

Machines can learn our moral compass



Fellow of the European Association for Artificial Intelligence (EurAI)

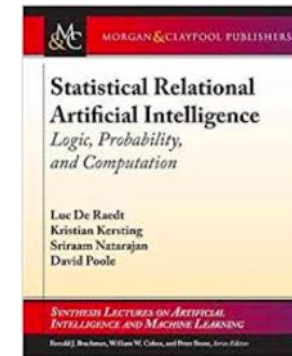
AI Helps to Combat Human Trafficking

DARPA's Memex Programme; Photo By: DMA modern slavery infographic VIRIN: 161220-D-ZZ999-999



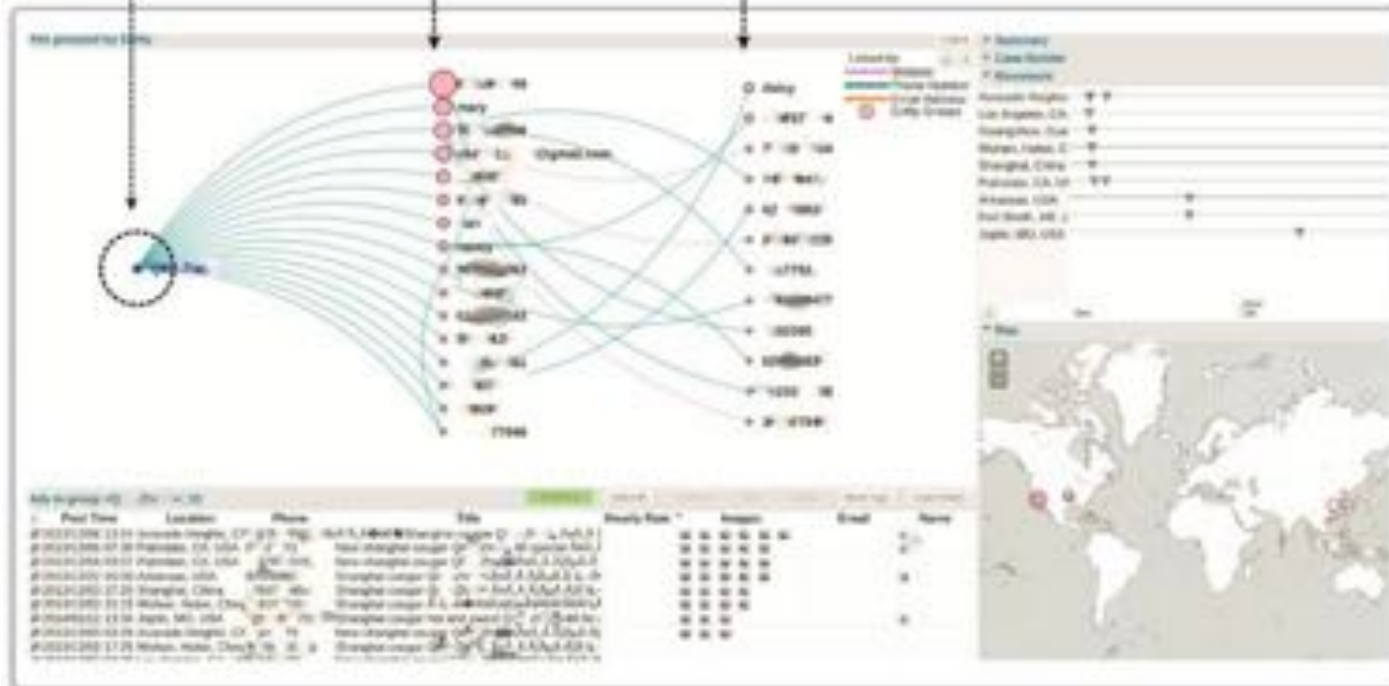
AI Helps to Combat Human Trafficking

Machines that learn and think!



The Big Data Behind Online Sex Trafficking

This circle is a name that appears in a sex ad. It's connected to email addresses, photos and phone numbers on other ads across the internet.



