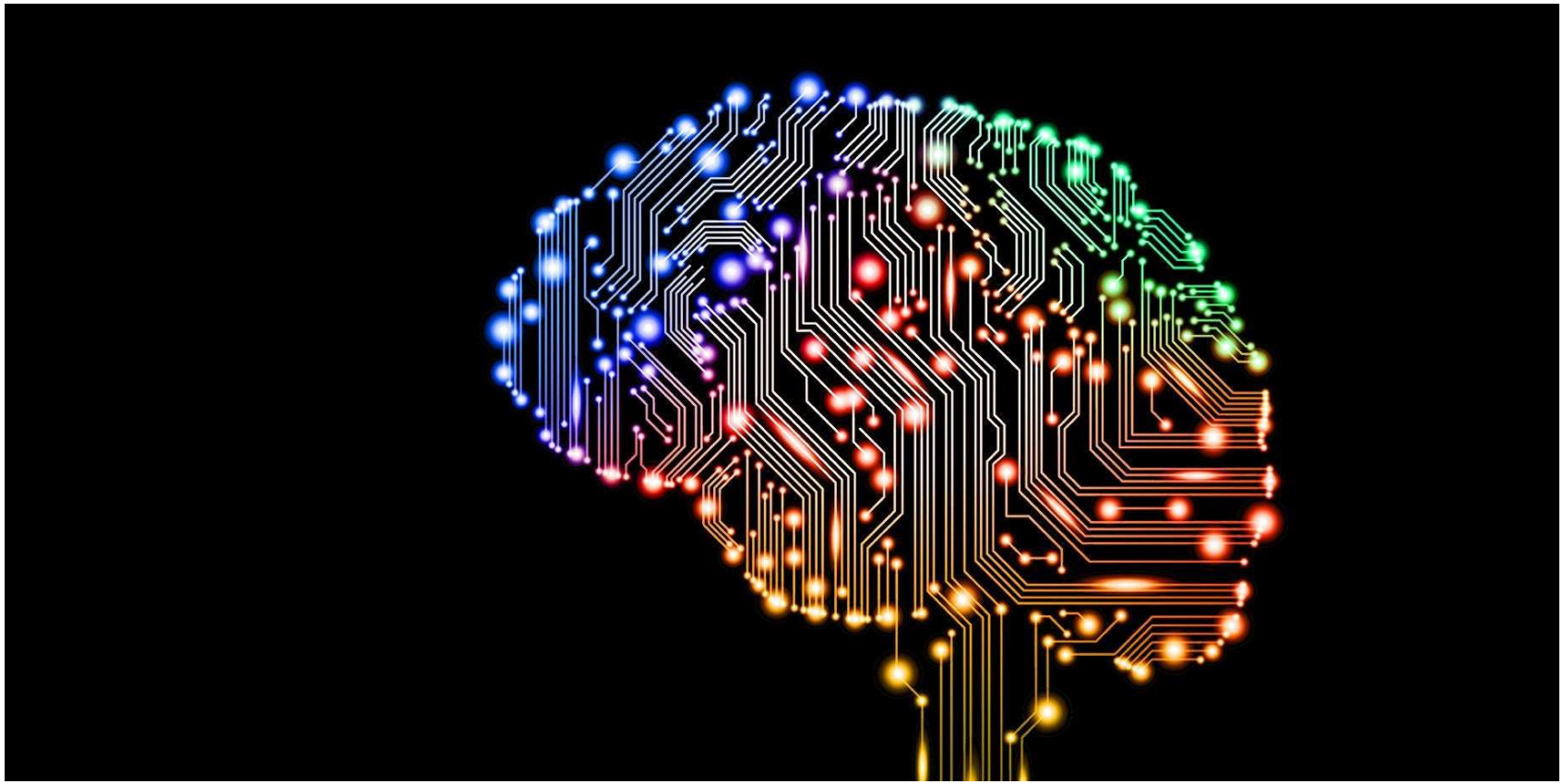
Künstliche Intelligenz

Kolloquium Zivilgesellschaft, 10. November 2017

Kaspar Etter
www.superintelligence.ch



Einleitung

Worüber reden wir?

