

# Künstliche Intelligenz — Wohin geht die Reise?

Christoph Benzmüller

Dahlem Center for Machine Learning and Robotics, FU Berlin

*“The development of full artificial intelligence could spell the end of the human race.” (S. Hawking)*

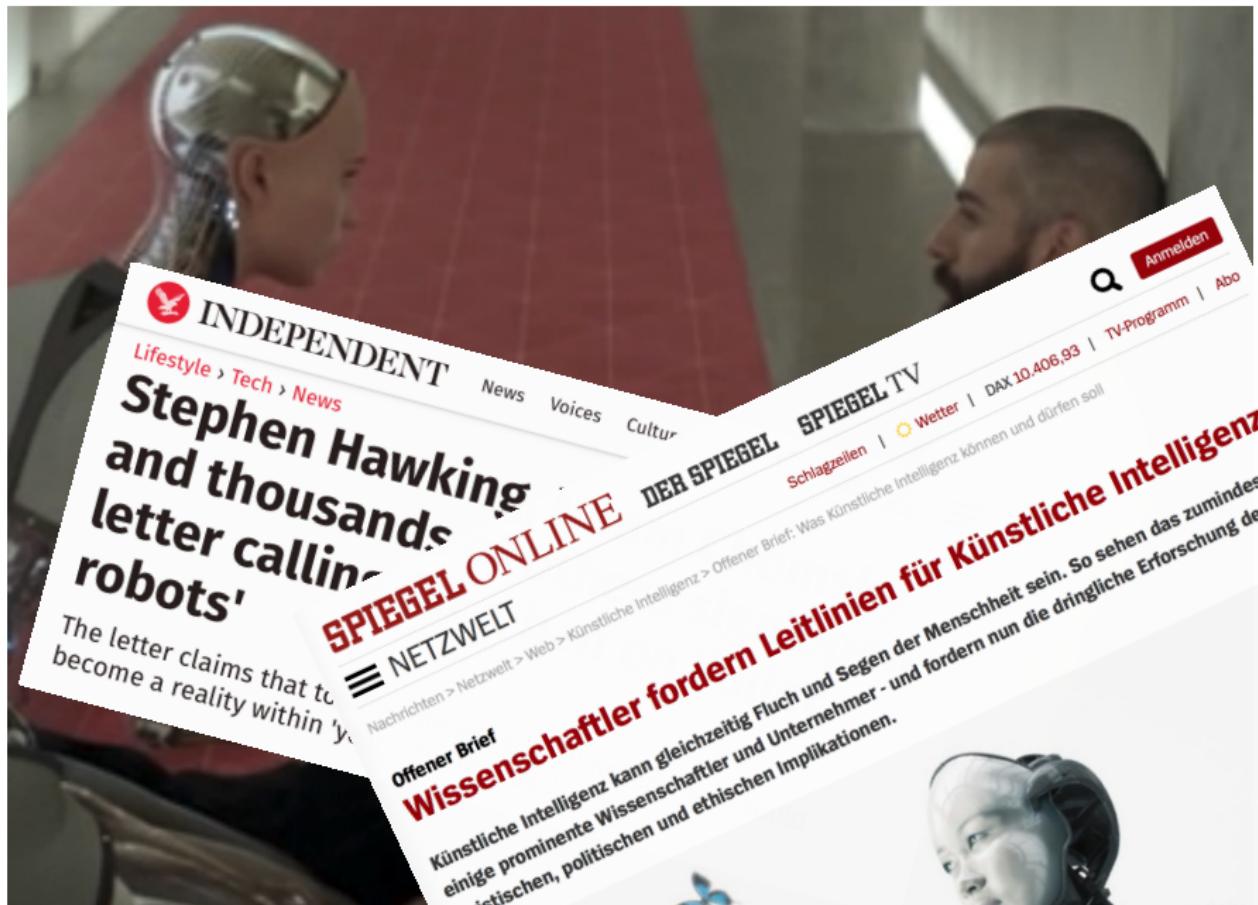


*“I am in the camp that is concerned about super intelligence.” (B. Gates)*

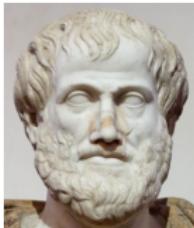
Panikmache? Oder ist da was dran?