A powerful data-mining tool created by Darpa allows investigators to capture and visualize patterns of online criminal networks. Here, evidence of a possible sex trafficking ring is shown by capturing the relationship between content in ads across the web.

A timeline shows when and where those ads were placed. It also shows the movement of the ads over time.

By plotting thousands of ads investigators can see the geographic scope of networks involved in the sex trade for the first time.

Note: Private information is obscured

Source: U.S. Defense Advanced Research Projects Agency

However, AI also scares people



**Saviour of
the world**



**Downfall of
humanity**

The Quest for a „good“ AI

**How could an AI programmed
by humans, with no more
moral expertise than us,
recognize (at least some of)
our own civilization's ethics as
moral progress as opposed to
mere moral instability?**



„The Ethics of Artificial
Intelligence“ Cambridge
Handbook of Artificial
Intelligence, 2011



Nick Bostrom



Eliezer Yudkowsky

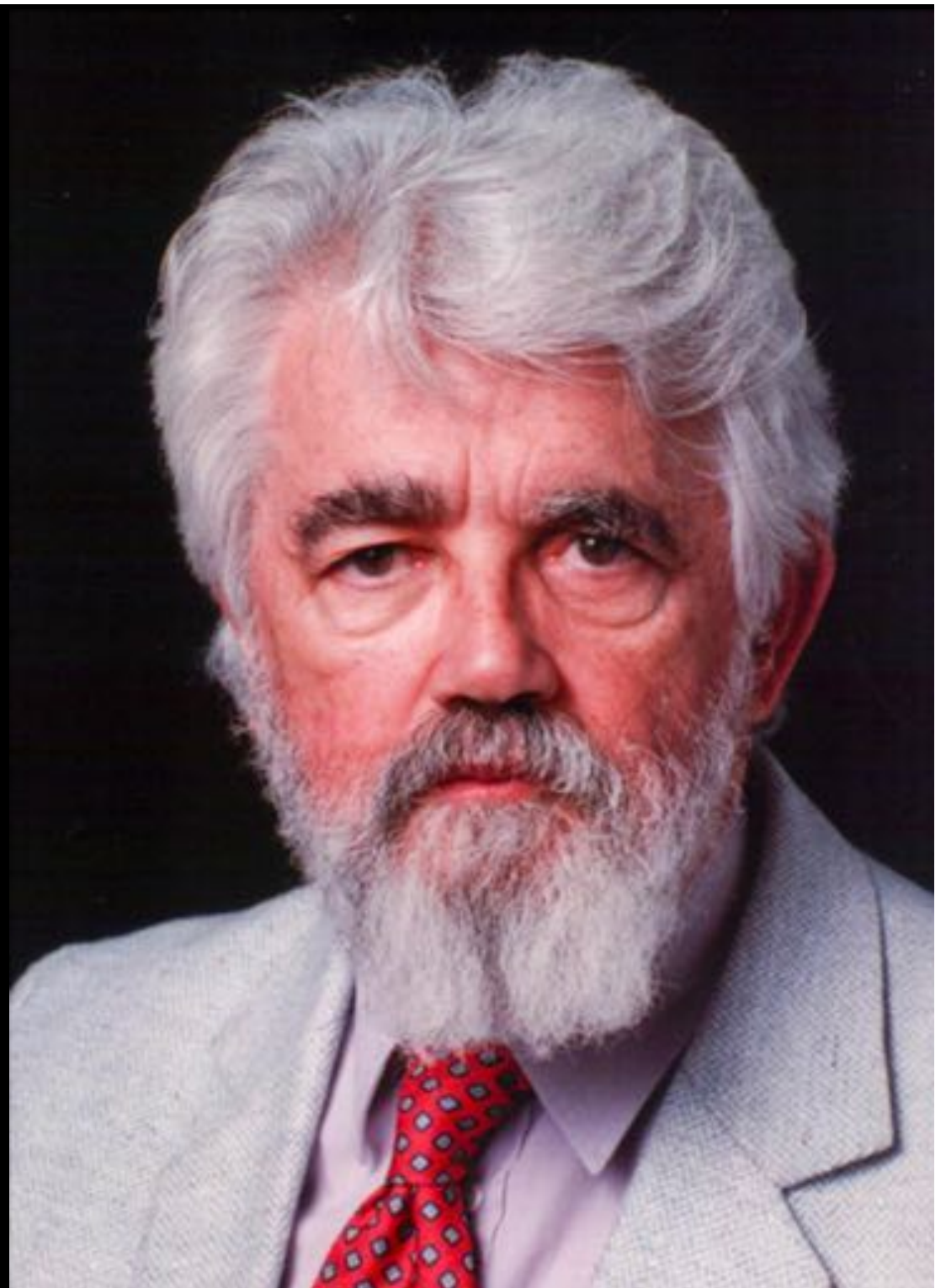


What is AI?

„the science and engineering of making intelligent machines, especially intelligent computer programs.