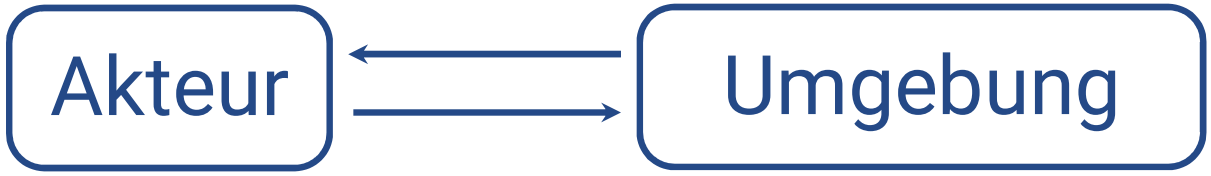
Intelligenz

«Intelligenz misst die Fähigkeit eines Akteurs, seine Ziele in einem weiten Spektrum von unbekannten Umgebungen zu erreichen.»

$$\text{Intelligenz} = \frac{\text{Optimierungskraft}}{\text{Ressourceneinsatz}}$$

Universal Intelligence
arxiv.org/pdf/0712.3329.pdf

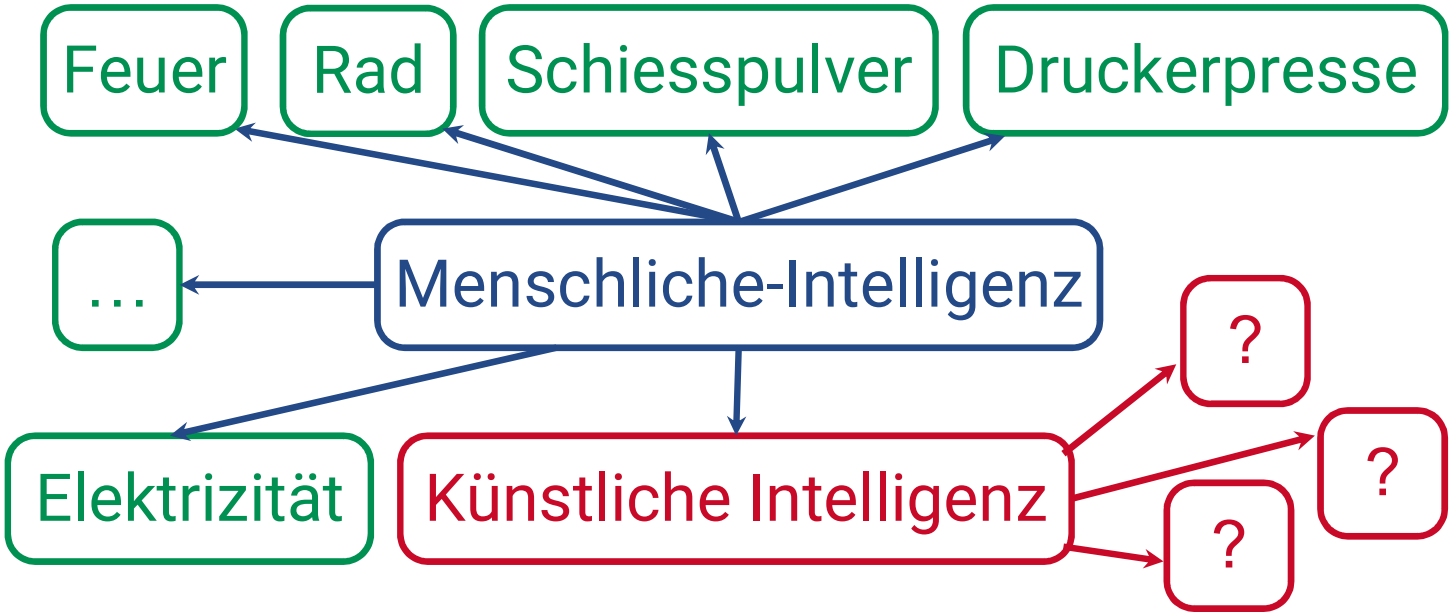
Interaktion



Lernen, voraussagen, bewerten und planen!

