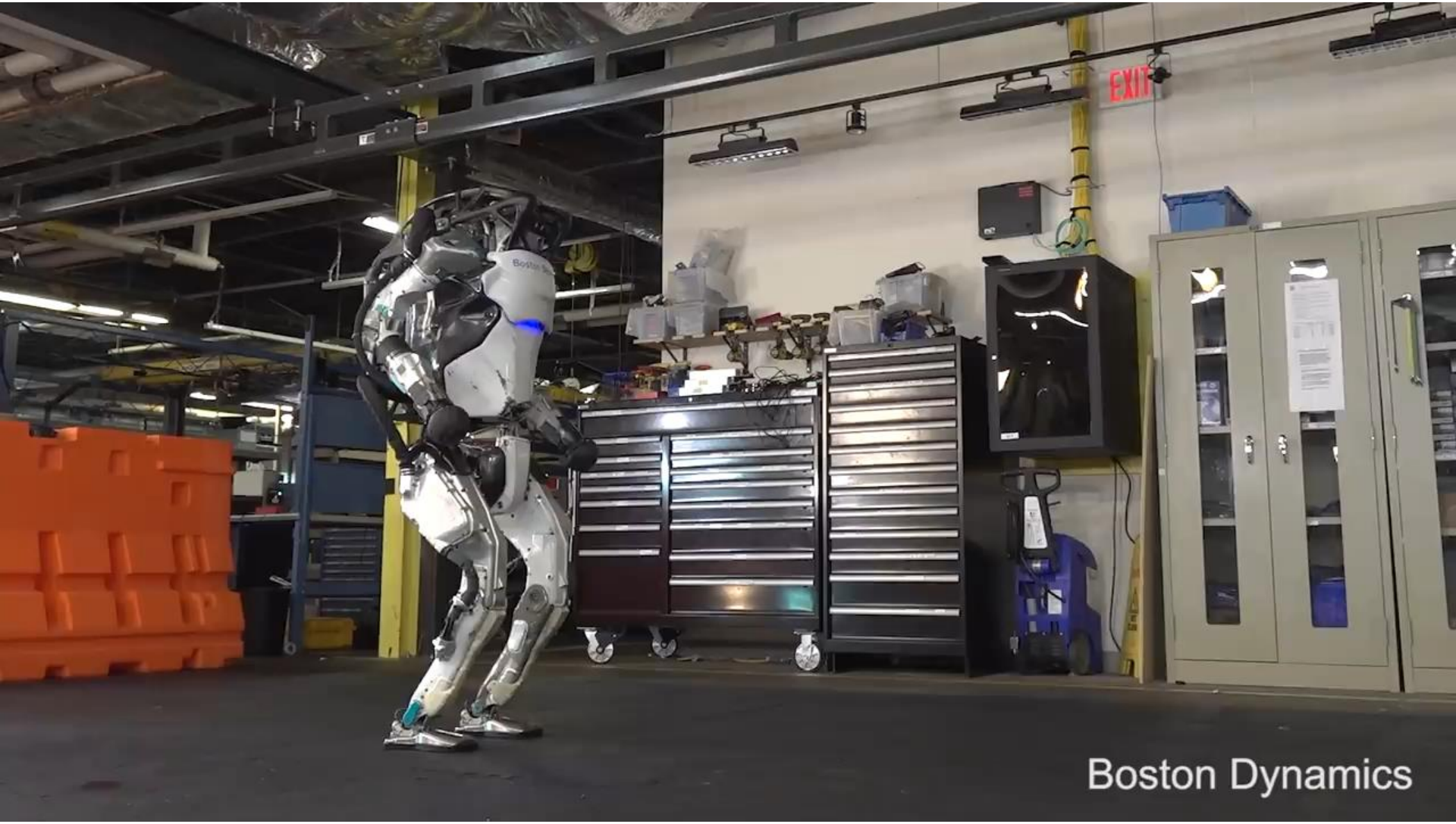
Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun

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

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

Maastricht University

KATLEEN GABRIELS
REGELS VOOR BOTS
ETHIEK IN TIJDEN VAN AI



Maastricht University

KATLEEN GABRIELS
REGELS VOOR BOTS
ETHIEK IN TIJDEN VAN AI

Ethische problemen

1. ‘Verborgen moraliteit’

2. Moraliteit ≠ beslissingsboom

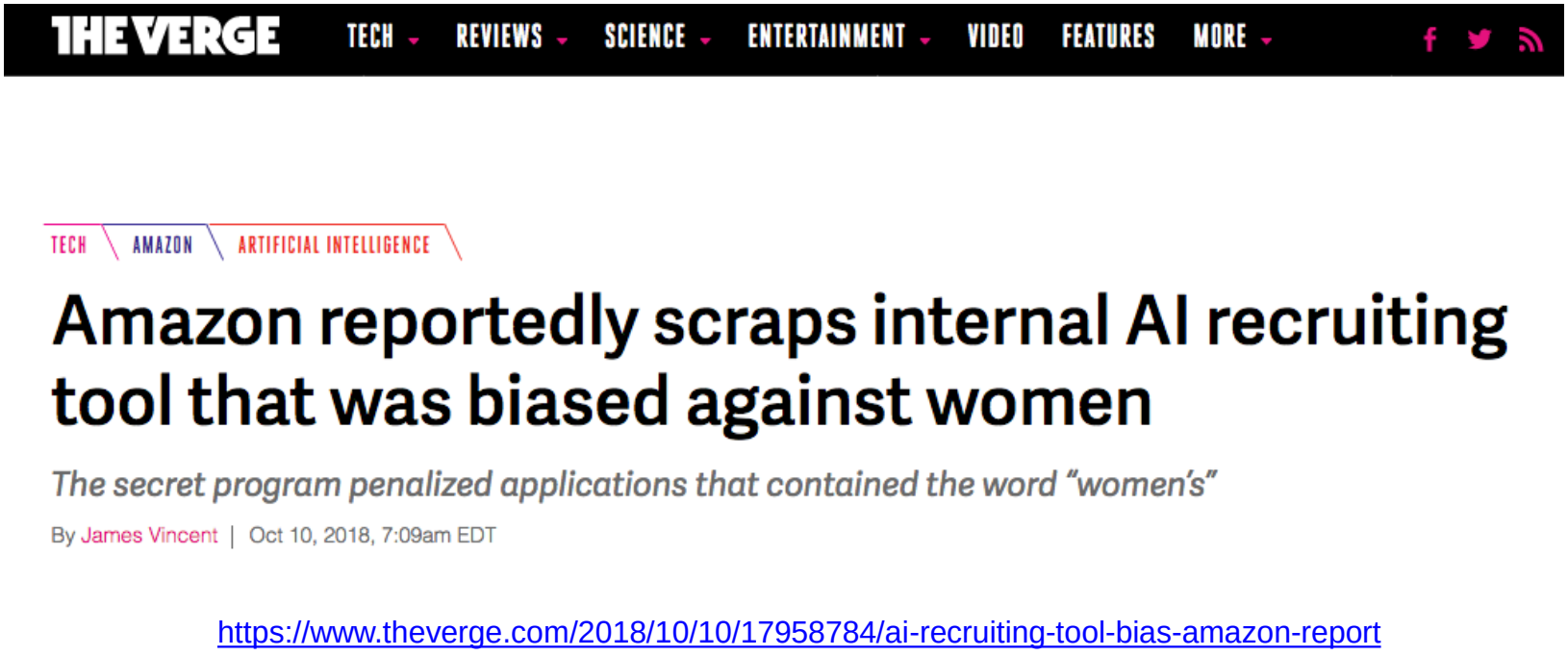
1. ‘Verborgen moraliteit’

KU LEUVEN

Maastricht University

KATLEEN GABRIELS
REGELS VOOR BOTS
ETHIEK IN TIJDEN VAN AI

“Algorithms are opinions embedded in code” (Cathy O’Neil)



https://www.theverge.com/2018/10/10/17958784/ai-recruiting-tool-bias-amazon-report