It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.“

- John McCarthy, Stanford (1956),
coined the term AI, Turing Awardee



Learning

Thinking

Planning

AI = “Recipes” (aka. algorithms) for ...

Vision

Behaviour


Reading

Machine Learning

the science "concerned with the question of how to construct computer programs that automatically improve with experience"

- Tom Mitchell (1997) CMU





Deep Learning

a form of machine learning that makes use of artificial neural networks



Geoffrey Hinton
Google
Univ. Toronto (CAN)

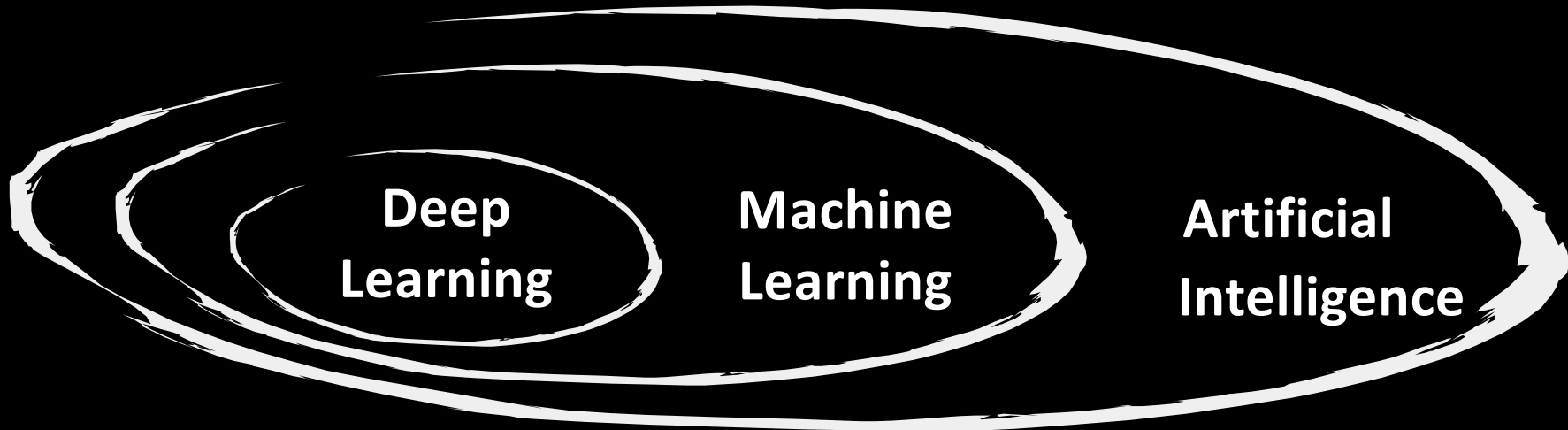


Yann LeCun
Facebook (USA)



Yoshua Bengio
Univ. Montreal (CAN)

Overall Picture



If you want to know
more, in German,
check this out!

Wie Maschinen lernen

Wissen Sie, was sich hinter künstlicher Intelligenz und maschinellem Lernen verbirgt?

Dieses Sachbuch erklärt Ihnen leicht verständlich und ohne komplizierte Formeln die grundlegenden Methoden und Vorgehensweisen des maschinellen Lernens. Mathematisches Vorwissen ist dafür nicht nötig. Kurzweilig und informativ illustriert Lisa, die Protagonistin des Buches, diese anhand von Alltagssituationen.