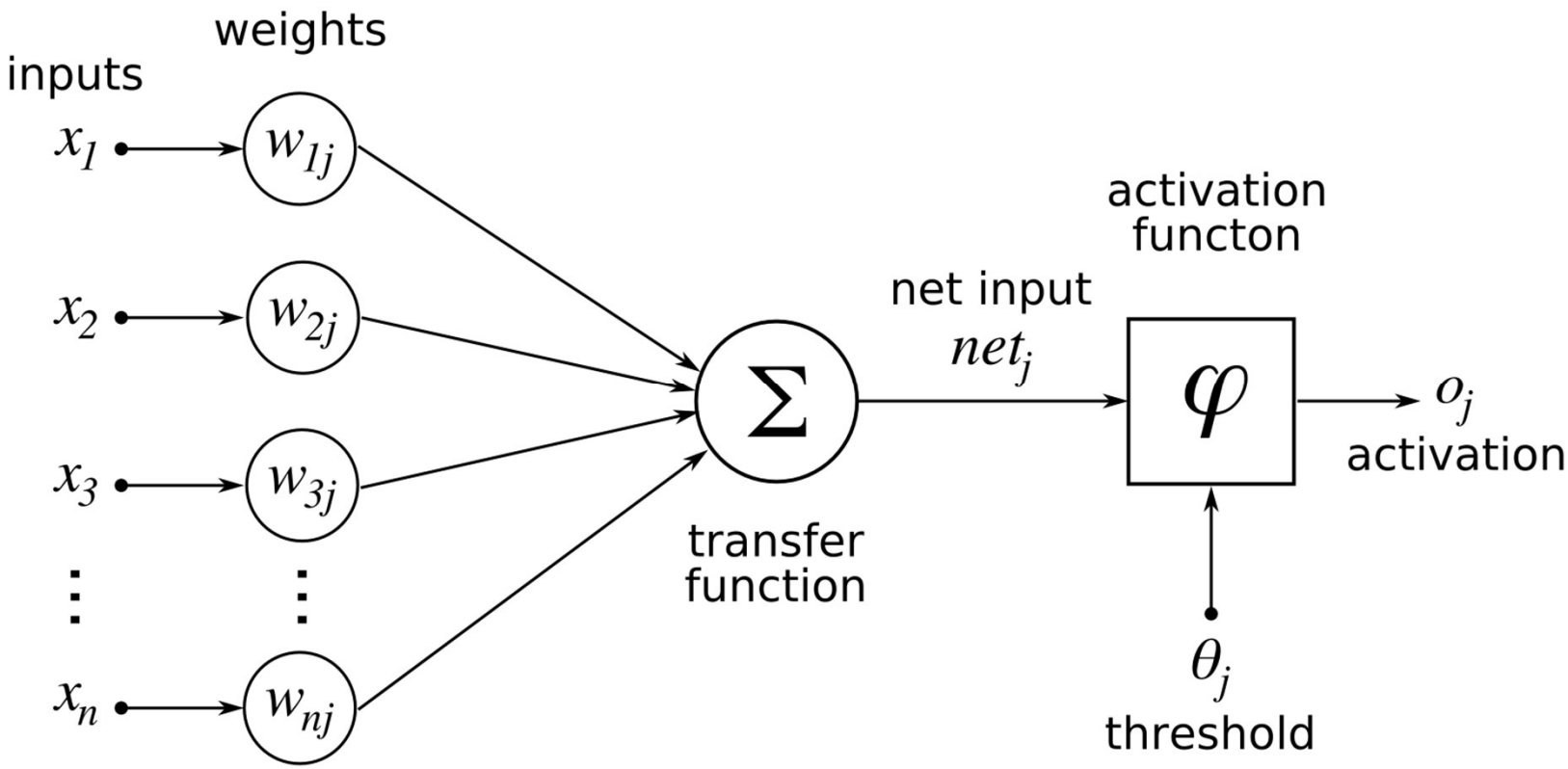
Marcus Hutter: Universal Artificial Intelligence
www.youtube.com/watch?v=I-vx5zbOOXI

Künstliche Intelligenz (KI)



KI ist eine Technologie ohne Vergleich!

Our Final Invention
www.jamesbarrat.com



Maschinelles Lernen

Wie funktioniert das?

Lernen

- Modelle mit Vorhersage-Kraft aus Daten erstellen
- Algorithmen werden besser durch «Erfahrung»
- Praktisch, wenn man selber keine Lösung kennt

- **Maschinelles Lernen**: Lernen und Vorhersagen
- **Data Mining**: Neue Zusammenhänge entdecken

Machine Learning
en.wikipedia.org/wiki/Machine_learning

Rückkoppelung

Kategorien von KI je nach Art der Rückkoppelung:

- Überwachtes Lernen: Aus einer Menge von Inputs und Outputs eine Regel finden
- Unüberwachtes Lernen: Struktur im Input finden ohne Trainingsdaten
- Selbstverstärkendes Lernen: Ein Ziel in einer dynamischen Umgebung erreichen