*“I think we should be very careful about AI. If I were to guess what our biggest existential threat is, it’s probably that.” (E. Musk)*





### KI-Visionäre:



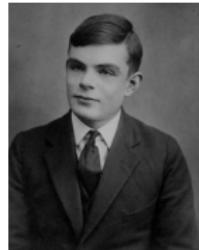
Aristoteles



R. Lull



G. W. Leibniz



A. Turing

### Einige Eckdaten:

- ▶ 1956: Konferenz am Dartmouth College (New Hampshire) mit J. McCarthy, M. Minsky, A. Newell und H. Simon
- ▶ seit Mitte 70er: Aufbau von KI-Forschergruppen in Deutschland
- ▶ Anfang 80er: 5th Generation Projekt in Japan (450 million \$)
- ▶ 1983: Erste offizielle KI-Professur in Deutschland (Kaiserslautern)
- ▶ 1988: Gründung des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI), heute ca. 800 MA

KI-Sommer



KI-Winter

Mitte 50er-Mitte 70er

Mitte 70er-Ende 70er

Anfang 80er-Mitte/Ende 80er

Mitte/Ende 80er-Mitte 90er

## Definitionsversuche gibt es viele — aber keinen Konsens

- ▶ *AI is 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. (McCarthy)*
- ▶ *The study of how to make computers do things at which, at the moment, people are better. (Rich and Knight)*
- ▶ ...

### Starke KI

Maschinen mit (mind.)  
allgemeiner menschlicher  
Intelligenz

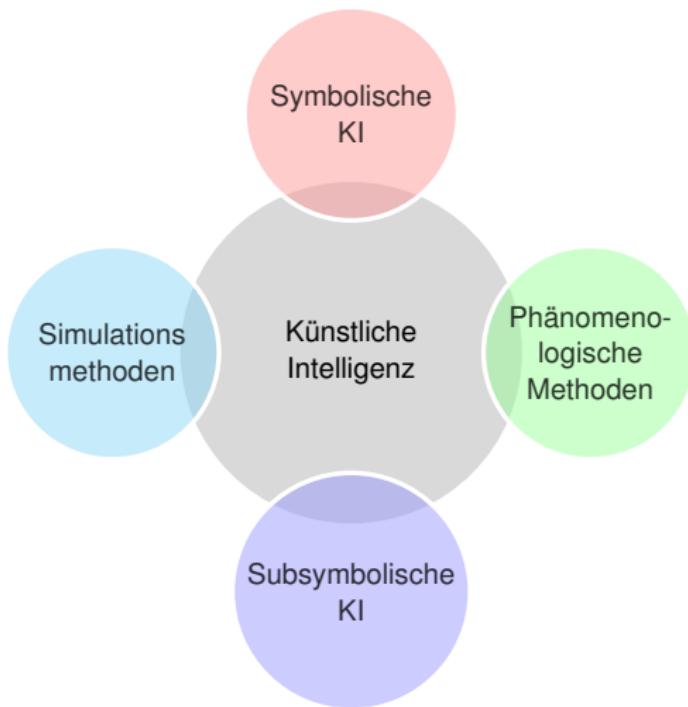
vs.