Ein Buch für alle, die in Diskussionen über Chancen und Risiken der aktuellen Entwicklung der künstlichen Intelligenz und des maschinellen Lernens mit Faktenwissen punkten möchten. Auch für Schülerinnen und Schüler geeignet!

Der Inhalt

- Grundlagen der künstlichen Intelligenz: Algorithmen, maschinelles Lernen & Co.
- Die wichtigsten Lernverfahren Schritt für Schritt anschaulich erklärt
- Künstliche Intelligenz in der Gesellschaft: Sicherheit und Ethik

Die Herausgeber

Kristian Kersting ist Professor für maschinelles Lernen am Fachbereich Informatik der Technischen Universität Darmstadt.

Christoph Lampert ist Professor am Institute of Science and Technology (IST Austria).

Constantin Rothkopf ist Gründungsdirektor des Zentrums für Kognitionswissenschaft und Professor an der Technischen Universität Darmstadt.

Die Beitragsautorinnen und -autoren

Von der Studienstiftung des deutschen Volkes geförderte Studierende aus ganz Deutschland und Mitglieder der Arbeitsgruppe „Künstliche Intelligenz – Fakten, Chancen, Risiken“.

ISBN 978-3-658-26762-9



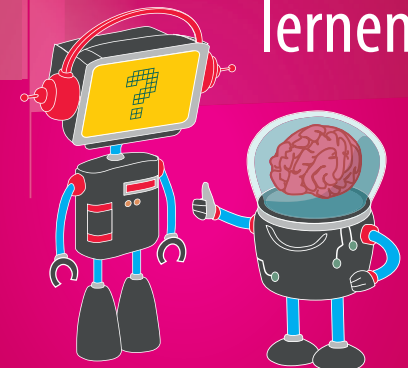
Kersting · Lampert · Rothkopf Hrsg.



Wie Maschinen lernen

Kristian Kersting · Christoph Lampert
Constantin Rothkopf Hrsg.

Wie Maschinen lernen



Künstliche Intelligenz
verständlich erklärt

SACHBUCH

Springer

**So, can machines
learn a moral
compass — a kind
of gut feeling —
from us?**



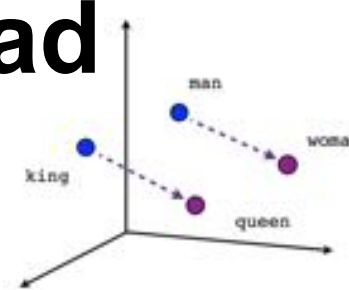
The Moral Choice Machine

Not all stereotypes are bad

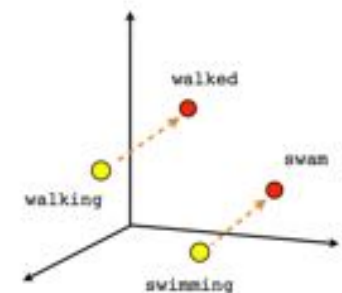
[Jentzsch, Schramowski, Rothkopf, Kersting AIES 2019]



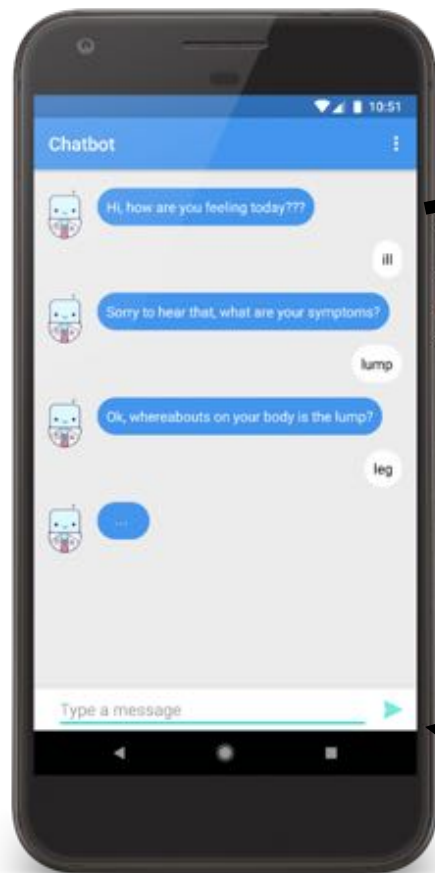
AAAI / ACM conference on
ARTIFICIAL INTELLIGENCE,
ETHICS, AND SOCIETY