Supervised Learning
en.wikipedia.org/wiki/Supervised_learning

Aufgaben und Modelle

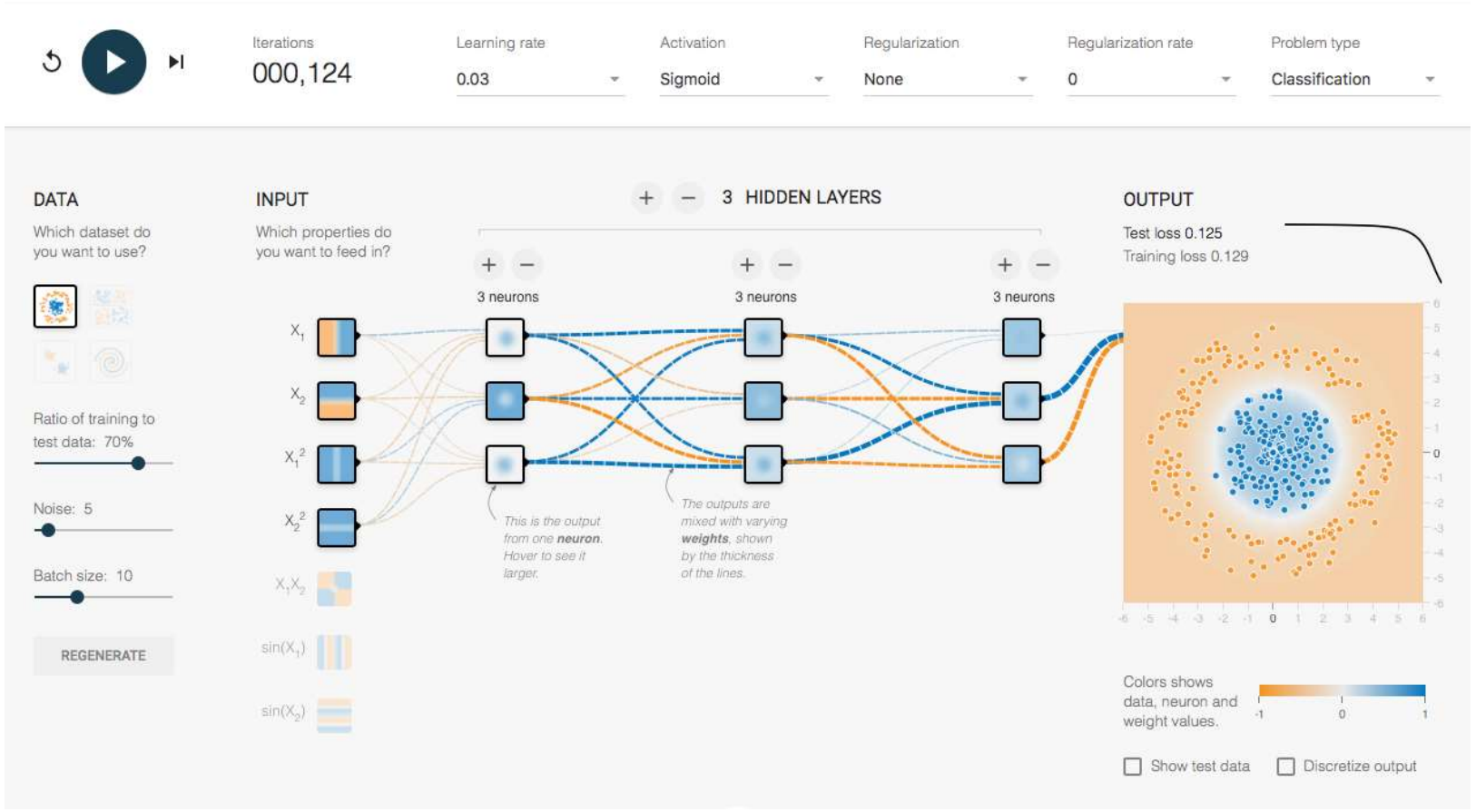
- **Klassifizierung**: Zuordnung von unbekanntem Input zu diskreten Kategorien (überwacht)
- **Regression**: Abbildung von Input auf kontinuierliche Werte (überwacht)
- **Gruppierung**: Automatische Aufteilung einer Menge von Inputs in verschiedene Gruppen (unüberwacht)
- **Reduktion der Dimensionen**: Vereinfachung der Inputs nach Gemeinsamkeiten (unüberwacht)