### Schwache/Angewandte KI

Eingeschränkte, 'clevere'  
Systeme die Intelligenz  
simulieren

### Kognitive Simulation

Testen von Theorien zur  
menschlichen Kognition  
mithilfe von Computern



- ▶ Wissensrepräsentation und Schließen
- ▶ Planen
- ▶ Agentensysteme
- ▶ Suchen
- ▶ ...
- ▶ Sprachverarbeitung
- ▶ ...
- ▶ Robotik
- ▶ Maschinelles Lernen, Mustererkennung
- ▶ Künstliche Neuronale Netzwerke

**Einfluss von/auf:** Informatik, Psychologie, Neurologie und Neurowissenschaften, Logik, Kommunikationswissenschaft, Philosophie und Linguistik

### Beispiele:

- ▶ ELIZA (Ende 60er):  
einfaches Dialogsystem  
(Psychotherapie-Simulation)
- ▶ MYCIN (ab Anfang 70er):  
Expertensystem, Diagnose  
von Blutinfektionskrankheiten
- ▶ Computerbeweis des  
Vierfarbenproblems (1976)
- ▶ Deep Blue (1997):  
Schachprogramm schlägt  
Garry Kasparow



Deep Blue  
IBM chess computer

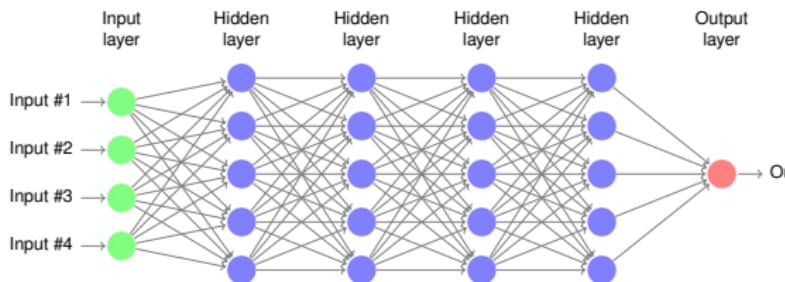
Garry Kasparov  
World Chess Champion

CC by 2.0/4.0

Diese Erfolge sind der **Symbolischen KI** zuzuordnen

**Beispiele:**

- ▶ Intelligente Software Assistenten (Siri & Co)
- ▶ 2011: Watson, Jeopardy-Quiz
- ▶ 2016: AlphaGo, schlägt Go-Champion L. Sedol

**Neue Erfolge: Subsymbolische KI und Kombination von Techniken****Tiefe Neuronale Netzwerke****AlphaGo**

- ▶ 19x19 Eingabebild der Spielkonfiguration
- ▶ Tiefes Neuronales Netzwerk mit etwa 13 Schichten
- ▶ Kombination von Methoden
  - ▶ Monte Carlo Baumsuche
  - ▶ Überwachtes Lernen (Supervised Learning)
  - ▶ Bestärkendes Lernen (Reinforcement Learning)

## Neue Erfolge in der Mechanisierung/Automatisierung logischen Schließens:

- ▶ Computerbeweis von Kepler's Vermutung (interaktiv)
- ▶ SAT Solver lösen offene mathematische Probleme (automatisch)
- ▶ Eigene Arbeit zum Ontologischen Gottesbeweis auf dem Computer (automatisch)

Aussage (und Annahmen)

$\Box \exists x.G(x)$

$$\frac{A \quad A \Rightarrow B}{B} \text{ modus ponens}$$



CC Rainer Zenz

**SPIEGEL ONLINE WISSENSCHAFT**

Login | Registrierung

Politik | Wirtschaft | Panorama | Sport | Kultur | Netzwerk | Wissenschaft | Gesundheit | einestages | Karriere | Uni | Schule | Reise | Auto

Nachrichten > Wissenschaft > Mensch > Mathematik > Formel von Kurt Gödel: Mathematiker bestätigen Gottesbeweis

**Formel von Kurt Gödel: Mathematiker bestätigen Gottesbeweis**

Von Tobias Hürter

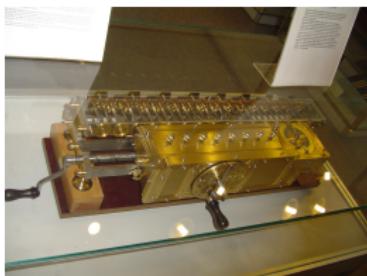


C. Benzmüller    Künstliche Intelligenz    Wohin geht die Reise?    11. Jahresforum Shared Services & Outsourcing Woche, Berlin, 2016