2. Moraliteit ≠ beslissingsboom

- *Artificial Moral Agent*

nature
International journal of science

Article | Published: 24 October 2018

The Moral Machine experiment

Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff ✉, Jean-François Bonnefon ✉ & Iyad Rahwan ✉

Nature **563**, 59–64 (2018) | Download Citation ↓

https://www.nature.com/articles/s41586-018-0637-6

“O, als het eens zo eenvoudig was! …dat er gewoon duistere lieden op aarde rondliepen die arglistig duistere praktijken bedreven, en dat wij hen alleen maar van de overige mensen hoefden af te scheiden en te vernietigen. Maar de streep die het goede van het kwade scheidt, loopt dwars door het hart van ieder mens. En wié zal er een stuk van zijn hart willen vernietigen?”

De Goelag Archipel
Aleksandr Solzjenitsyn

Philos. Technol.
DOI 10.1007/s13347-017-0284-0

COMMENTARY

The German Ethics Code for Automated and Connected Driving

Christoph Luetge[1]

Received: 24 August 2017 / Accepted: 29 August 2017

“Als je iets ziet dat technisch aanlokkelijk is, dan ga je je gang en maak je het, en pas nadat je je technische succes hebt gehad, discussieer je over wat je ermee gaat doen.
Zo ging het met de atoombom”

J. Robert Oppenheimer

Ethics by design

een ethisch gefundeerde AI,
waarbij ethiek een inherent deel van het ontwerp vormt

van reactief naar proactief

Maastricht University

Mogelijkheden

Moeilijkheden

Ethisch Technologisch Aspectenonderzoek

Maastricht University

Principles for Accountable Algorithms and a Social Impact Statement for Algorithms

Principles for Accountable Algorithms

Automated decision making algorithms are now used throughout industry and government, underpinning many processes from dynamic pricing to employment practices to criminal sentencing. Given that such algorithmically informed decisions have the potential for significant societal impact, the goal of this document is to help developers and product managers design and implement algorithmic systems in publicly accountable ways. Accountability in this context includes an obligation to report, explain, or justify algorithmic decision-making as well as mitigate any negative social impacts or potential harms.

We begin by outlining five equally important guiding principles that follow from this premise:

Algorithms and the data that drive them are designed and created by people -- There is always a human ultimately responsible for decisions made or informed by an algorithm. "The algorithm did it" is not an acceptable excuse if algorithmic systems make mistakes or have undesired consequences, including from machine-learning processes.

https://www.fatml.org/resources/principles-for-accountable-algorithms

What's in the Toolkit:

- **A checklist of 8 risk zones** to help you identify the emerging areas of risk and social harm most critical for your team to start considering now.
- **14 scenarios** to spark conversation and stretch your imagination about the long-term impacts of tech you're building today.
- **7 future-proofing strategies** to help you take ethical action today.

Toolkit & checklist

https://ethicalos.org/

Maastricht University

KATLEEN GABRIELS
REGELS VOOR BOTS
ETHIEK IN TIJDEN VAN AI

Algorithmic Impact Assessment

The AIA is a questionnaire designed to help you assess and mitigate the risks associated with deploying an automated decision system. The AIA also helps identify the impact level of your automated decision system under the proposed Directive on Automated Decision-Making. The questions are focused on your business processes, your data, and your system design decisions.

Publisher - Current Organization Name: Treasury Board of Canada Secretariat

Licence: Open Government Licence - Canada

Resources

Resource Name	Resource Type	Format	Language	Links
Algorithmic Impact Assessment	Tool	HTML	English	Access
Algorithmic Impact Assessment	Tool	HTML	French	Access
Algorithmic Impact Assessment (Old)	Tool	XLS	English	Access
Algorithmic Impact Assessment (Old)	Tool	XLS	French	Access

https://open.canada.ca/data/en/dataset/748a97fb-6714-41ef-9fb8-637a0b8e0da1