Statistical Classification
en.wikipedia.org/wiki/Statistical_classification

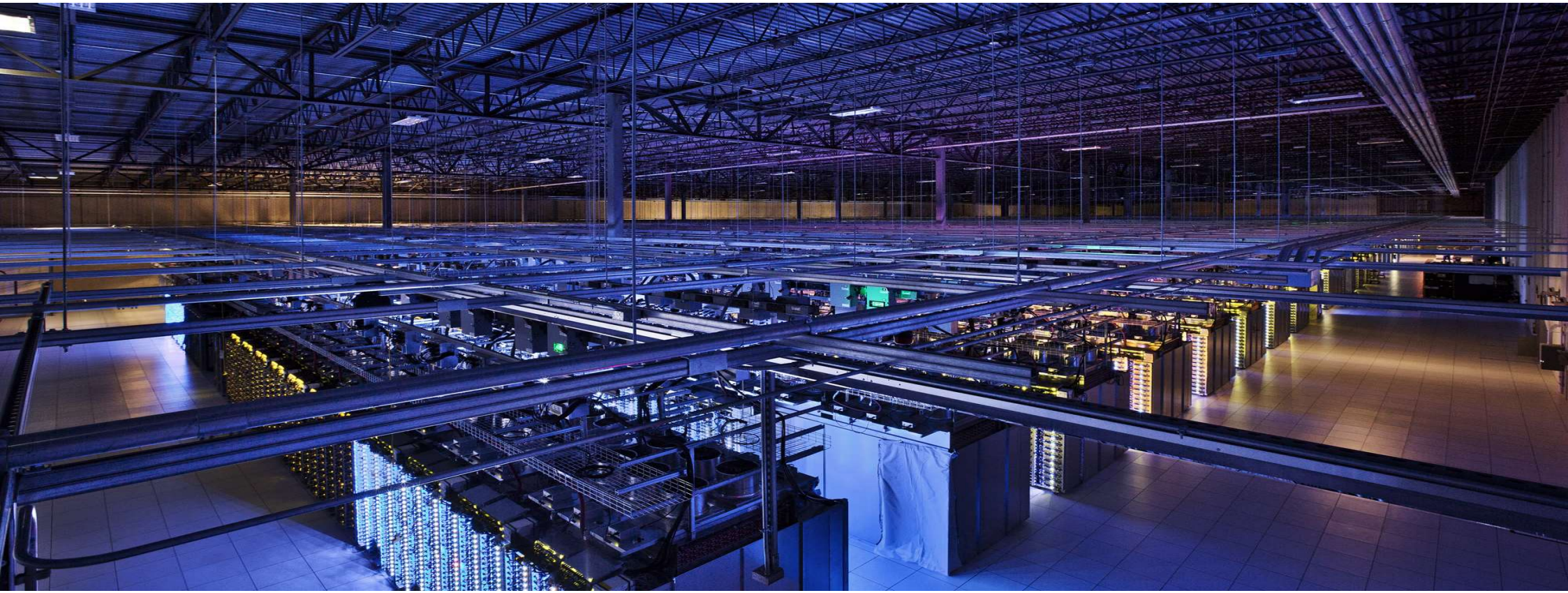
Reinforcement Learning (CNN)

Google develops self-learning computer program
www.theguardian.com/technology/2015/feb/25/[…]

Trainiere dein eigenes Neuronales Netzwerk



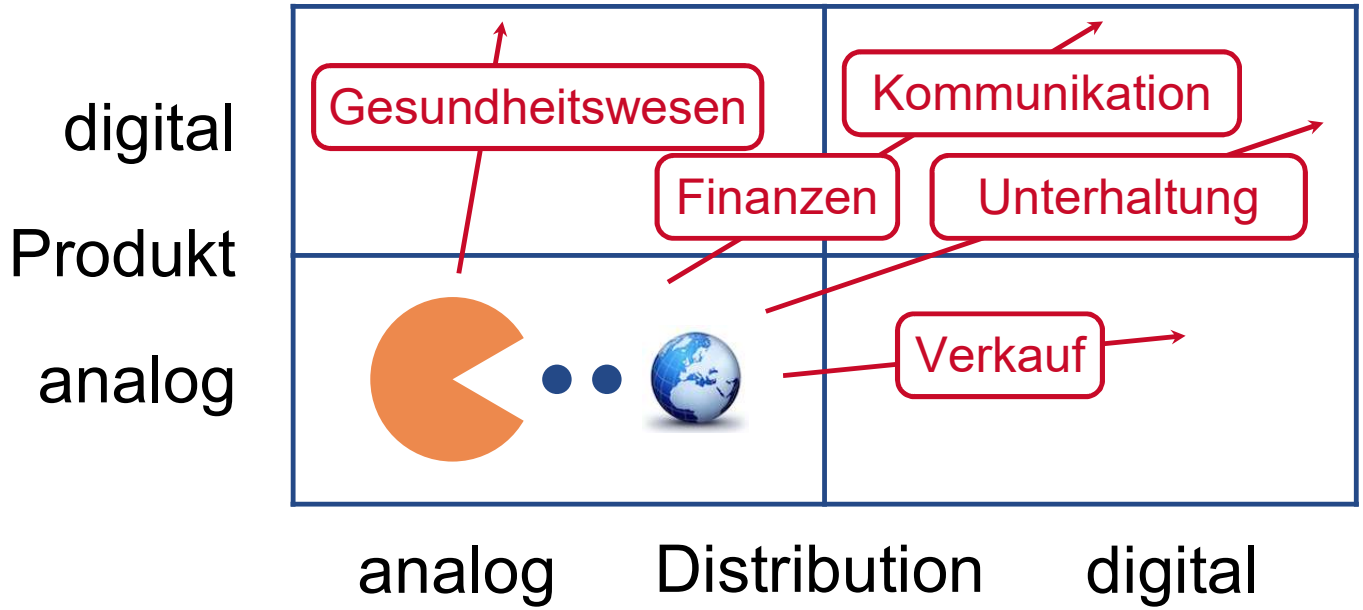
Neural Network in Browser
playground.tensorflow.org



Aktuelle Trends

Wohin geht die Reise?