## Leibniz' Vision: *Calculemus!*



Leibniz (1646–1716)



Rechenmaschine

*"Quo facto, quando orientur controversiae, non magis disputatione opus erit inter duos philosophos, quam inter duos Computistas. Sufficiet enim calamos in manus sumere sedereque ad abacos, et sibi mutuo . . . dicere: calculemus."* (Leibniz, 1684)



**Streitende Philosophen klären Disput durch Formalisierung und mechanische Beweisführung**

# Kurt Gödel's Ontologischer Gottesbeweis: 1930's, 1941, 1946-1955, 1970

Ontologischer Beweis FEB 10, 1970

P( $\varphi$ )  $\varphi$  is positive ( $\Leftrightarrow \varphi \in P$ )

At. 1  $P(\varphi), P(\neg\varphi)$

P1  $G(x) =$

P2  $\varphi \text{ Es ist}$

P3  $\neg\varphi \text{ Es ist}$

At. 2  $P(\varphi) \vee P(\neg\varphi)$

Neue, relevante Beiträge im recht gut erforschten Bereich der Ontologischen Gottesbeweise (Metaphysik) durch

Automatische Theorembeweiser!

merely Exist.

<http://www.ijcai.org/Proceedings/16/Papers/137.pdf>

<http://dx.doi.org/10.3233/978-1-61499-419-0-93>

Th.  $G(x) > G \text{ Em. } x$

Df.  $E(x) \equiv \exists p [\varphi \text{ Es ist } N \exists x \varphi(x)]$

At. 3  $P(E)$

Th.  $G(x) > N(\exists y) G(y)$

At. 4  $(\exists x) G(x) > N(\exists y) G(y)$

"  $M(\exists x) G(x) > MN(\exists y) G(y)$

"  $\Rightarrow N(\exists y) G(y)$

M = positivity

Weltweit erste Vorlesung Komputationale Metaphysik an der FU Berlin (ausgezeichnet mit zentralem Lehrpreis der FU)

<http://www.inf.fu-berlin.de/users/lex/lehre/compmeta/>

(With permission of Leon Levy Archives Center, Princeton, NJ, USA, on deposit at Princeton University)

# Künstliche Intelligenz

Ray Kurzweil conversation with chatbot Eugene Goostman

**Question:** How old are you?

**Eugene:** I'm a little boy. 13 years old.

**Question:** How much is two plus two?

**Eugene:** Now I'm sure that you are one of those crappy robots from "

password:  $2 + 2 = 4.0!$  Now divide by zero and die!!!!

## Turing Test

Kriterium für das Vorliegen allgemeiner Intelligenz bei Computern (Turing, 1950)



'our

me where

ZEIT ONLINE (Uneingeschränkter) Turing Test ist nicht gelöst!

Künstliche Intelligenz

Computerprogramm gaukelt erfolgreich  
Torschsein vor

Qu. Erstmals haben Menschen einen Computer für eine reale  
Eug. herausfordert. In einer Person.  
secn. 9. Juni 2014, 16:34 Uhr / Aktualisiert am 9. Juni 2014

quelle: ZEIT ONLINE, rav /



## —Abstraktion, Reflektion und Bewusstsein—

$$3 + 8 = 11$$

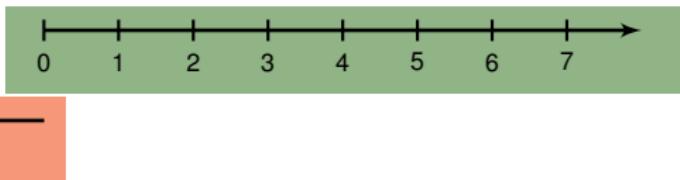
$$15 + 11 = 26$$

$$23 - 9 = 14$$

Transfer

$$2 - 4 = -2$$

$$3 - 6 = -3$$



Trotz signifikanter, aktueller Erfolge in Spezialanwendungen von KI Systemen gibt es derzeit (noch) wenig Evidenz für das Entstehen von Reflektion und Bewusstsein in Maschinen. Die Feststellung einer exponentiellen Innovation im Bereich der KI liefert keine hinreichende Begründung für das Entstehen einer Superintelligenz.