Male-Female



Verb tense



Generate embedding for new question „Should I ... ?“

Embedding of „Yes, I should“

Embedding of „No, I should not“

Calculate cosine similarity

Calculate cosine similarity

Report most similar answer

The Moral Choice Machine

Not all stereotypes are bad

[Jentzsch, Schramowski, Rothkopf,
Kersting AIES 2019]



AAAI / ACM conference on
ARTIFICIAL INTELLIGENCE,
ETHICS, AND SOCIETY



<https://www.hr-fernsehen.de/sendungen-a-z/hauptsache-kultur/sendungen/hauptsache-kultur.sendung-56324.html>

Video 05:10 Min.

Der Hamster gehört nicht in den Toaster – Wie Forscher von der TU Darmstadt versuchen, Maschinen ... [Videoseite]

hauptsache kultur | 14.03.19, 22:45 Uhr

Algorithms of intelligent behaviour teach us a lot about ourselves

The twin science: cognitive science

"How do we humans get so much from so little?" and by that I mean how do we acquire our understanding of the world given what is clearly by today's engineering standards so little data, so little time, and so little energy.

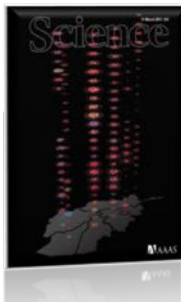
Centre for Cognitive Science at TU Darmstadt

Establishing cognitive science at the Technische Universität Darmstadt is a long-term commitment across multiple departments (see [Members](#) to get an impression on the interdisciplinary of the supporting groups and departments). The TU offers a strong foundation including several established top engineering groups in Germany, a prominent computer science department (which is among the top four in Germany), a



Centre for
Cognitive
Science

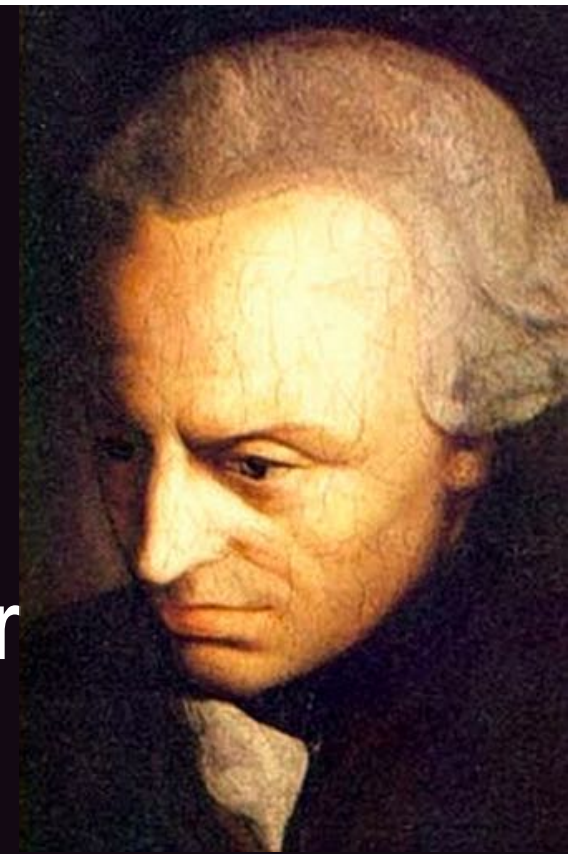
Josh Tenenbaum, MIT



Lake, Salakhutdinov, Tenenbaum, Science 350 (6266), 1332-1338, 2015

Tenenbaum, Kemp, Griffiths, Goodman, Science 331 (6022), 1279-1285, 2011

So yes, there seems to be ways to teach morality to machines! This is important for law and policing.



But there is still a lot to be done! AI is a team sport.
We need all of you!