«Software is eating the world»



… und KI wird die Datenberge verdauen!

Marc Andreessen
on.wsj.com/1w2FbVs

State of the Art

Deep Blue: 1997

Stanley: 2005

IBM Watson: 2011

Schmidhuber: 2011

Checkers	Superhuman
Backgammon	Superhuman
Othello	Superhuman
Chess	Superhuman
Crosswords	Expert Level
Scrabble	Superhuman
Bridge	Equal to Best
Jeopardy!	Superhuman
Poker	Varied
FreeCell	Superhuman
Go	Strong Amateur

Superhuman

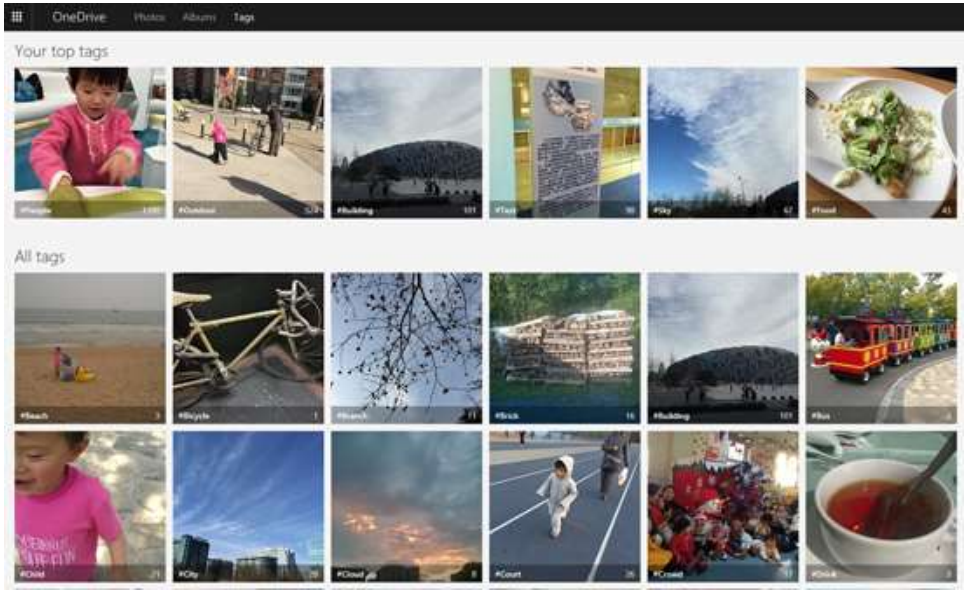
How bio-inspired learning keeps winning competitions
www.kurzweilai.net/how-bio-inspired-deep-learning-[…]

Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image
A person riding a motorcycle on a dirt road.	Two dogs play in the grass.	A skateboarder does a trick on a ramp.	A dog is jumping to catch a frisbee.
A group of young people playing a game of frisbee.	Two hockey players are fighting over the puck.	A little girl in a pink hat is blowing bubbles.	A refrigerator filled with lots of food and drinks.
A herd of elephants walking across a dry grass field.	A close up of a cat laying on a couch.	A red motorcycle parked on the side of the road.	A yellow school bus parked in a parking lot.