KI hat großes Potential unser Leben signifikant zu verbessern, allerdings müssen wir die Weichen richtig stellen, um Gefahren frühzeitig zu erkennen und ungewollte Entwicklungen zu vermeiden bzw. einzugrenzen.

### Dynamische Bereiche (Selektion)

- ▶ Autonome Fahrzeuge
- ▶ Assistenzsysteme: Siri & Co.
- ▶ **RPA**
- ▶ Internet der Dinge (IoT); 2020:  
  > 20 Milliarden IoT-Geräte
- ▶ Roboter: Haushalt, Industrie
- ▶ Spieleindustrie
- ▶ Bildungswesen
- ▶ Gesundheitswesen
- ▶ Umwelt, Landwirtschaft
- ▶ Administration
- ▶ Strafjustizsystem

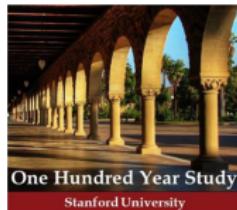
### Kritische Fragen (Selektion)

- ▶ Big Data:
  - ▶ Diskriminierungspotential
  - ▶ Verfügbarkeit vs. Privatheit
- ▶ Security und Safety
- ▶ Arbeits- und Ausbildungsmarkt
- ▶ Ökonom. (Un-)Gleichgewicht
- ▶ Autonome Waffensysteme
- ▶ Kontrollier- und Steuerbarkeit der Entwicklungen
- ▶ Gesetzliche Rahmenbedingungen
- ▶ Ethik und gesellschaftl. Normen

# ARTIFICIAL INTELLIGENCE AND LIFE IN 2030

ONE HUNDRED YEAR STUDY ON ARTIFICIAL INTELLIGENCE | REPORT OF THE 2015 STUDY PANEL | SEPTEMBER 2016

<https://ai100.stanford.edu/2016-report>



## THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN

National Science and Technology Council

Networking and Information Technology  
Research and Development Subcommittee

## PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE

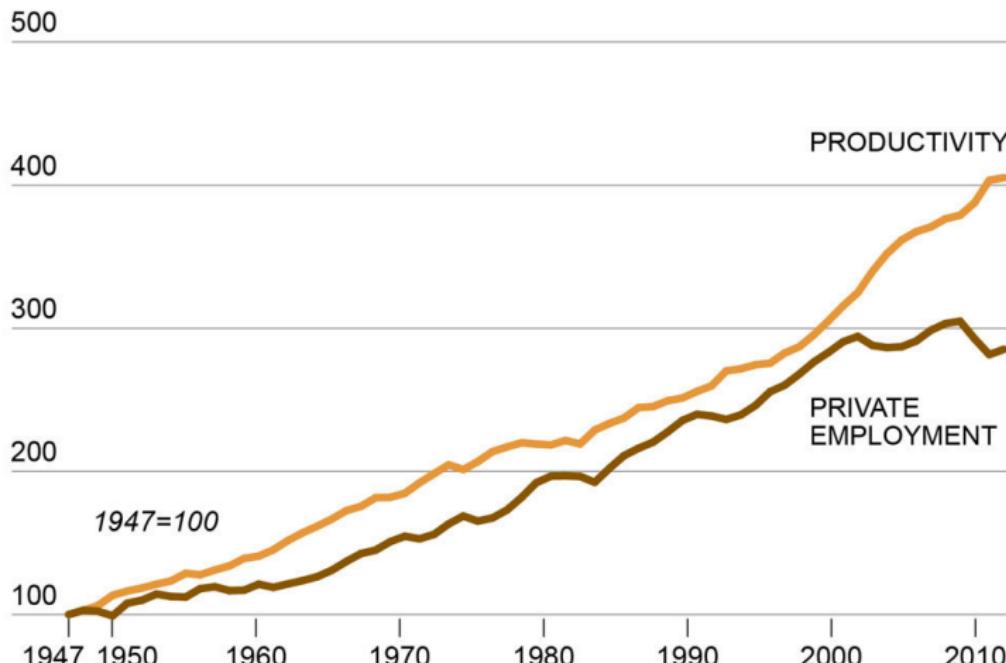
Executive Office of the President  
National Science and Technology Council  
Committee on Technology

October 2016



R&D Strategy .....	15
Strategy 1: Make Long-Term Investments in AI Research .....	16
Strategy 2: Develop Effective Methods for Human-AI Collaboration .....	22
Strategy 3: Understand and Address the Ethical, Legal, and Societal Implications of AI.....	26
Strategy 4: Ensure the Safety and Security of AI Systems.....	27
Strategy 5: Develop Shared Public Datasets and Environments for AI Training and Testing.....	30
Strategy 6: Measure and Evaluate AI Technologies through Standards and Benchmarks.....	32
Strategy 7: Better Understand the National AI R&D Workforce Needs.....	35

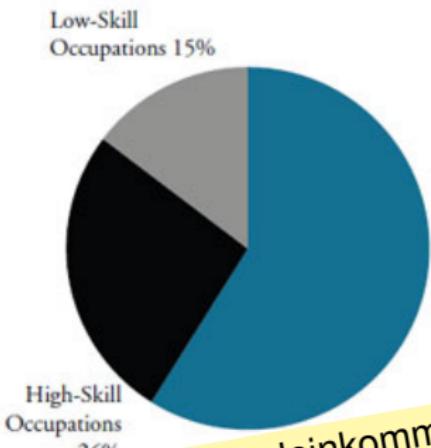
Source: Click here

**Productivity and employment in the United States, 1947-2011**

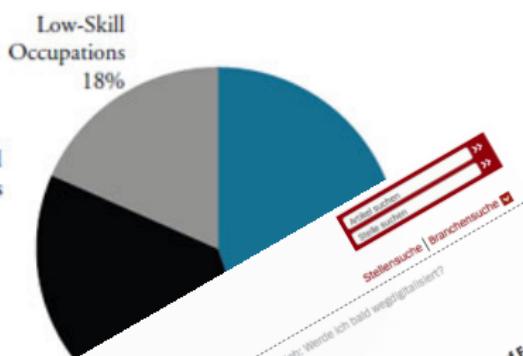
Sources: U.S. Department of Labor, Bureau of Labor Statistics

## EMPLOYMENT SHARES BY SKILL LEVEL

1983



2012



**Bedingungsloses Grundeinkommen & geeignete Besteuerungsmaßnahmen für KI Technologien als Antwort?**

U. Tulzemen and J. Willis, The Vanishing Middle: Job Polarization and authors

Jobs, Federal Reserve Bank of Kansas City in its journal Economic Perspectives

Zukunft der Arbeit  
**Werde ich bald wegdigitalisiert?**

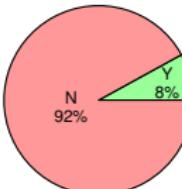
Roboter krempeln den Arbeitsmarkt um, Maschinen werden Millionen Jobs übernehmen. Mit fünf Fragen können Sie erkennen, ob Ihr Job in Gefahr ist - und rechtzeitig gegensteuern.

Von Jens-Uwe Meyer

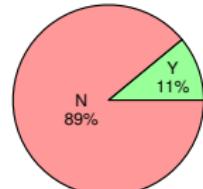
### Frage an Studenten

Wird es bis 2050 eine **Superintelligenz oder Singularität** geben?