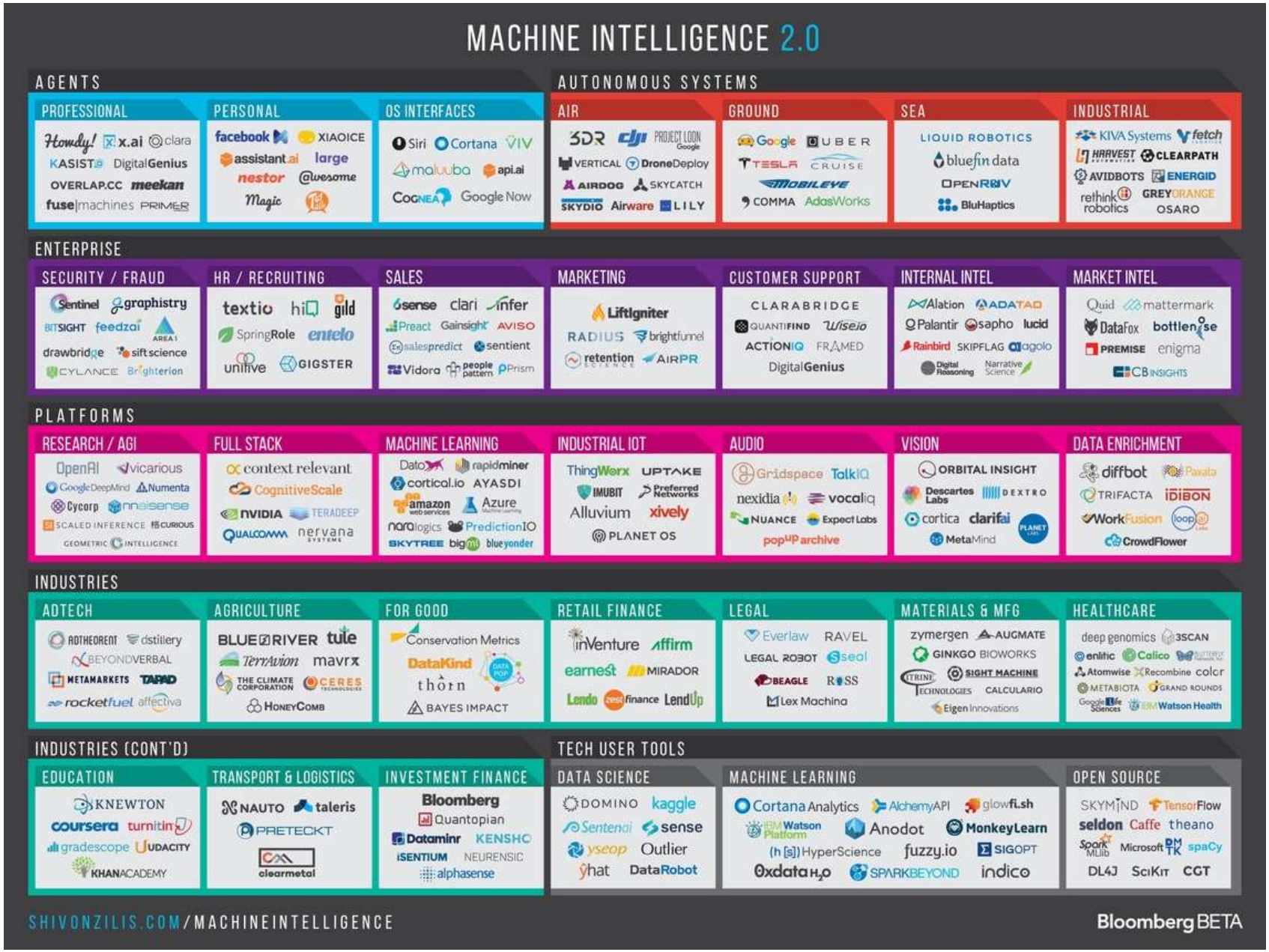
Google Research Blog, November 2014
googleresearch.blogspot.ch/2014/11/[…]

Übermenschliche Bilderkennung

- Mit gefalteten neuronalen Netzen
- 1.2 Mio. Trainingsbilder, ~ 30 Ebenen



Inside Microsoft Research, February 2015
blogs.technet.com/b/inside_microsoft_research/[…]