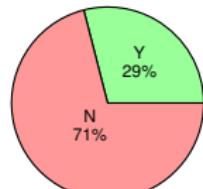
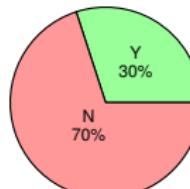
Tag 1 ( $\sigma = 64$ )



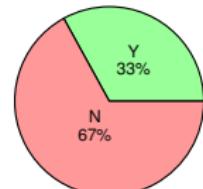
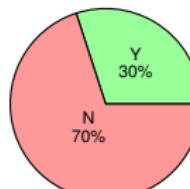
Tag 2 ( $\sigma = 54$ )



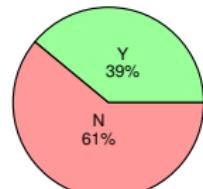
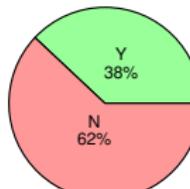
Können Maschinen **Gefühle und Emotionen** entwickeln?



Brauchen wir spezielle **Rechte für Maschinen**?



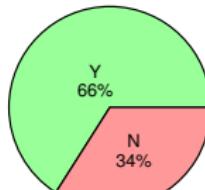
Können Maschinen "**Bewusstsein**" erlangen?



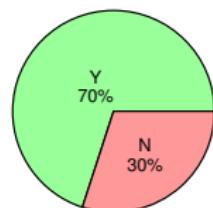
### Frage an Studenten

Werden wie bis 2050 überwiegend **autonome Fahrzeuge** haben?

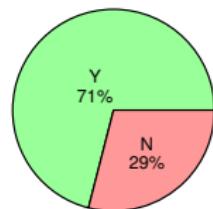
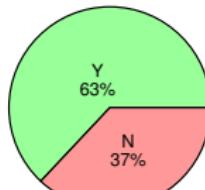
Tag 1 ( $\sigma = 64$ )



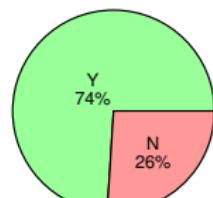
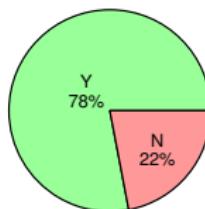
Tag 2 ( $\sigma = 54$ )



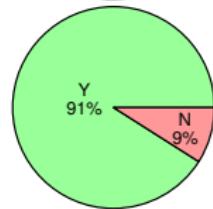
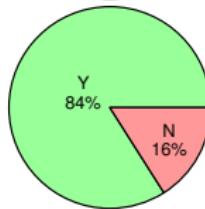
Sollten wir KI Technologien in bestimmten Bereichen **bannen**?



Können Maschinen "**Intelligenz**" erlangen?



Brauchen wir spezielle **Gesetze und ethische Regeln** für Maschinen?



## Beschleunigte Entwicklung von KI-Technologien

- ▶ Exponent. Innovation: Großer Wandel innerhalb einer Generation
- ▶ Aber keine Evidenz für Superintelligenz
- ▶ Menschliche Fähigkeiten von Maschinen in vielen Einzeldomänen
- ▶ Große Chancen
- ▶ Große Herausforderungen

## Mögliche Maßnahmen

- ▶ Bildung und Ausbildung
- ▶ Gut informierte Entscheidungsträger
- ▶ Investition in KI-Forschung & KI-Entwicklung
- ▶ Investition in flankierende Maßnahmen
- ▶ Entwicklung einer KI-Ethik
- ▶ Sachliche Aufklärung in den Medien

## Autonome Maschinen in der realen Welt

- ▶ Gesetzliche Rahmenbedingungen
- ▶ Verantwortung und Haftbarkeit bei Fehlverhalten und Schäden
- ▶ Trennung von Entscheidungsfindung und Entscheidungsbewertung?
- ▶ Integration in bestehende Normen-/Wertesysteme
- ▶ Modellierung einer ethisch-moralischen Wertesysteme für Maschinen?
- ▶ Morale Dilemma, konfliktäre Wertesysteme
- ▶ Gesetze zum Schutz von Maschinen?

## Verwendung von "Big Data" in KI-Systemen

- ▶ Diskrimierung
- ▶ Privatheit von Daten
- ▶ Zugang
- ▶ Fairness

Notwendig: Stärkung der flankierenden Forschung zu solchen Themen