MACHINE INTELLIGENCE 2.0
AGENTS
PROFESSIONAL
Howdy! x.ai clara
KASISTO DigitalGenius
OVERLAP.CC meekan
fuse|machines PRIMER
PERSONAL
facebook XIAOICE
assistant.ai large
nestor @wesome
Magic
OS INTERFACES
Siri Cortana VIV
maluuba api.ai
COGNEA Google Now
AUTONOMOUS SYSTEMS
AIR
3DR DJI PROJECT LOON
VERTICAL DroneDeploy
AIRDOG SKYCATCH
SKYDIO Airware LILY
GROUND
Google UBER
TESLA CRUISE
MOBILEYE
COMMA AdasWorks
SEA
LIQUID ROBOTICS
bluefin data
OPENROV
BluHaptics
INDUSTRIAL
KIVA Systems fetch
HARVEST CLEARPATH
AVIDBOTS ENERGID
rethink robotics GREYORANGE
OSARO
ENTERPRISE
SECURITY / FRAUD
Sentinel graphistry
BITSIGHT feedzai AREA 1
drawbridge siftscience
CYLANCE Brighterion
HR / RECRUITING
textio hiQ gild
SpringRole entelo
unitive GIGSTER
SALES
6sense clari infer
Preact Gainsight AVISO
salespredict sentient
Vidora people pattern Prism
MARKETING
LiftIgniter
RADIUS brightfunnel
retention SCIENCE AIRPR
CUSTOMER SUPPORT
CLARABRIDGE
QUANTIFIND Wise.io
ACTIONIQ FRAMED
DigitalGenius
INTERNAL INTEL
Alation ADATAD
Palantir sapho lucid
Rainbird SKIPFLAG algolo
Digital Reasoning Narrative Science
MARKET INTEL
Quid mattermark
DataFox bottlenose
PREMISE enigma
CB INSIGHTS
PLATFORMS
RESEARCH / AGI
OpenAI vicarious
Google DeepMind Numenta
Cycorp nnaisense
SCALED INFERENCE CURIOUS
GEOMETRIC INTELLIGENCE
FULL STACK
context relevant
CognitiveScale
NVIDIA TERADEEP
QUALCOMM nervana
MACHINE LEARNING
Dato rapidminer
cortical.io AYASDI
amazon web services Azure
narrologics PredictionIO
SKYTREE bigml blue yonder
INDUSTRIAL IOT
ThingWorx UPTAKE
IMUBIT Preferred Networks
Alluvium xively
PLANET OS
AUDIO
Gridspace TalkIQ
nexidia vocaliq
NUANCE Expect Labs
popUP archive
VISION
ORBITAL INSIGHT
Descartes Labs DEXTRO
cortica clarifai
MetaMind PLANET LABS
DATA ENRICHMENT
diffbot Paxata
TRIFACTA iDIBON
WorkFusion loop
CrowdFlower
INDUSTRIES
ADTECH
ADTHEORENT dstillery
BEYONDVERBAL
METAMARKETS TAPAD
rocketfuel affectiva
AGRICULTURE
BLUE RIVER tule
TerrAvion mavrx
THE CLIMATE CORPORATION CERES TECHNOLOGIES
HONEYCOMB
FOR GOOD
Conservation Metrics
DataKind DATA POP
thorn
BAYES IMPACT
RETAIL FINANCE
inVenture Affirm
earnest MIRADOR
Lendo zestfinance LendUp
LEGAL
Everlaw RAVEL
LEGAL ROBOT seal
BEAGLE ROSS
Lex Machina
MATERIALS & MFG
zymergen AUGMATE
GINKGO BIOWORKS
TRINE TECHNOLOGIES SIGHT MACHINE
CALCULARIO
Eigen Innovations
HEALTHCARE
deep genomics 3SCAN
enlitic Calico
Atomwise Recombine color
METABIOTA GRAND ROUNDS
Google Life Sciences IBM Watson Health
INDUSTRIES (CONT'D)
EDUCATION
KNEWTON
coursera turnitin
gradescope UDACITY
KHANACADEMY
TRANSPORT & LOGISTICS
NAUTO taleris
PRETECKT
CMA clearmetal
INVESTMENT FINANCE
Bloomberg
Quantopian
Dataminr KENSHO
iSENTIUM NEURENSIC
alphasense
TECH USER TOOLS
DATA SCIENCE
DOMINO kaggle
Sentenai sense
yseop Outlier
ŷhat DataRobot
MACHINE LEARNING
Cortana Analytics AlchemyAPI glowfi.sh
IBM Watson Platform Anodot MonkeyLearn
(h[s]) HyperScience fuzzy.io SIGOPT
0xdata H2O SPARKBEYOND indico
OPEN SOURCE
SKYMIND TensorFlow
seldon Caffe theano
Spark MLlib Microsoft DMTK spaCy
DL4J SciKit CGT
SHIVONZILIS.COM/MACHINEINTELLIGENCE
Bloomberg BETA

Ausblick

Was kommt als Nächstes?

Automation schafft Überfluss

Viele Industrien werden sich zu reinen KI-Industrien wandeln. Dadurch werden die Kapitalerträge zunehmen, die Löhne hingegen abnehmen.

Das wäre an sich eine positive Entwicklung – aber wir sind nicht darauf vorbereitet!

Erik Brynjolfsson
go.ted.com/biXe

Ein rationaler Akteur wird versuchen…

… funktional zu bleiben (deckt sich mit «Eigeninteresse»)

… seine Ziele aufrechtzuerhalten (Zielbewahrung)

… Materialien zu erwerben (Ressourcenakkumulation)

… klüger zu werden (Intelligenzexplosion)

Das Problem mit der KI ist nicht Böswilligkeit, sondern von den Menschen abweichende Zielsetzungen und bessere Entscheidungen.

Stephen M. Omohundro: The Basic AI Drives
selfawaresystems.[…].com/2008/01/ai_drives_final.pdf

Auswirkungen von (fortgeschrittener) KI

- Kurzfristig (heute bis in 15 Jahren): Klassifizierung/Assistenz, Überwachung/Geheimdienst, autonome Systeme (Autos, Waffen, etc.) > KIs als moralische Akteure
- Mittelfristig (in 15 bis 40 Jahren): Automatisierung & Arbeitslosigkeit, Kooperation & Wettrüsten, Kontrolle & Prinzipal-Agent-Problem, Orakelproblem > unbeabsichtigte Effekte
- Langfristig (in 40 bis 100 Jahren): Durchbrüche in Wissenschaft mit KIs, Rekursive Selbstverbesserung/Ausreisser-KI, künstliches Bewusstsein > KIs als moralische Patienten

Potential risks from advanced artificial intelligence
www.openphilanthropy.org/blog/potential-risks-[…]

Aus grosser Macht folgt grosse Verantwortung!

Kaspar Etter
www.superintelligence.ch